Boundary markers in spontaneous Hungarian speech

Beke, András – Gósy, Mária – Horváth, Viktória

Research Institute for Linguistics, Hungarian Academy of Sciences
33 Benczúr Street, Budapest, Hungary

{beke.andras,gosy.maria,horvath.viktoria}@nytud.mta.hu

Abstract: The aim of this paper is an objective presentation of temporal features of spontaneous Hungarian narratives, as well as a characterization of separable portions of spontaneous speech. Ten speakers' spontaneous speech materials taken from the BEA Hungarian Spontaneous Speech Database were analyzed in terms of hierarchical units of narratives (durations, speakers' rates of articulation, number of words produced, and the interrelationships of all these). We conclude that (i) the majority of speakers organize their narratives in similar temporal structures, (ii) thematic units can be identified in terms of certain prosodic criteria, (iii) there are statistically valid correlations between factors like the duration of phrases, the word count of phrases, the rate of articulation of phrases, and pausing characteristics, and (iv) these parameters exhibit extensive variability both across and within speakers.

Keywords: articulation tempo, pauses, durations, F0, thematic units, phrases

1 Introduction

Temporal characteristics of spontaneous speech are affected by a number of factors. The aim of the present study is an objective presentation of temporal features of spontaneous narratives including a characterization of the phrases in the narratives. An attempt is made at defining various units of spontaneous narratives and capturing objective acoustic-phonetic properties of boundary marking. We try to identify the factors determining the articulation rate of portions of speech within and across speakers and to find out whether the acoustic-phonetic parameters we analyze make up a characteristic pattern, and if they do, how they can be described.

Klatt [1] listed seven factors that determine the temporal patterns of speech: extralinguistic factors (the speaker's mental or physical state), discourse factors (position within discourse), semantic factors (emphasis and semantic novelty), syntactic factors

adfa, p. 1, 2011.

(phrase-final lengthening), morphological factors (word-final lengthening), phonological and phonetic factors (stress, phonological length distinctions), and physiological factors (segment-internal temporal structure). Additional factors may also play a role, like topic of discourse, speech type, speech situation, speech partner [2]. An analysis of tempo in Dutch interviews confirmed the distinct role of phrase length [3]. Dialect also seems to be a crucial factor, as shown by an analysis of speech rate in 192 speakers of American English from Wisconsin and North Carolina [4]. Similar results emerged from an analysis of 267 hours of spontaneous dialogues produced by Dutch speakers living in the Netherlands and in Belgium [5]. Both of the last-mentioned papers claim, in addition, that men tend to speak faster than women do, and that young speakers' speech rate is faster than that of older speakers. Some data gathered from speakers of (American) English partly contradict this, however: in a spontaneous speech material of nearly two hundred speakers, the speech tempo of forty-year-olds turned out to be the fastest, as opposed to both younger and older groups of speakers [4]. Significant differences were found between the speech rates of neutral spoken texts vs. ones produced in various joyful or sorrowful states of mind [6]. An increase of the speech rate may be caused by the fact that the speaker considers the given portion of the message less important; but it can also be due to some external factor like the behavior of the interlocutor.

The transformation of the speaker's ideas into speech may become slower due to conceptual planning becoming hesitant, construction of the utterance becoming difficult, or lexical selection becoming riddled by competitive lexemes at the given point. In the phrases of spontaneous Italian narratives, the tempo of syllables has been measured, and compared between pre-stress and post-stress positions [7]. The results showed that after phrasal stress, the tempo increased (by some 65%), while in pre-stress positions, such increase was only by 33%. The decrease of speech rate, on the other hand, where it occurred, was 15% in a post-stress position and 40% before the stressed syllable. It can be concluded that the temporal properties of a longer stretch of spontaneous speech are not constant and not independent of other prosodic properties of speech like stress, or intonation [8].

Inter-speaker variation is significant; but large variability can also be found across utterances of one and the same speaker. In spontaneous English conversations, for instance, 33% large changes were attested in speech rate with one of the speakers [9].

Data from perceptual experiments make it probable that speakers tend to employ general features as boundary markers of thematic units (TU) and of phrases, ones that can also be used in decoding. Thematic units are portions of discourse exhibiting coherence of content that are appropriately structured both syntactically and prosodically [10, 11]. In determining phrases within spontaneous narratives or dialogues, on the other hand, primarily rises and falls of speech melody, as well as stress relationships are taken into consideration [12]. So-called idea units (brief coherent spontaneous text segments) are taken to be 2 seconds long on average, corresponding to roughly 6 English words.

It has been claimed that the acoustic-phonetic marking of prosodic boundaries is not universal and that prosodic boundaries do not necessarily coincide with either syntactic or semantic boundaries in Danish spontaneous speech [13]. In addition, pauses

do not inevitably occur at prosodic boundaries and pauses themselves should not be considered to be boundary markers. Perceivable changes of speech melody and rhythm at boundaries seem to provide cues for boundary identification.

Speech tempo also seems to be a factor influencing boundary patterns [14]. The quantification of speech tempo that provides a single value for a spontaneous utter-ance or for a longer spontaneous speech sample seems to be insufficient, irrespective of whether articulation rate is considered in itself or various types of pauses are also taken into account [15]. Speech tempo values are extremely rough indicators of the nature of spontaneous speech and are not suitable to characterize long narratives or to make comparisons across speakers, dialects, languages or even speech situations. An articulation rate value (without pauses) or a speech tempo value including pauses as contributing to the overall rate of spontaneous speech are not informative enough since they do not show the changes within various parts/units of the speech samples. Speakers continuously adjust their speech rate to cognitive and environmental changes. The underlying adaptive processes unfold in time and involve continual changes in speaking tempo. A timekeeper is hypothesized to reflect the temporal structure of articulation events, thereby establishing a frame of reference for the tim-ing of successive motor commands [16].

This paper intends to reveal the internal tempo changes based on segmentation into thematic units and phrases in spontaneous speech. Analysis focuses further on the interactions of the duration of phrases, the word count of phrases, the rate of articulation of phrases, and pausing characteristics. There are three main research questions: (i) how thematic units and phrases can be defined in spontaneous narratives, (ii) what the interrelations are among various acoustic-phonetic cues that define phrases, and (iii) whether there are universal temporal patterns in spontaneous speech or, on the contrary, individual characteristics show totally different temporal structures in the processing of spontaneous utterances.

The findings of the present research will throw new light on temporal properties of spontaneous narratives, on covert processes of speech planning and pinpoint universal and individual characteristics, features characterizing several speakers and single speakers, respectively. We hypothesize that (i) spontaneous narratives can be segmented into units defined by acoustic-phonetic parameters: these are thematic units that are further segmentable into phrases, (ii) phrases exhibit characteristic temporal patterns, and (iii) thematic units are mostly universal but can also be taken to be based on individual peculiarities to some extent.

2 Subjects, material, method

For this study, we used 10 interviews of the BEA Hungarian Spontaneous Speech Database [17] in which the participants talk about their job, family, and hobbies. Five of the speakers are female, and five are male; all of them native speakers of Hungarian from Budapest; aged between 22 and 35.

The total material is 57 minutes long (3–8 minutes per informants), and was annotated in Praat 5.1 [18] at several levels (thematic units and phrases encoded orthographically and in phonetic transcription, and sound-level annotation). In the case of voiced segments, the first period was taken to be the boundary. Using a Praat script, we automatically extracted fundamental frequency (F0) and intensity. (We sampled both at every 200 ms.) The initial criterion of the definition of thematic units (TU) was that the interviewer opened a new topic by each question, that is, the preceding portion of text was a unit semantically, syntactically, and prosodically, as well. The interviewer started a new topic only when the speaker indicated, verbally or in some other manner, that s/he did not want (or could not) say anything more. Within thematic units, we separated phrases by either or both of the following two criteria: (i) an utterance flanked by (silent or filled) pauses on both sides, and/or (ii) a radical change both in fundamental frequency and intensity.

We automatically determined the occurrence and duration of all labeled silent and filled pauses, and of all phrases, and calculated automatically the rate of articulation, defined as the number of segments per total articulation time. The corpus included a total of 7863 words. The informants uttered an average of 177 words per minute. For statistical analyses, we used the SPSS 13.0 program (analysis of variance, correlation analysis).

3 Results

Description of the results will be organized in five subsections of temporal analysis which concern silent and filled pauses, temporal properties of thematic units, and phrases as well as articulation tempo.

3.1 Silent and filled pauses

Our analyses have confirmed that phrases can be reliably defined in terms of pauses. The corpus included 1326 silent pauses, of a mean duration of 510 ms (SD: 405 ms). The shortest pause took 23 ms, and the longest took 3036 ms. The number and durations of pauses found with individual speakers exhibited extensive variability (Fig. 1).

The duration of silent pauses was significantly different across speakers ($F(9,1326) = 17.422$; $p < 0.001$). The number of filled pauses was 260 in the corpus. Their mean duration was 323 ms (SD: 153 ms). The shortest filled pause took 20 ms, and the longest one took 720 ms. Statistical analysis confirmed significant differences across speakers ($F(9,219) = 6.704$; $p < 0.001$), but a post-hoc test showed that the difference was only significant between a single speaker (speaker 4 in Fig. 1) and all the others. Correlation analysis showed that pausing exhibited individual differences across speakers; if the speech of a speaker was characterized by longer silent pauses, s/he also tended to produce longer filled pauses ($R^2 = 0.643$; $p = 0.045$).

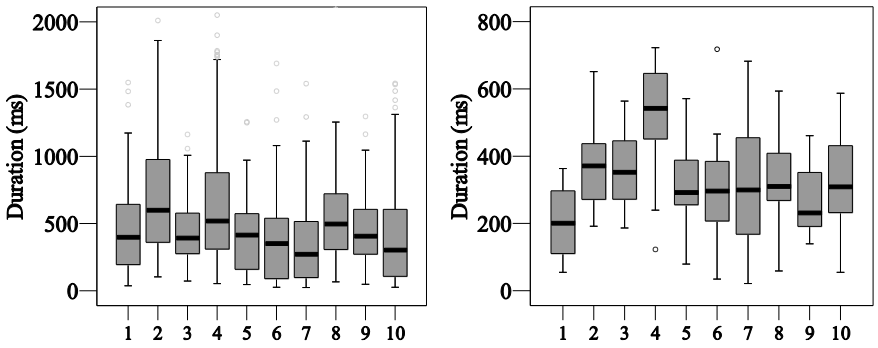

Fig. 1. Duration of silent (left panel) and filled pauses (right panel) (1-5=females, 6-10=males)

3.2 Temporal properties of thematic units

With 60% of the speakers, the narrative could be segmented into three thematic units; the rest of the speakers produced 5 or 6 thematic units. Starting a new topic as the criterion for thematic unit boundaries was correlated with changes in fundamental frequency and intensity; thus, TU boundaries were predictable.

The mean duration of TUs was 56 s (SD: 48 s). The distribution of durations was lognormal (Fig. 2), meaning that most duration figures fell between zero and 100 s, and that the curve decreased in a protracted manner.

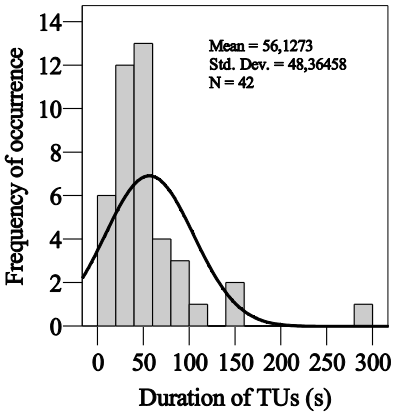

Fig. 2. The distribution of duration of Tus

In the duration of thematic units, with two exceptions, there were no significant differences across speakers (Fig. 3). TU durations of speakers 2 and 3 significantly differed, according to post-hoc tests, from the data of all the other speakers ($F(9,302) = 5.485$; $p < 0.001$). These informants produced far longer thematic units than the others did (Table 1).

The position of TUs within the narratives may have influenced their duration. For an analysis of this, we only considered narratives that contained three thematic units, given that the duration of these units did not exhibit significant differences. The trend was that TUs get shorter as the end of the narrative draws nearer (Fig. 4).

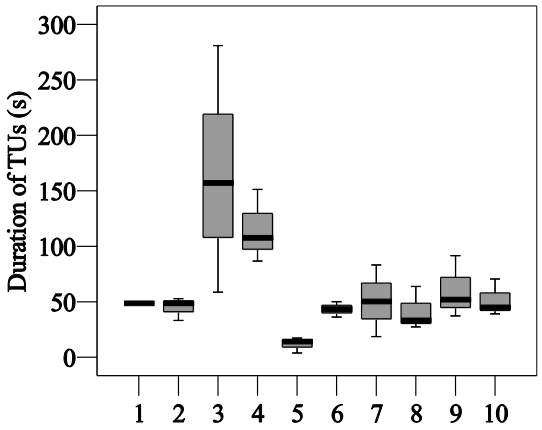

Fig. 3. The duration of TUs in individual speakers' narratives (1-5=females, 6-10=males)

Table 1. Duration of thematic units in individual speakers' narratives (f = female, m = male)

Speakers	Mean (s)	Standard deviation (s)	Minimum (s)	Maximum (s)
1f	44.95	10.40	33.15	52.75
2f	165.67	111.36	58.62	280.89
3f	115.34	32.87	86.71	151.23
4f	24.88	21.67	3.75	76.52
5f	43.35	6.95	36.21	50.11
6m	49.26	23.09	18.63	83.26
7m	43.35	22.03	21.92	70.04
8m	60.43	28.08	37.24	91.64
9m	39.32	15.06	24.55	54.65
10m	52.65	12.04	39.18	70.59

Hungarian speakers produce almost 20 words less in a minute than English speakers do; the relevant figure for English is 196 words per minute [2]. This difference is obviously due to the fact that Hungarian, being an agglutinative language, has longer words (the average syllable count of Hungarian words in spontaneous speech is 3.5). The mean number of words per thematic unit was 245 (SD: 199), irrespective of whether they were content words or function words.

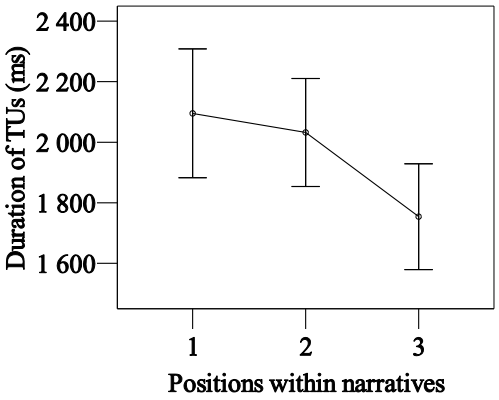

Fig. 4. Duration of TUs in various positions within narratives (1 = initial; 2 = medial; 3 = final)

3.3 Fundamental frequency and intensity of thematic units

F0 changes seem to have a role in the separation of various phrases (and other units) in spontaneous speech. Findings confirmed this separation role using automatic methods [19, 20]. Results of the present study show that F0-values are higher at the beginning of a TU (in the case of about 70% of all speakers) than at the end of a TU (the difference ranges between 6 Hz and 41 Hz), see Table 2. The intensity values revealed similar interrelations: 90% of all speakers produced higher intensity at the beginning of TUs than at their end.

Table 2. Values of F0 at the beginning and end of TUs (f = female, m = male)

Speakers	Thematic units	Mean F0 (Hz)	F0-range (Hz)
1f	beginning	199.3	21.9
	end	159.7	46.5
2f	beginning	191.2	6.7
	end	150.8	55.3
3f	beginning	181.3	13.4
	end	157.4	15.7
4f	beginning	222.8	22.5
	end	186.2	14.8
5f	beginning	191.2	6.7
	end	150.8	55.3
6m	beginning	145.0	32.6
	end	101.7	0.8
7m	beginning	156.9	38.3
	end	138.2	33.0

8m	beginning	124.6	8.8
	end	131.3	48.5
9m	beginning	134.4	12.1
	end	128.5	11.1
10m	beginning	139.3	21.9
	end	114.7	46.5

3.4 Temporal properties of phrases

The number of phrases was 1394 in our material. Their number within TUs was not independent of whether the TU was initial, medial, or final in the narrative. Medial thematic units consisted of fewer phrases than the preceding or following ones (Fig. 5). The duration differences of phrases within thematic units were significant ($F(9,1394) = 11.175$; $p < 0.001$). Their variability was larger across speakers than that of the duration of thematic units. Speakers can be classified into two groups, one group produced relatively short phrases, while the other group produced relatively long ones.

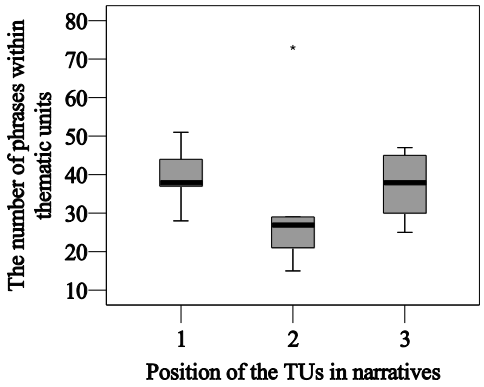

Fig. 5. The number of phrases within thematic units (in six speakers' material)

The position of thematic units within narratives also affected the length of phrases (Fig. 6). Narrative-final TUs were realized in shorter duration than the preceding ones ($F(2,750) = 3.277$; $p = 0.038$).

3.5 Word counts in TUs and in phrases

We established the word count of each TU, irrespective of whether they were content words or function words. The mean number of words per TU was 245 (SD: 199). The smallest number was 147 words/min in a TU, and the largest was 206 words/min. The results show minor differences across TUs of the same speaker; but across speakers, the differences are larger.

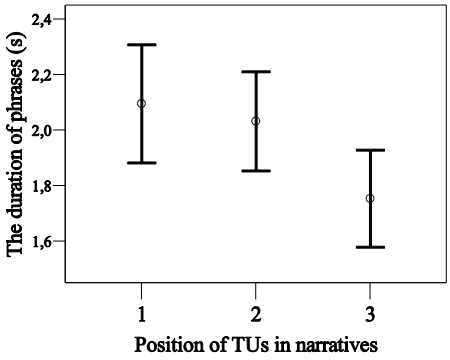

Fig. 6. The duration of phrases in terms of the position of TUs (1= initial; 2=medial; 3=final)

The average word count in phrases within thematic units was 5.8 words (SD: 4.7, minimum: 3.4, maximum: 8.1). The average word count of phrases is lognormal, and exhibited significant differences depending on which TU the given phrase occurred in. The phrases of third thematic units contained fewer words on average than those of first and second ones (1st TU = 6.2 words; 2nd TU = 6.1 words; 3rd TU = 5.1 words; $F(2,750) = 4.313$; $p = 0.014$). That is, towards the end of a narrative, it was not only the case that the thematic units got shorter, but also the phrases they contained were shorter and consisted of fewer words. We found strong linear correlation between the number of words in a phrase and its duration ($R^2 = 0.8603$; $p < 0.001$). This means that the longer the duration of a phrase the more words it consists of (Fig. 7).

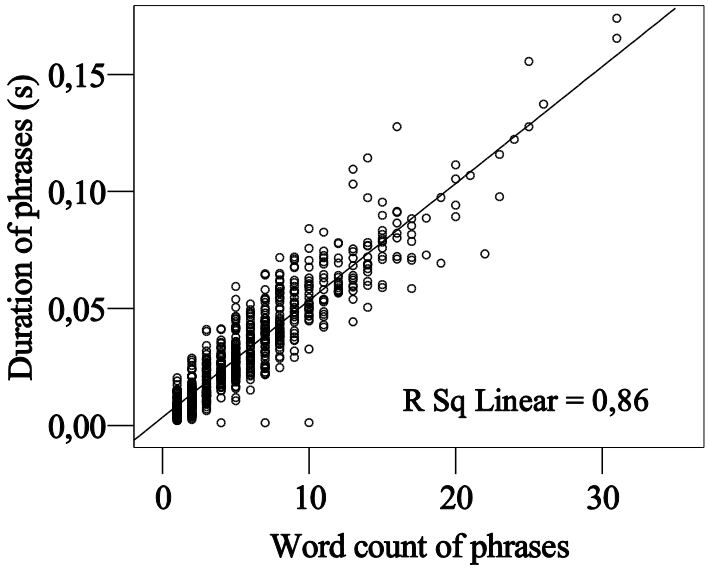

Fig. 7. The correlation between the duration and word count of phrases

3.6 Rate of articulation

The slowest speaker exhibited a mean rate of articulation of 11.7 sounds/s (SD: 3.1), the fastest speaker exhibited 15.4 sounds/s (SD: 6.5). Statistical analyses confirmed significant differences of rate of articulation across speakers ($F(9,1387) = 13.168$; $p < 0.001$). However, a post-hoc test showed that three speakers differed from nearly all other speakers.

Among speakers producing three thematic units, we found two different tendencies in tempo changes across TUs. With three of them, the mean rate of articulation accelerated in the second TU compared to the first, and then got slower toward the end of the narrative. With the other three, on the contrary, the rate of articulation was slower in the second TU than in the first, and then a strong acceleration occurred toward the end of the narrative (Fig. 8).

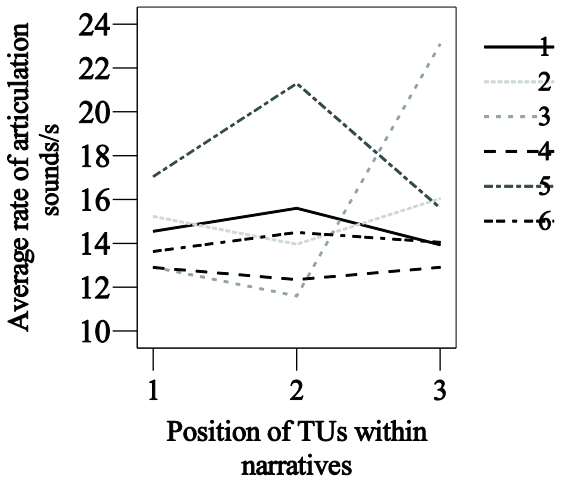

Fig. 8. Average rate of articulation in individual TUs

Given that the rate of articulation changes continuously in the narratives, we performed continuous time analysis of the rate of articulation of phrases. As compared to the mean rate of articulation of the whole narrative, extremely fast and extremely slow values were both found in the individual phrases (Fig. 9).

4 Conclusions

Spontaneous speech corpora make it possible to perform a thorough analysis of temporal properties of spontaneous speech. The mean tempo values can only be a point of departure, followed by detailed analyses of the complex temporal patterns of spontaneous utterances. In the present series of investigations, we determined thematic units and phrases, and gave objective values of the parameters measured. We found that (i) the

majority of speakers (60% in our case) organized their narratives in similar temporal structures, (ii) thematic units could be identified in terms of certain prosodic criteria, (iii) we found statistically valid correlations across factors like the duration of phrases, F0 changes, the word count of phrases, the rate of articulation of phrases, and pausing characteristics, and (iv) these parameters exhibited extensive variability both across and within speakers. The results of the present study speak in favor of the claim that changing temporal structures within spontaneous narratives indicate well segmentable units across speakers.

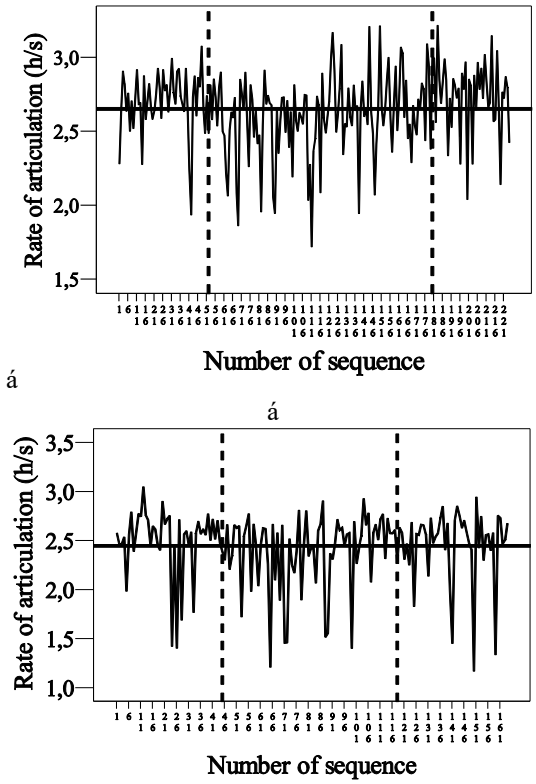

Fig. 9. Rate of articulation in two speakers' narratives (the horizontal line represents the average rate of articulation of the whole narrative; the vertical lines indicate the boundaries of TUs)

According to our data, speakers create TUs of roughly similar duration in their narratives, that is, we can assume the existence of a kind of "internal time control" as part of covert speech planning processes that determines how long speakers may dwell on a given topic in a non-conversational situation. This control function probably takes several factors into consideration, including the listener's assumed level of interest, the amount of information to be shared with the interlocutor, selection, avoidance of certain details, etc. While filled pauses did not differ in length in a statistically relevant manner,

silent pauses did. This can be due to physiological factors like the regulation of breathing, but obviously a number of other factors play a role in how long silent pauses a speaker produces. Pauses, being generally accepted boundary markers, appear to be language specific in both their occurrence and phonetic properties [21, 22]. Narrative-medial TUs tend to consist of fewer phrases than the TUs before and after them. This can be due to the fact that the speaker tends to elaborate the first topic in relatively more detail, requiring more thought and speech planning, a fact that emerges in the production of a higher number of phrases. In the second topic, the speaker employs strategies of narrative construction more easily, speaks more concisely, and produces fewer phrases. In the case of the third topic, however, the speaker appears to lose interest, find solitary speech production inconvenient, or simply get tired, given that in everyday communication the construction of lengthier narratives is not typical.

All those factors may result in the fewer phrases that characterize the last TUs of narratives. The objective temporal data reflect the same pattern. Rate of articulation is expected to exhibit great variability both across and within speakers. The rate of articulation of individual speakers follows two clear tendencies, in which the second thematic unit has a crucial role. But the appearance of extreme values characterizes all phrases.

Our first hypothesis, according to which units defined by acoustic-phonetic parameters can be determined within spontaneous narratives, was confirmed. Thematic units were getting shorter towards the end of the narratives, whereas in terms of the number of words involved, there was no statistically confirmed difference across TUs.

Our second hypothesis was that the phrases making up the thematic units would exhibit particular temporal patterns. This was also confirmed. The duration of phrases showed a lot more variability across speakers than that of thematic units did. It appears, then, that phrases primarily exhibit speaker-dependent properties. Their duration is affected by where exactly they occur within a thematic unit. A strong correlation was found between the number of words in a phrase and its duration, confirming the claim that in longer phrases the speaker indeed produces more words than in shorter ones.

In our third hypothesis, we stated that the properties of thematic units are universal to a larger extent than they are speaker specific. On the basis of our results, this statement has to be qualified. Although the temporal organization of narratives exhibits a number of universal properties, individual properties may override these in interesting ways [23].

Narrative-internal tempo changes may depend on a number of further factors. The present paper demonstrated some objective characteristics of the ways narratives are organized, including properties that are true of speakers in general and those that characterize them individually.

References

1. Klatt, D.: Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. In: Journal of the Acoustical Society of America 59, 1208–1221 (1976)

2. Yuan, J., Liberman, M., Cieri, C.: Towards an integrated understanding of speaking rate in conversation. In: Proc. of the 9th International Conference on Spoken Language Processing. pp. 541–544. Pittsburgh, PA. (2006)
3. Quené, H.: Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In: Proc. of Interspeech 2005. pp. 2457–2460. Lisbon, Portugal (2005)
4. Jacewicz, E., Fox, Allen, R., Lai, W.: Between-speaker and within-speaker variation in speech tempo of American English. Journal of the Acoustical Society of America 128, 839–850 (2010)
5. Verhoeven, J., De Pauw, G., Kloots, H.: Speech rate in a pluricentric language: a comparison between Dutch in Belgium and the Netherlands. Language and Speech 47, 297–308 (2004)
6. Schnoebelen, T.: Variation in speech tempo: Capt. Kirk, Mr. Spock, and all of us in between. In: Proc. of 36th Conference on New Ways of Analyzing Variation: Diversity, Interdisciplinarity, Intersectionality. San Antonio, Texas (2010)
7. Cutugno, F., Savy, R.: Correlation between segmental reduction and prosodic features in spontaneous speech: the role of tempo. In: Proc. of the XIVth International Conference of the Phonetic Sciences. pp. 471–474. San Francisco (1999)
8. Keller, E., Port, R. Speech timing: Approaches to speech rhythm. In: Proc. of the XVIth International Conference of the Phonetic Sciences. pp. 327–329. Saarbrücken (2007)
9. Chafe, W.: Prosody and emotion in a sample of real speech. In: Fries, P. H., Cummings, M., Lockwood, D., Spruiell, D. (eds.) Relations and functions within and around language. pp. 277–315 Continuum, London (2002)
10. Swerts, M., Geluykens, R., Terken, J.: Prosodic correlates of discourse units in spontaneous speech. In: Proc. of the International Conference on Spoken Language Processing. pp. 421–424. Banff (1992)
11. Georgakopolou, A., Goutsos, D.: Discourse analysis: an introduction. Edinburgh University Press, Edinburgh (2004)
12. Botinis, A., Gawronska, B., Katsika, A., Panagopoulou, D.: Prosodic speech production and thematic segmentation. PHONUM 9, 113–116 (2003)
13. Grønnum, N.: A Danish phonetically annotated spontaneous speech corpus (DanPASS). Speech Communication 51, 594–603 (2009)
14. Laver, J.: Principles of phonetics. Cambridge University Press, Cambridge (1994)
15. Jessen, M.: Forensic reference data on articulation rate in German. Science and Justice 47, 50–67 (2007)
16. Schwartze, M., Keller, P. E., Patel, A. D., Kotz, S. A.: The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. Behavioral Brain Research 216, 685–691 (2011)
17. Gósy, M.: BEA - a multifunctional Hungarian spoken language database. The Phonetician 51–62 (2012)
18. Boersma, P., Weenink, D.: Praat: doing phonetics by computer. (2010) (http://www.fon.hum.uva.nl/ praat/download_win.html)
19. Künzel, H. J., Masthoff, H. R., Köster, J. P.: The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition. Science & Justice 35, 291–295 (1995)
20. Sztahó, D., Imre, V., Vicsi, K.: Érzelmek automatikus osztályozása spontán beszédben. In: Tanács A., Vincze, V. (eds.) VII. Magyar Számítógépes Konferencia. pp. 61–274. Szegedi Tudományegyetem, Szeged (2010)
21. Zellner, B.: Pauses and the temporal structure of speech. In: Keller, E. (ed.) Fundamentals of speech synthesis and speech recognition. pp. 41–62. John Wiley, Chichester (1994)

22. Tseng, S-Ch.: Linguistic markings of units in spontaneous Mandarin. In: Huo, Q., Ma, B., Chang, E-S., Li, H. (eds.) Chinese spoken language processes. pp. 43–54. Springer, Singapore (2006)
23. Russo, M., Barry, W. J.: Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo, Proc. Fourth Conference on Speech Prosody, May 6-9 2008, pp. 419–422. Campinas, Brazil (2008)

Acknowledgements

This research was supported by the Hungarian National Scientific Research Fund (OTKA), project No. 108762.