Eötvös Loránd University

Faculty of Arts

ABSTRACT BOOKLET

Váradi Viola

A SPONTÁN BESZÉD SZEGMENTÁLÁSA PRODUKCIÓS ÉS PERCEPCIÓS SZEMPONTBÓL

Segmentation of Spontaneous Speech from the Aspects of Production and Perception

Doctoral School of Linguistics

Head of the school: Prof. Dr. Vilmos Bárdosi CSc

Program of Applied Linguistics

Head of the program: Prof. Dr. Mária Gósy DSc

Members of the PhD-committee:

Head of the committee: Prof. Dr. Nyomárkay István, professor emeritus

Reviewers: Prof. Dr. Keszler Borbála DSc, professor emeritus

Dr. Horváth Viktória PhD

Secretary of the committee: Dr. Raátz Judit PhD

Supplemantory members of the committee: Dr. Adamikné Jászó Anna DSc;

Prof. Dr. Olaszy Gábor DSc; Dr. Menyhárt Krisztina PhD

Supervisor: Prof. Dr. Gósy Mária DSc

Budapest, 2013.

1. INTRODUCTION

One of the main questions of speech research is to what units spontaneous speech can be segmented into, and what acoustic phonetic features differentiate these from each other. The productory features and the perceptual characteristics of the segmentation of spontaneous speech is of great importance to reveal the strategies of both the speaker and the listener, since proper segmentation is an elementary requirement of understanding and interpretation of fluent speech (Gósy 2003).

The abilities of segmentation and grouping have been proved to be important parts of the perceptual mechanisms (Fon 2002). A clear example for this is the transformation of two-dimensial information into three dimensions during visual perception. According to Marr's three-steps model (1982) the segmentation of visual objects is based on the intensity change of simple physical signs, like that of light. Segmentation plays an important role in the auditive perception, as well. The identification of words, sentences, or larger units does not cause any problem for typical speakers. However, not all speakers apply consciously the articulatory sings for segmentation during their speech. Inexperienced speakers' speech is often segmented by physiological phenomena, like breathing.

This leads in Shriberg's and her colleagues' theory (2000) to call segmentation a challenging task. Children during acquiring their mother tongue face the problem of segmentation, as well: In enlargement their vocabulary, the segmentation of speech into words and larger units plays a crucial role. Speakers are not always aware of what segmental and suprasegmental features they apply for segmentation, i.e. the articulatory execution is usually subconscious and its interspeaker differences are large (Kohler 1983).

Segmentation plays a crucial role in both speech production and speech perception. The most common speech production model is that of Levelt (1989, 1999). This serial structured model is based on the results of experiments that applied mental cronometric tasks. The model is built up from five larger phases of two systems (rhetoric/semantic/syntactic and phonological/phonetic ones). According to this model, production of utterances does not mean combining words but utterances are planned in intonation phrases built up from phonological phrases. The conceptual planning stands of two main processes: the macroplanning and the microplanning. During macroplanning, the speaker decides what he or she wants to say and what the reason and goal of their speech is. The speaker accesses the relevant information, and sets the focus of the discourse during this phase. Microplanning means the linguistic shaping of the message. The speaker's perspective is chosen in this phase, i.e. the person whose point of view the speaker takes in the conversation.

One of the most important questions of speech perception research is how the listener perceives the continuously changing acoustic stimuli, how the analogue sign chain is transformed into the chain of discrete speech units, how her or his understanding and interpretation of the message is performed (Pisoni and Luce 1987). Cutler and Clifton's speech perception model (1999) introduces the entire operation of the mechanism. It takes both the suprasegmental features and the process of segmentation into consideration. Four phases of speech perception are distinguished by this model: i) decoding of the pieces of information about speech, ii) segmentation of speech that corresponds to the identification of

word boundaries in this model, iii) word identification and interpretation of the utterance, iv) integration into the discourse model.

The units that can be used to describe the structure of speech are not only a question in applied linguistic research but in theoretical approaches as well. Determination of the suprasegmental units is a basic requirement of the analysis of the prosodic structure. Speech sound is the basic unit in the analysis of the segmental structure but in the prosodic research, consensus have been met regarding neither the terminology nor the definitions (Markó 2009).

Hierarchical structure is most often hypothesized in the prosodic descriptions of speech (Gussenhoven 2004, Roca–Johnson 2005, Varga 1994, Hunyadi 2002). Proceeding from top to bottom, the following levels can be distinguished: utterance, intonation phrase, phonological phrase, phonological word, foot, syllable. The question arises if this structure is applicable on spontaneous speech as well.

Listeners can rely on various potential boundary mark keys during the segmentation of speech. None of these keys is sufficient or clean-cut boundary marker on its own, but their complex combination is typical (Shriberg et. al 2000, Warner et. al 2004, Gósy–Kovács 2008). This is the consequence of the characteristics of speech productory mechanisms: speakers realize the prosodic segmentation of their speech in various ways. (Frazier et al. 2003).

This PhD-thesis aims to describe segmentation of spontaneous speech from the point of view of both the production and perception, and also to reveal the connections of the two processes. I shall try to approach to understanding speech segmentation through the analysis of prosodic units, the segmentation into utterances and into paragraphs. I attempt to reveal data on the characteristics of the segmentation processed, and the strategies applied by speakers and listeners through the empirical analysis of a large speech corpus.

2. THE QUESTIONS AND MAIN HYPOTHESES OF THE PHD-THESIS

In my research, I try to find the answer for the following questions:

(i) What acoustic phonetic features describe the prosodic units in Hungarian?

(ii) What role do pauses and prosody play in the identification of virtual sentences?

(iii) Does the speech type or mode influence the characteristics of virtual sentences?

(iv) Does the segmentation strategy change through the life-span?

(v) What acoustic phonetic features describe those units of Hungarian spontaneous speech that are larger than utterances?

In my research I have analyzed the segmentation processes in Hungarian speech production and perception. One of my main goals is the analysis of the acoustic phonetic realization of the prosodic units of Hungarian speech from diverse points of view. Only a few studies have been made on Hungarian prosodic unit analysis (Olaszy 2006, Szaszák and Beke 2012), thus this PhD-thesis attempts to fill this lack. The characteristics of spontaneous speech are compared to those of read speech in order to be able to draw conclusions on characteristics of the production of spontaneous speech.

The other main goal of my research is the traditional linguistic categories based phonetic analysis of the segmentation mechanism in Hungarian speech. My analysis concerns the effect

of speech mode, speech type and age on the sentence-level segmentation strategies. With the analysis of the paragraph-level segmentation, I aim to reveal data on the acoustic phonetic characteristics that drive the segmentation into units that are longer than utterances.

I set the following hypotheses:

i) The existence of prosodic units, i.e. speech phrase and intonation phrase, is verifiable in speech, and it has temporal and prosodic correlates.

ii) Since the differences in the planning processes of reading aloud and spontaneous speech, I hypothesize these units to realize differently in the two speech modes.

iii) Greater variability is assumed in the realizations of the above mentioned prosodic units in spontaneous speech than in reading aloud.

iv) In my assumption speech mode and speech type have influence on the listeners' sentence-level segmentation strategies.

v) I also hypothesize that also in spoken language, there are coherent units that are longer than the utterances and that during segmentation into these units, i.e. thought-units, listeners rely on prosodic and temporal factors to a great extant.

3. METHODS

The analysis of the prosodic units, that of the virtual sentences both in reading aloud and spontaneous speech, and of paragraphs were carried out on recordings chosen from a Hungarian speech database, BEA. The database aims to collect a large number of high-quality recorded speech material from speakers of various ages, qualifications and occupations under the same laboratory conditions and circumstances appropriate for phonetic analysis. The development of the database started in 2007. It has been recorded at the Phonetic Laboratory of the Research Institute for Linguistics of the Hungarian Academy of Sciences. It contains spontaneous speech, reading aloud of sentences and a text, as well, as sentence repetition and interpreted speech. The recorded spontaneous speech samples are various, the speakers i) are interviewed on their everyday life and hobbies, ii) they tell their opinion on a topic of a current interest and participate in a three-member conversation on an everyday topic. The recordings are always carried out at the same place, under the same circumstances, in a silent-treated room. The speech material is recorded digitally, directly to computer with the means of GoldWave at 44.1 kHz, 16 bit, 86 Kbytes/s, mono (Gósy 2012).

In order to analyze the utterances in different spontaneous speech types, I recorded a speaker talking about her experiences, telling a tale and describing her apartment under the circumstances and conditions of BEA. The speaker was a 27 years old woman with a diploma of master degree. She had neither speech disorders, nor any hearing loss. She was aware of being recorded.

In the analysis of the prosodic units the reading aloud and the spontaneous narrative of eight (four female, four male) speakers of the BEA database were used. All speakers are monolingual and use standard Hungarian, their speech and hearing are unimpaired. Their mean age is 28 years.

The perception of virtual sentences in both read and spontaneous speech was carried out using perception tests. 26 women participated in this experiment. The listeners had no hearing

loss, speak standard Hungarian. They all were students of Hungarian linguistics and literature at the time of the experiment. Their mean age was 22 years.

The effect of age and speech mode on the sentence-level segmentation was investigated also via a perception test. Altogether 54 (28 female and 26 male) listeners' answers were analyzed. Their age was between 13 and 54 years (mean: 25 years). The subjects were divided into three age groups: children (mean age: 14 years), young adults (mean age: 22 years), middle-aged adults (mean age: 40 years).

In the perception test carried out in order to analyze the paragraph-level segmentation, 45 female students participated. Their mean age was 24 years, all had no hearing impairment. They were monolingual standard Hungarian speakers.

The applied audio recordings of each experiments were manually annotated by me using Praat software (Boersma and Weenink 2007, 2012). The annotation was based on the spectrogram, the oscillogram and continues auditive monitoring. The acoustic analyses were carried out with the means of the Praat software, as well.

The listeners of the perception tests were instructed to mark the places in the written form of the text with a vertical line where they judge to have heard utterance- or paragraph-ending. Each speech sample was played twice to the subjects.

Statistic analysis was carried out with the means of SPSS 13.0. Since several statistical analysis requires normal distribution the Shapiro–Wilk normality test was run for each data set. In case of normal distribution parametric tests (one way ANOVA, general linear model with Tukey post hoc test and repeated measures) were run to reveal information statistical features of the data sets, while in case of non-normal distribution non-parametric tests (Mann–Whitney, Kruskal–Wallis, Spearman correlation analysis) were applied.

4. PROSODIC UNITS IN HUNGARIAN SPEECH

The mean duration of the speech phrases was 2550 ms in reading aloud, while in the spontaneous narratives their length was shorter, 1674 ms. The mean of the duration of the intonation phrases was 915 ms in the first speech mode, while in the second, they were somewhat shorter again, 884 ms. The difference of the duration of the intonation phrases in the two speech modes was not significant, while that of the speech phrases was.

The duration of the speech phrases and the intonation phrases was analyzed speaker by speaker, as well, in both speech modes. The mean duration of these was counted by adding up the total speech time in one speaker's speech in one speech mode, and divided by the number of the given phrases. The duration of the speech phrases was longer in the case of each speaker in reading aloud than in the spontaneous speech.

The articulation tempo was analyzed in the intonation and speech phrases of both the reading aloud and spontaneous speech (Fig. 1). The results showed that the mean of these values was 12.87 sounds/s in spontaneous speech, while it was higher, 14.46 sounds/s in reading aloud. The dispersion of these values was higher in the first speech mode (σ = 3.59 sounds/s), than in the latter one (σ = 1.36 sounds/s). This means that the variability of the articulation rate in reading aloud was less variable than in the spontaneous speech.

The mean of the articulation rate of the intonation phrases in the reading aloud was 14.46 sound/ s, while in the spontaneous speech it was somewhat lower, 13.98 sounds/s. This

difference is significant according to the statistical analysis. The variability of the articulation rate of the intonation phrases in spontaneous speech was also higher (σ = 3.99 sounds/s) than in reading aloud (σ = 2.01 sounds/s). The articulation rate of the intonation phrases was realized in a larger interval (reading aloud: 18.41 sounds/s, spontaneous speech: 27.85 sounds/s) in both speech modes than that of the speech phrases (reading aloud: 9.45 sounds/s, spontaneous speech: 22.13 sounds/s).

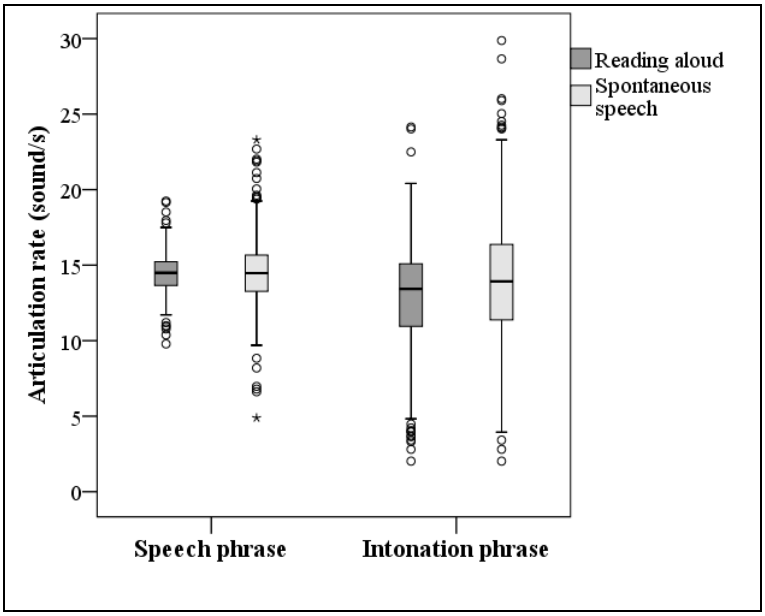

Figure 1. Articulation rate (sounds/s) in the intonation and speech phrases of reading aloud and spontaneous speech

The variability of the articulation rate was analyzed in the intonation phrase[1]. The following factors were hypothesized to influence this characteristics: the position of the word in the intonation phrase, the position of the intonation phrase in the speech phrase, the number of syllables of the word, the parts of speech type of the word, the category of the word, speaker. The two-, three- and four-words long intonation phrases were analyzed in the research. GLM (general linear model) model was built up from the five above mentioned factors to interpret the results of the intonation phrases of the three different lengths. The resulting model was found to be efficient, because the 80.2% of the variability in the two-word long, 77.9% of that in the three-words long and 63.4% of that in the four-words long intonation phrases could be interpreted by the model. Including the speaker as an influencing factor had decreased the interpreting strength of the model, thus it was left out during the further analysis. The position of the intonation phrase in the speech phrase was not significant in any cases. The position of the word in the intonation phrase itself had an important effect on the results in the case of the three-word long phrases. According to the η^2-data, the parts of speech type of the word was the strongest influencing factor in the results. However, further

[1] The analysis of the variability of articulation rate in the speech and intonation phrases was done together with András Beke (Váradi and Beke, in press).

analysis is required to figure out what roles the non-equal distribution of the various parts of the speech and the differences of the length of these plays in the results.

The model built on the results of the used spontaneous corpus is less effective than that built on the read corpus. The included factors are able to interpret only lower percent of the variability of the articulation rate in the intonation phrases than in read speech. Only 11.9% of the variability of the articulation rate could be interpreted by the applied five factors in the three- and four-words long intonation phrases and 11.5% of that in the two-word long ones. Including the category of parts of speech had decreased the interpreting strength of the model, this it was left out in the further analysis. The speaker itself was proven to be the strongest influencing factor in the three models of the different word-lengths bas on the η^2-data – opposed to reading aloud where including the speaker as an interpreting factor decreased the strength of the model. This difference between the two speech modes suggests that the interspeaker difference is higher in the case of spontaneous speech than in reading aloud. However, it should be kept in mind that the two speech modes did not only differ in their planning processes, but in their language material, as well.

The way of the variation of articulation rate was also analyzed both in reading aloud and spontaneous speech. A tendency curve was laid on the articulation rate values of the words that the intonation phrase in question consists of. The shape of this curve was described and analyzed. This way we were able to analyze the speeding (accelerando) and slowing (rallentando) in the intonation phrase. The two-, three- and four-words long ones were analyzed in the two speech modes. In reading aloud, the rallentando tendency was the most frequent realization. In the most of the intonation phrases, the tempo slowed down at the end that the listeners can apply as a cue during segmentation. In opposite, in spontaneous speech this tendency did not apply, the tendencies were either speaker-specific. This interspeaker difference of the realized articulation rate structures suggests that the timing structures of spontaneous and non-spontaneous speech are different.

The pitch ranges of the speech phrases and intonation phrases were also compared in both reading aloud and spontaneous speech. The pitch ranges of the speech phrases were significantly different in the two speech modes. In reading aloud their mean was 1.9 Hz, while in spontaneous speech it was 1.8 Hz. The data were more variable in the spontaneous speech (σ = 0.64 Hz), than in the other speech mode (0.51 Hz).

5. SENTENCE-LEVEL SEGMENTATION IN HUNGARIAN

5.1. Sentence-level segmentation with regard to the speech mode

The least number of marked sentence boundaries by one listener was 11 in the case of the read text. The highest number was 14. The 57.6% of the listeners perceived 13 sentences in the read text that also originally consisted of 13 sentences. The mean of the marked boundaries was 12.73. The number of the marked utterances was more variable, it ranked from 4 to 19.

Because of the difference of the duration of the two texts, not the number of perceived boundaries was applied in my comparison, but the highest number of boundaries marked by the same speaker was set to 100% (Fig. 2).

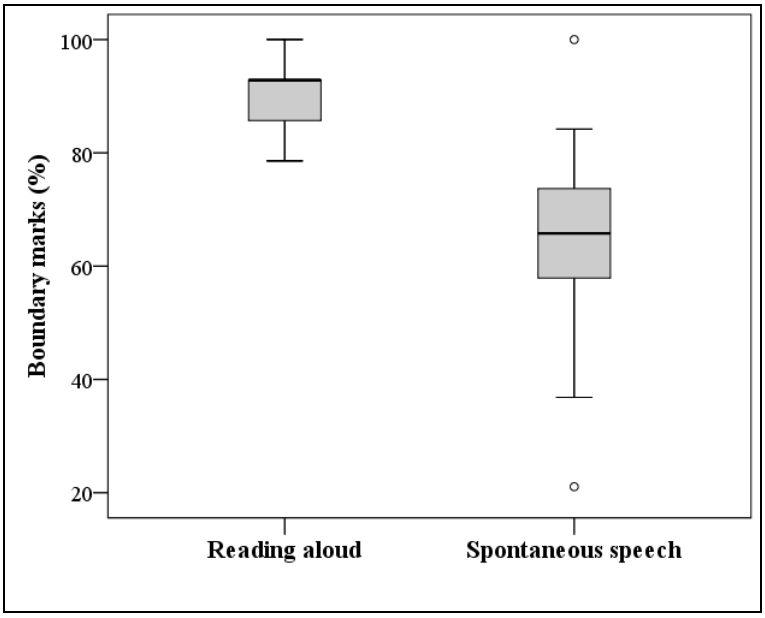

Figure 2. Sentence-boundary marks in the two speech modes

As the box plots show the number of marked boundaries are less variable in reading (SD = 5.56) than in the spontaneous speech (SD = 16.10). The mean is 64% in the first, and 91% in the second. The difference of the two speech modes is significant.

The boundary marks set at the end of speech phrases were analyzed in both speech modes. In reading aloud, 70% of the ends of the speech phrases were marked at least by one listener, while in the case of the spontaneous speech sample, only 23% of them.

There were only two speech phrase ends in reading that not all listeners marked as a boundary regardless of the duration of the pause following it. The correlation of the length of the pause and the number of boundary marks was proven to mid-strong and significant according to the Spearmen correlation analysis. In spontaneous speech no such interrelation was found according to the Pearson correlation analysis.

The effect of pitch movement on the sentence-boundary perception was analyzed, as well. The characteristics of pitch were analyzed at the places where at least one listener had marked perceived boundary. In reading aloud, floating pitch triggered the most often sentence-end perception (53.75%), while in spontaneous speech pitch decrease did (45.26%). Boundary mark set inside a pitch movement was found only in the spontaneous speech; however, it appeared only rarely (6.93%). Boundary marks often were set after creaky voice in reading aloud (38.44%).

The text points were further analyzed where more than the 75% of the listeners ($\geq$ 20) marked boundary. In reading aloud, these points were the 92.31% of the points where at least one listener marked a boundary, while in spontaneous speech they made up only the 32% of them. All text points in reading, and all-but-one in spontaneous speech, where at least 75% of the listeners marked boundary, appeared at speech phrase end. Floating pitch contour triggered the most often boundary marking by more than 75% of the listeners in both speech modes. 40.65% of these points appeared after decreasing pitch in spontaneous speech. In reading pitch decrease did not trigger boundary mark by more than 75% of the listeners (8.39%), while the effect of creaky voice was the other way around: it often triggered

boundary mark by the most speakers in reading (41.29%) but did not in spontaneous speech at all.

5.2. Sentence-level segmentation with regard to age and speech type

At least one listener marked boundary at 106 points in the experience telling, at 109 points in the tale and at 54 points in the description. The number of all boundary marks in the experience telling was 946, in the tale 1315 and in the description 766, i.e. the most marks were set in the longest, the least in the shortest. The same way as in the above described research, not the number of marked boundaries were used in the further analysis, but the highest number of marks set by the same speaker was set as 100%. The 100% was equal to 44 marks in the experience telling, 53 in the tale and 25 in the description. The effect of the listener's age on the utterance-perception was analyzed (Fig. 3).

The most boundaries in the experience telling was marked by the middle aged speakers (56.44%), the least was by the children (33.71%). The dispersion of the data was also different among the age groups. The lowest variability was found in the children's results (σ = 9.71), the middle aged listeners' markings were more variable than that (σ = 16.10), and the largest interlistener difference was found in the young adults' group (σ = 21.50). The boundary marks in the tale showed similar tendencies (middle aged listeners: 54.53%, young adults: 43.40%, children: 42.70%). The interlistener variability was the highest in the middle aged group (σ = 19.85), it was lower in the young adults' group (σ = 14.59), and the lowest in the children's (σ =12.95). The most boundaries were marked by the middle aged listeners in the description, as well (63.06%), the young adults marked less (54.94%), and the children did the least again (47.19%). The variability of the data in this speech type was similar to that in the experience telling (children: σ = 12.72, middle aged listeners: σ = 18.08, young adults: σ = 23.06). The differences among the listeners' groups are not significant in any speech types, but the difference among the speech types is significant in all age groups according to the repeated measure ANOVA.

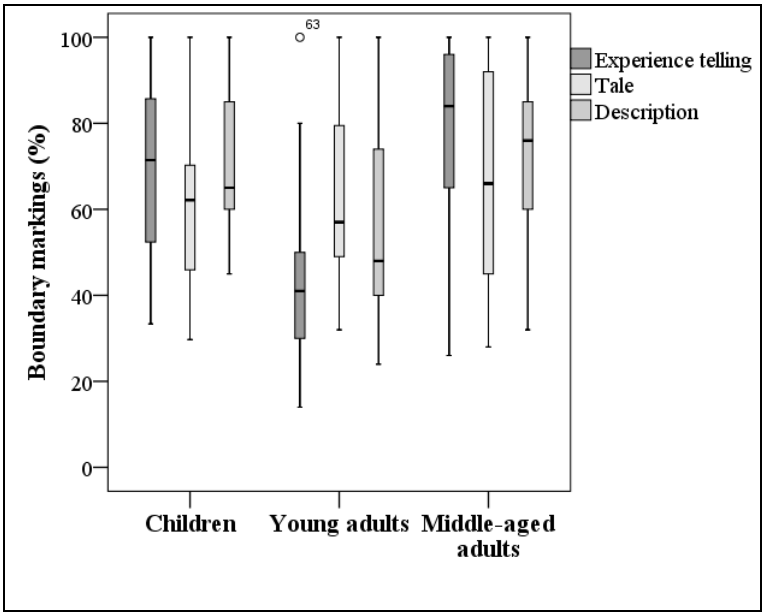

Figure 3. Sentence boundary marks in the three age groups and in the three speech types

The number of marks was analyzed in all texts with regard to the speech type and the age groups. The number of listeners was set to 100% in each age group (children: 18, young adults: 19, middle aged listeners: 17) and the number of marks put at the same text point was calculated in ratio compared to the number of the subjects in the group.

In the experience-telling, the mean of marks at the same text points was 28.45% in the children's, 29.07% in the young adults' and 25.16% in the middle aged listeners' group. The highest frequency of appearance was found in each speech type at points where only one subject had put a boundary mark (from the youngest to the oldest group this frequency was: 32.03%, 26.87%, 31.33%). The age groups did not show any significant differences.

The young adults put the most boundary point in the tale, too (36.76%). There is no difference between the children's and the middle aged group (32.84% and 32.93%, respectively). The marks at points where only one subject had put a boundary mark were the most frequent again; however, they did not appear as frequently as in the previously described speech type. The tendency was though the same: it added up the 25% of all boundary marks in the children's, 22.58% in the young adults', and 24.14% in the middle aged group's answers, and no significant difference was detected among the age groups.

The less variability in boundary marks was found in the description. The 36.70% of the children, 43.29% of the young adults and 36.55% of the middle aged listeners had put boundary mark at the same text point on average. The difference among the age groups followed the tendencies of the other two speech types: The young adults had put the highest number of boundary marks, and there was no difference between the other two groups. Also the most frequent appeared type of boundary marks was where only one listener had put a boundary mark (children: 31.58%, young adults: 12.90%, middle aged group: 21,43%). The young adults and the middle aged listeners had marked the least one-listener-perceived boundaries in this speech type, while the children had in the experience-telling. All three age groups had put the most one-listener-perceived boundary mark in the tale. The difference between the age groups reached the statistically significant level neither in this speech type.

The appearance of boundary marks put by most subjects (more than 75% of the listeners in the group: children ≥ 14, young adults ≥ 15, middle aged group: ≥ 13) was analyzed in this experiment, as well. All occurred at pauses, at speech phrase end.

The most text point where most ($\geq$ 75%) of the listeners in a group marked boundary appeared the most often in the description regardless of the age group itself (children: 34.78%, young adults: 28.00%, middle aged group: 27.27%). The experience-telling showed the less frequent occurrence of this kind of text points (young adults: 11.63%, middle aged listeners: 4.44%, children: 0%). In the tale, it appeared in the highest ratio in the eldest age group's answers (27.04%), while it was less frequent in the young adults' ones (25.00%), and the children put the less frequently their marks at the same text points (15.38%).

At least one listener marked boundary at the 64% of the speech phrase ends in the experience-telling and in the description and at 72% of those in the tale. The distribution of boundary marks at these places was also analyzed with regard to the age of the listeners.

The children marked boundary at 53.23% of the speech phrase ends, the young adults did at the 51.79% and the middle aged listeners did at the 55.56% of them in the case of the experience-telling. The number of people who put mark at the same phrase end was also calculated. Its mean was 15.14% in the children's, 22.07% in the young adults' and 20.30% in

the middle aged listeners' group. The 28.45% of the children, 42.76% of the young adults and 35.97% of the middle aged listeners perceived sentence boundary at the boundary marks set at phrase ends.

The middle age group put the most often boundary mark at the phrase ends in the tale (64.38%), while the children marked boundary less frequently at these points (60.27%) and even less boundary marks were found in the young adults' answers at phrase ends (58.90%). Also the eldest aged listeners put most often boundary marks after pauses (31.27%), the second eldest group marked the 27.40% of the pauses as boundary, and the children did only the 26.03% of them. The most boundary marks that were placed at the same text point by at least the 75% of the listeners followed the same tendency. The 43.18% of the children's boundaries at speech phrase ends was marked at least by the 75% of them, 46.51% of that of the young adults and 48.56% of the middle age group's. The intergroup difference is neither in this speech type significant.

The description showed a different tendency than the above described two speech types. The young adults marked boundary at the 57.14% of the speech phrase ends, the children at the 54.28% and the eldest group at the 51.43% of them. The 31.43% of the children's boundary marks was set at speech phrase end, the 31.26% of the second eldest age group's, while only the 28.57% of the middle aged listeners' marks appeared at speech phrase ends. The 61.78% of the boundaries marks at speech phrase ends was set at least by the 75% of the listeners in the middle aged group, this ratio was lower, 57.89% in the children's answers and only 50% in the third age group. The difference among the listener groups did not reach the level of significance.

The results for the three speech types were compared in the three age groups. In the case of the children, the three speech types differed significantly. The experience-telling was proved to differ significantly from both the tale and the description according to the Tukey post hoc test. Also in the case of the middle aged group, the general results were found to be significantly different, however, according to the Tukey post hoc test, only the difference of the experience-telling and the description reached the level of significance. In the third age group, no statistical difference was found among the speech types.

The interrelation of the duration of the pauses and the frequency of boundary marks was also analyzed. I hypothesized that longer pauses will trigger the perception of end, completeness in more listeners, i.e. the longer the pause is the more marks appear at the same text point. No text points were found in the case of the experience-telling in any of the age groups where all listeners had put boundary mark. The duration of the pause and the number of the boundary marks at the same text point showed a middle-strong correlation according to the Spearman correlation analysis in each age group.

The interrelation of the pause length and the boundary marks was analyzed in the tale, as well. It also showed a middle-strong correlation according to the Spearman test.

The description gave somewhat different results. In the case of the young adults, no correlation, while in the other two age groups, significant strong correlation was found in the pause duration and the number of boundary marks.

The type of pitch movement was analyzed in all three age groups in all three speech types at all text points where at least one speaker had marked boundary. In experience-telling, the most of the boundaries appeared after raising pitch in all three age groups (children: 47.67%,

young adults: 26.96%, middle aged group: 42.00%). Over the fourth of the marks occurred after floating pitch in all age groups. Boundary marks appeared less frequently after decreasing pitch in all subject groups (children: 19.38%, young adults: 23.44%, middle age listeners: 16.29%). Boundary marks were rarely put at quickly rising pitch, creaky voice and laughter or inside the pitch contour.

The most of boundary marks appeared after floating pitch in all three age groups in the tale (from the youngest to the eldest: 38.87%, 34.10% and 40.45%). Decreasing pitch also often triggered boundary perception (in the same order: 34.78%, 27.42% and 31.83%). The boundary marks appeared the third most often after rising pitch (in the same order: 18.16%, 25.80% and 21.15%).

Over the half of children's boundary marks (52.59%), over the third of that of the young adults (38.71%) and nearly the half of those in the third age group (42.15%) were set after rising pitch. The children and the middle aged groups more often set boundary mark after decreasing pitch (18.33% and 21.07%, respectively), than after floating pitch (16.33% and 20.69%, respectively). Young adults put boundary marks after both types (22.58%).

6. PARAGRAPH-LEVEL SEGMENTATION IN HUNGARIAN

In the experiment where the listeners' task was to mark paragraph boundaries 76 text points were marked at least by one of them and the total number of marks set was 579. The mean of the number of paragraph boundaries marked by one speaker was 12.87 (SD = 7.39). These results show high interlistener differences (Fig. 4). The mean duration of the paragraphs was 34 s.

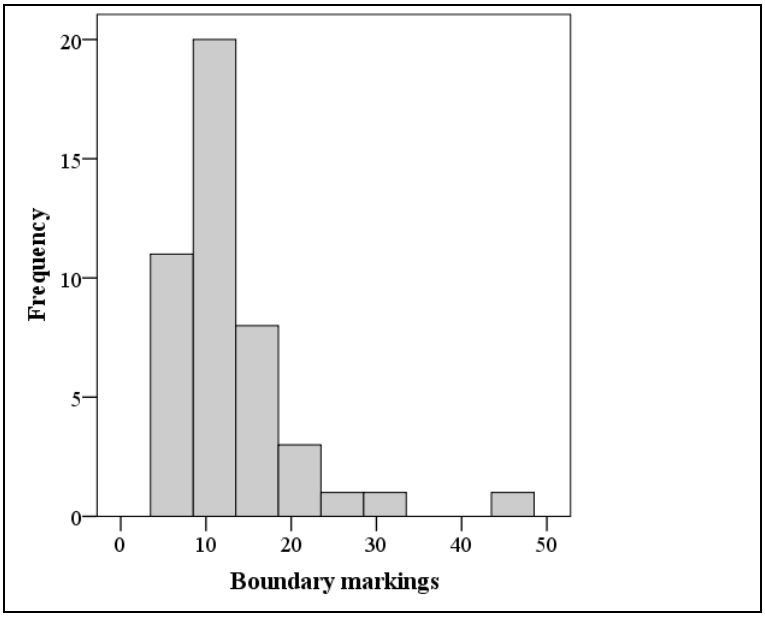

Figure 4. The distribution of the number of paragraph boundary marks set by one speaker

The lowest number of paragraph boundary set by the same speaker is six (it appeared three times). The highest number of that is 45 (seven times more marks than the lowest, it appeared once). This results means that there were listeners who considered speech parts as a paragraph that were longer than one minute, while others segmented the text into near 10 s long parts.

23 out of the 76 boundary places (30%) appeared at text points where no pause appeared. In the most of these cases the boundary was marked only by one listener. The highest number of paragraph boundary set inside speech phrases by one speaker was 10. That also means that only the 22.22% of the listeners marked together at such places.

Boundary marks set at pauses were also most often marked only by one listener but at the same time the number of marks set at the same place by several subjects was more frequent than in the case of those set inside a speech phrase.

The interrelation of pause duration and boundary marks was also analyzed. This interrelation did not reach the level of significance: longer pauses did not trigger marks by more listeners than shorter pauses did.

Most of the boundary marks (98%) were found at the end of syntactic structures. That means that grammatical complexity is a relevant characteristic of thought-units in spontaneous speech. Only six out of the 76 marked boundary points were set inside a syntactic unit that means 11 markings. Four out of the 11 markings (five listeners marked at the same place, the other places were marked only by one subject) were set at text points where no pauses appeared. Two further of these marks were put at text point where a short, shorter than 500 ms pause was to be found (two markings at a 272 ms long pause, one at a 417 ms long one).

The interrelation of pitch movement and paragraph boundary marks was analyzed, as well. The boundary marks at the different types of pitch contours were calculated in the ratio of all marks. Near the half of the boundary marks were to be found at creaky voice. Marks appeared less often after floating (26%) and rising pitch (22%), and only 4% of them were set after decreasing pitch.

The 32% of the text points where at least one listener marked boundary were marked only by one listener. That adds up the 4% of all boundary marks. 80% of the 24 text points where only a couple of the listeners marked boundary was to be found at the end of a syntactic structure. 45.83% of these non-consensual boundary marks were put inside a speech phrase. The most of these appeared at 200 ms long or shorter pauses and the longer the pause was the less of the non-consensual marks appeared. The most of the text points where only one listener marked paragraph boundary ended with floating pitch (45.8%) while these were less frequent at decreasing pitch and creaky voice. 29.2% of the text points marked by only one listener ended with rising pitch.

No text points were marked by all subjects; however 41% of them were marked at least by 75% of the listeners (34 people).

All text points where at least 75% of the subjects marked as a boundary was followed by a pause that was at least 500 ms, was the end of a syntactic structure. Four time creaky voice, twice floating pitch was detected at these points. Besides the syntax and prosody, semantic and pragmatic features also played important roles in paragraph-end percept: the speaker changed topic at each of these points.

7. THESES OF THE PHD-DISSERTATION

The goal of my research was to explore the processes of segmentation. The research was carried out on the analyses of production and perception of spoken language. The results suggest the following theses, conclusions:

1. The realization of speech phrases is more variable than that of intonation phrases. Intonation phrases can be considered as the fundamental element of the suprasegmental structure.

2. Both the production and the perception of virtual sentences are affected by the speech mode. Sentence boundary percept is more stable in reading aloud than in spontaneous speech, i.e. marks at a place are more often set by the most listeners at the same text point in the first one. This difference may be resulted by the difference of the speech planning processes of the two speech modes.

3. The characteristics of the speech types affect sentence-level segmentation. During segmentation, listeners rely heavily on features that are typically realized differently in the speech type in question than in the others.

4. Segmentation strategies are rather similar in the 8^{th} graders', in the young adults' and in the middle aged listeners' perception. The skills that lead to adult-alike segmentation strategies seem to be acquired solidly by the age of 13-14 years. Generational differences don't affect the segmentation strategies.

5. The units longer than sentences are to be found in spontaneous speech. Listeners' cognitive ideas about paragraphs in spontaneous speech differ from each other's.

References

Boersma, Paul – Weenink, David 2007. Praat: doing phonetics by computer. (Version 5.0). http://www.fon.hum.uva.nl/praat/download_win.html

Boersma, Paul – Weenink, David 2012. Praat: doing phonetics by computer. (Version 5.3.32.). http://www.fon.hum.uva.nl/praat/download_win.html

Cutler, Anne – Clifton, Charles Jr. 1999. Comprehending spoken language: a blueprint of the listener. In Brown, Colin M. – Hagoort, Peter (eds.): *The Neurocognition of Language*. Oxford University Press, Oxford. 123–166.

Fon, Yee-Jean Janice 2002. *A Cross-linguistic Study on Syntactic and Discourse Boundary Cues in Spontaneous Speech*. Dissertation, The Ohio State University.

Frazier, Lyn – Clifton, Charles Jr. – Carlson, Katy 2003. Don't break, or do: Prosodic boundary preferences. *Lingua* 1. 1–25.

Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszédkutatás 2003*. 19–43.

Gósy Mária 2012. BEA – A Multifunctional Hungarian Spoken Language Database. *The Phonetician* 104–105. 51–62.

Gósy, Mária – Kovács, Magdolna 2008. Virtual sentences of spontaneous speech: Boundary effects of syntactic-semantic-prosodic properties. In: Gardner-Bonneau, Daryle – Blanchard, Harry E. (eds.): *Human Factors and Voice Interactive Systems*. Springer, New York–London. 361–379.

Gussenhoven, Carlos 2004. *The Phonology of Tone and Intonation*. Cambridge University Press, Cambridge.

Hunyadi László 2002. *Hungarian Sentence Prosody and Universal Grammar*. Peter Lang, Frankfurt am Main–Berlin–Bern–Bruxelles–New York–Oxford–Wien.

Kohler, Klaus J. 1983. Prosodic boundary signals in German. *Phonetica* 40. 89–134.

Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. The MIT Press, Cambridge (MA).

Levelt, Willem J. M. 1999. Producing spoken language: a blueprint of a listener. In Brown, C. – Hagoort, P. (eds.): *The Neurocognition of Language*. Oxford University Press, Oxford. 83–122.

Marr, David 1982. *Vision*. Freeman, San Fransisco.

Olaszy Gábor 2006. Prozódiai szerkezetek jellemzése a hírfelolvasásban, a mesemondásban, a novella és a reklámok felolvasásában. *Beszédkutatás 2006*. 21–50.

Pisoni, David B. – Luce, Paul A. 1987. Acoustic-phonetic representations in word recognition. *Cognition* 25(1-2). 21–52.

Roca, Iggy – Johnson, Wyn 2005. *A Course in Phonology*. Blackwell, Oxford.

Shriberg, Elizabeth – Stolcke, Andreas – Hakkani-Tür, Dilek – Tür, Gökhan 2000. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication* 32. 127–154.

Szaszák György – Beke András 2012. Statisztikai módszerek alkalmazása a szintaktikai szerkezet és a beszédjel prozódiai szerkezetének feltérképezéséhez olvasott és spontán beszédben. In: Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 236–250.

Váradi Viola – Beke András megjelenőben. Az artikulációs tempó variabilitása felolvasásban.

Varga László 1994. A hanglejtés. In Kiefer Ferenc (szerk.): *Strukturális Magyar Nyelvtan 2., Fonológia*. Akadémiai Kiadó, Budapest. 468–549.

Warner, Natasha – Butler, Lynnika – Arai, Takayuki 2004. Intonation as a speech segmentation cue: Effects of speech style. *9th Conference on Laboratory Phonology*. University of Illinois, Urbana. 37–42.

Publications related to the dissertation

Váradi Viola 2008. A virtuális mondatok műfaji meghatározottsága. *Beszédkutatás 2008*. 134–147.

Váradi Viola 2009. A spontán beszéd mondatai a műfajok és az életkor tekintetében. In Keszler Borbála – Tátrai Szilárd (szerk.): *Diskurzus a grammatikában – grammatika a diskurzusban*. Tinta Könyvkiadó, Budapest. 65–73.

Váradi Viola 2009. Határjelzés a felolvasásban és a spontán beszédben. In Gecső Tamás – Sárdi Csilla (szerk.): *A kommunikáció nyelvészeti aspektusai*. Tinta Könyvkiadó, Budapest. 278–283.

Váradi Viola 2012. A felolvasás és a spontán beszéd tagolhatósága prozódiai egységek jellemzése alapján. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 108–120.

Váradi Viola 2012. Bekezdések a spontán beszédben. In Markó Alexandra (szerk.): *Az anyanyelv-elsajátítástól a zöngekezdési időig*. ELTE és MTA Nyelvtudományi Intézete, Budapest. 304–315.

Váradi Viola – Beke András megjelenőben. Az artikulációs tempó variabilitása felolvasásban.

Presentations related to the dissertation

A spontán beszéd mondatai a műfajok tekintetében, Új nézőpontok a magyar nyelv leírásában, Budapest. 2008. november

Határjelzés a felolvasásban és a spontán beszédben, A kommunikáció nyelvészeti aspektusai, Budapest. 2008. november

A gondolatdarabolás működése, XIII. Balatonalmádi Pszicholingvisztikai Nyári Egyetem, Balatonalmádi. 2011. május

Virtuális határok a spontán beszédben. A felolvasás és a spontán beszéd tagolhatósága prozódiai egységek jellemzése alapján, Beszéd, adatbázis, kutatások workshop, Budapest. 2012. szeptember