Disfluencies in Spontaneous Narratives and Conversations in Hungarian

András Beke, Mária Gósy, Viktória Horváth, Dorottya Gyarmathy, Tilda Neuberger

Department of Phonetics, Research Institute for Linguistics, HAS, Hungary

beke.andras@nytud.mta.hu,gosy.maria@nytud.mta.hu,
gyarmathy.dorottya@nytud.mta.hu,horvath.viktoria@nytud.mta.hu,
tilda.neuberger@nytud.mta.hu

Abstract

Disfluencies refer to monitoring and repairing errors in speech production processes that either occur on the surface (like false starts) or do not occur but are indicated by, for example, filled pauses. Although there are various factors that can trigger disfluencies, their effects might not be uniform in various speech styles. Speakers use diverse strategies to overcome their speech planning problems. They apply the strategies that have been found to be successful earlier on. Therefore these strategies may be connected with a specific speech style. Our hypothesis is that frequent disfluencies would show different patterns in the two presently analyzed speech styles. We analyzed eight types of disfluencies produced by 50 Hungarian-speaking young adults in narratives and conversations. Results show that the effect of speech style is more decisive in the occurrence of disfluencies than that of their types are. These findings can be applied as an additional factor in automatic speech style identification.

Keywords: disfluency types, speaker-specific patterns, speech style effects

1. Introduction

Various types of disfluencies – filled pauses, repetitions, fillers, false starts, etc. – received increasing attention in the past decades (e.g., Schriberg 1992, Nooteboom, Clark and Wasow 1998). This direction of research is closely connected with the interest of representatives of speech-related fields in spontaneous speech. Speaking spontaneously means the quasi-parallel (or simultaneous) operations of speech planning and execution (Levelt 1989) that might result in disharmony at various levels of the mechanism. The consequences of such disharmony appear in various types of disfluencies that can be traced in spontaneous speech samples as early as from the age of three (Hudson and Carla 2008). Disfluencies refer to monitoring and repairing real or assumed errors in speech production processes that either occur on the surface (like false starts) or do not occur but are indicated by, for example, filled pauses. Disfluencies that are produced in narratives and conversations might be characteristic of the speech styles themselves (e.g., Shriberg 2001, Bortfeld et al. 2001, Mooshammer et al. 2008). Although there are various factors that can trigger disfluencies, their effects might not be uniform in various speech styles. Speakers use diverse strategies to overcome their speech planning problems. They apply the strategies that have been found to be successful earlier on. The different communication situations, the participants of the conversations and the topic to be discussed may lead to different frequency patterns of disfluencies. Therefore these strategies may be connected with a specific speech style. The aim of this study is to describe the disfluency patterns characteristic of narratives and conversations. The questions raised here are whether (i) it is the speech style or the disfluency type that influences the occurrences of disfluencies to a larger extent, and (ii) the disfluencies speakers produce can be used in automatic classification of the two speech styles. Our hypothesis is that frequent disfluencies would show different patterns in the two analyzed speech styles. Therefore we analyzed the occurrences of disfluencies that turned out to be frequent in our material – filled pauses, word repetitions, restarts, filler words, false starts, false words, and anticipations – and were produced by the same speakers in narratives and conversations.

2. Subjects, material, method

50 Hungarian-speaking young subjects (25 females and 25 males, aged between 20 and 32) were randomly selected from the BEA Spontaneous Speech Database of Hungarian (Gósy 2012). 50 narratives (10 hours) and 50 conversations (7 hours) elicited from the same speakers were used in this study. 'Speaking time' in conversations refers to speech samples that were produced by the same subjects whose narratives were used. The speech material was manually annotated at the segmental level by two of the authors while another two of them double-checked the annotations including marked disfluencies in the speech flow. The occurrence of 7 types of disfluency was measured: filled pauses, word repetitions, restarts, filler words, false starts, false words, and anticipations. For the processing of disfluency phenomena we used software specifically written for this task. 5,336 instances of disfluency were found in the conversations and 6,957 instances were found in the narratives. The occurrences and the types of disfluencies were analyzed across speech styles on the one hand, and the automatic classification of the two spontaneous speech styles were carried out using the Wilcoxon signed-rank test (with Monte Carlo simulation) for statistical examinations, as well as Fisher Linear Discriminant analysis were used for automatic classification of the speech styles.

3. Results

Seven types of disfluency were selected for this study that were the most frequent ones in our speech samples. Four of them belonged to the category that is usually labeled by the term 'hesitations'; they were restarts, repetitions, fillers and filled pauses. The other three types of disfluencies were lexical and sublexical errors: anticipations, false starts and false word activations. Since there is no total agreement upon disfluency types and their names in the literature, we briefly define the types we used.

Filled pauses are produced in the majority of cases by *ö*-like (close to neutral vowel) and *m*-like sounds in Hungarian that can easily be differentiated from meaningful sound sequences (e.g., *volt például* ***öö*** *egy tükör* 'there was, for example, **öö** a mirror'). Filler words were identified when neither the grammatical structure nor the semantic context required the actual word uttered (e.g., *meglátjuk hogy mi történik* ***tehát*** *hogy* 'we will see what happens **well** that'). Unintended repetitions were carefully identified considering also the

semantic context of the neighboring elements (e.g., *én én recepciós állásokra pályáztam* 'I I applied for receptionist jobs').

False starts are those erroneous parts of the speech flow where the speaker interrupts the articulation of the wrong word (irrespective of the source problem) followed by another sound sequence forming the originally intended word (e.g., ***ké** három éve* 'for **t** three years'). Restarts are similar to false starts concerning the interruption of the word that is being articulated. However, as opposed to false starts, the continuation is the articulation of the same word in these cases (e.g., ***képre** képregények* '**com** comics'). False word activations were identified via syntactic and semantic analysis of the utterance they appeared in (e.g., *megtámadták a **várost** vagyis a várat* '/they/ attacked the **town** that is the castle). Anticipations are sublexical (phonological) speech errors where the speech sounds or words are spoken ahead of their time (e.g., ***gon** növényeket is kell gondozni* '**nurs** plants should be nursed').

Data showed that narratives and conversations were characterized by different patterns of both the distribution and the frequency of disfluency types. In the material analyzed, 9.47 instances of disfluencies occurred per minute in narratives and 11.31 instances per minute were found in conversations. The ratio of occurrence of the various types of disfluencies shows large differences both between the two speech styles and within the same speech style. There were four types of disfluencies (filler word, word repetition, restart, and false start) occurring more frequently in conversations while the remaining three (filled pause, false word and anticipation) occurred more frequently in narratives (Fig.1). Statistical results supported that speech style had an effect on the occurrences of disfluencies. However, no such effect was found for the factor of disfluency type, and there was no statistically significant result concerning their interaction, either. Various disfluency types were analyzed in pairs depending on speech style. Results show that significant differences were found in five types (filled pause, word repetition, false word, filler and restart) in their occurrences between narratives and conversations (hesitation: Z=-3.702, p<0.001; repetition: Z=-2.351, p=0.019; false word: Z=-2.749, p=0.006; filler: Z=-3.316, p=0.001; restart: Z=-2.860, p=0.004).

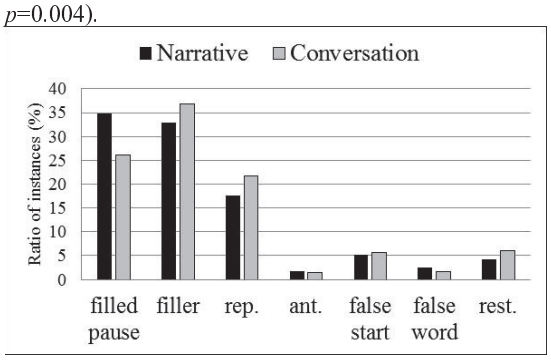


Figure 1: *Distributions of disfluency types depending on speech style (rep. = word repetition, ant. = anticipation, rest. = restart)*

In our material, instances of all disfluency types were found in 18 speakers, 7 types in 23 speakers, 6 types in 7 speakers, and 5 types in 2 speakers in narratives, while instances of all types were found in 15 speakers, 7 types in 17 speakers, 6 types in 13 speakers, and 5 types in 4 speakers and 4 types in 1 speaker in conversations.

Narratives and conversations are characterized by different patterns of distribution of disfluency types. Looking at the details, we found that practically no difference was found in proportions of anticipations, and false starts within each speech style. Filled pauses and false words seem to be more characteristic of narratives while word repetitions, fillers and restarts are characteristic of conversations to a larger extent than they are of narratives. Fillers, filled pauses and word repetitions occupy the largest ratios among disfluency types. Filled pauses indicate a decisive ratio in the pattern of narratives while fillers and word repetitions are more characteristic of the pattern of conversations.

3.1 Occurrences

Although the most frequent disfluency types were selected for this study, their real occurrences show large differences among various types (see Shriberg 2001, Eklund 2004). Table 1 shows the frequency characteristics of the eight types of disfluencies produced by 50 speakers in their spontaneous speech materials irrespective of the speech style. As expected, fillers and filled pauses turned out to be the most frequent disfluencies. Fillers *tehát* ('that is'), *akkor* ('so'), and *hát* ('well') were the most frequent ones in all speakers' speech materials while *ö*-like sounds with various durations were preferred in the case of filled pauses (see also Horváth 2010).

The next most frequent hesitation phenomenon was word repetition used to gain extra time for speech planning processes. Activation of the mental lexicon resulted in two types of speech errors, false starts and false words; however, their occurrences were much rarer than those of the hesitation phenomena mentioned above.

Table 1: *Occurrences of disfluencies in the spontaneous speech material.*

Types of disfluency	Frequency (occurrences/min)	
	mean	std. dev.
anticipation	0.20	0.19
filled pause	3.66	2.78
word repetition	2.24	1.71
false start	0.65	0.48
false word	0.24	0.24
filler	4.41	3.13
restart	0.61	0.50

3.2 Speech style

Disfluencies were analyzed depending on the two speech styles. Results show that there are four types of disfluencies (filler, word repetition, false start, restart and) occurring more frequently in conversations while the remaining three (filled pause, false word and anticipation) occurred more frequently in narratives. Fillers seem to be a better strategy for speakers to overcome their difficulties when they are supposed to speak for a relatively short time. On the contrary, filled pauses were preferred when speakers were forced to speak long in the narratives. Word repetition seems to be a good strategy in conversation to gain time for thinking and formulation of utterances on the one hand, and to signal to participants that the speaker wants to hold the floor, on the other hand. This latter function of word repetition is unnecessary in narratives. Restarts show larger differences between narratives and conversations than do false starts but both are more characteristic of conversation than of narrative. Although false words were not too frequent in our spontaneous speech material, they occurred more frequently in narratives. There

were no large differences in the occurrences of anticipations between narratives and conversations; however, anticipations were more frequent in narratives and less frequent in conversations (Fig. 2).

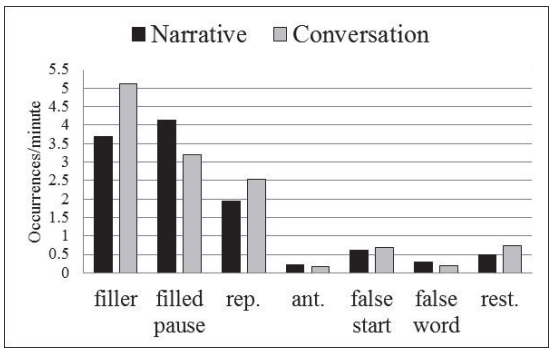


Figure 2: *Occurrences of disfluency types depending on speech style (rep. = word repetition, ant. = anticipation, rest. = restart)*

In narratives, 10.2 occurrences of all hesitation phenomena (std. dev.: 5.02) and 1.46 occurrences of lexical and sublexical errors per minute (std. dev.: 1.71) were found while 11.6 instances of hesitation phenomena (std. dev.: 5.32) and 1.47 instances of lexical and sublexical errors per minute (std. dev.: 2.01) occurred in conversations. Statistical results supported that speech style had an important effect on the occurrence of disfluencies (General Linear Model: $F(1, 799)=339.827$, $p=0.001$). However, no such effect was found for the factor of disfluency type, and there was no statistically significant result concerning their interaction, either. Speech style explains the occurrences of disfluencies in a relatively large proportion, in 63% (according to Partial Eta Squared).

Various disfluency types were analyzed in pairs depending on speech style. Results show that significant differences were found in five types (filled pause, word repetition, false word, filler and restart) in their occurrences between narratives and conversations (Table 2.).

Fisher Linear Discriminant Analysis (FLDA) revealed that there were three types of disfluency that were statistically significant depending on speech style (Table 1). The data suggest that restarts, filler words and false words (respectively) had the greatest discriminant power effect to discriminate speech style.

Table 2: *Statistical results of the FLDA*

Disfluency types	Tests of Equality of Group Means		
	Wilks' Lambda	F	p
anticipation	0.986	1.30	0.255
word repetition	0.971	2.85	0.094
false word	**0.953**	**4.82**	**0.030**
restart	**0.930**	**7.32**	**0.008**
filled pause	0.971	2.89	0.092
false start	0.994	0.58	0.446
filler word	**0.946**	**5.57**	**0.020**

The 3-fold cross-validation method was used to train and test our classifier. Correct automatic classification of narratives and conversations was 60%. To improve the system accuracy we applied dimension reduction using PCA (Principal Component Analysis). The 7-dimension data (number of dimension is the number of disfluency) were decreased by PCA to 3-dimension data, and the FLDA was re-applied. The classification accuracy can be increased from 60% to 64% using PCA. This means that we can achieve 4% relative improvement. The real improvement concerns the narrative classification by an 8% increase of accuracy. The classification result showed further improvement (by 4% both in narratives and in conversations) when the data had been normalized to the means and ranges within each speaker before they were reduced to 3-dimension data and were classified using LDA. The best accuracy yielded by normalized 3-dimensional feature reduced by PCA was 68%.

3.3 Gender and individual differences

Disfluencies produced by females and males were analyzed irrespective of speech style. Males produced more disfluencies per minute than females did (1.67 instances vs. 1.25 instances within a minute). In addition, instances of five types of disfluencies were more frequent in males than in females. Two of them (anticipation and false start) showed no gender differences in their occurrences. However, there was only one type of disfluency, filled pause, where significant difference was found depending on gender ($F(1, 99) = 10.909$, $p = 0.013$), see Fig. 3.

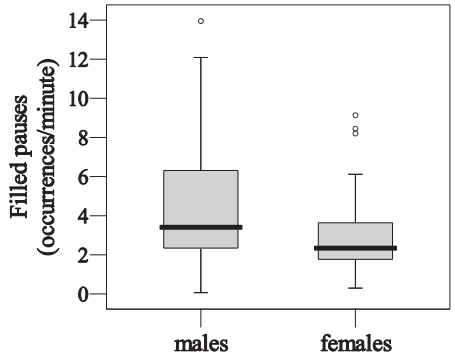
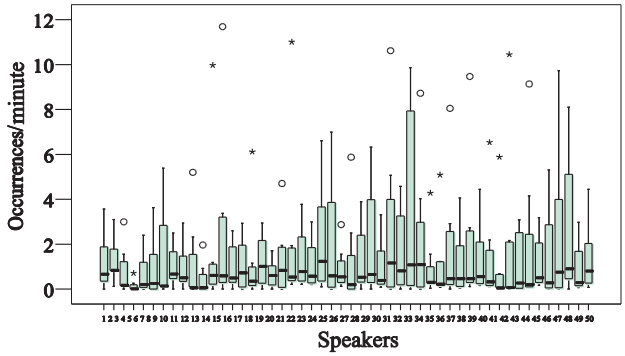


Figure 3: *Occurrences of filled pauses depending on gender*

The occurrence of disfluencies is not uniform across speakers (see Fig. 4); however, groups of speakers sharing similar frequency of certain disfluency types can easily be found. However, no individual speaker-specific disfluency patterns could be found in our large spontaneous speech material of 50 speakers.



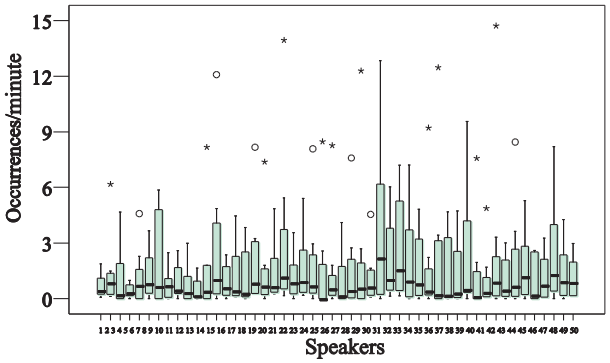


Figure 4: *Speaker-specific disfluency patterns depending on the occurrences of disfluencies in narratives (top) and in conversations (bottom).*

4. Conclusion

Eight of the most frequent disfluency types were analyzed in order to find out whether there is a speech style-specific pattern based on the frequency information of disfluencies. Two speech styles were considered to show their possible effects on the production of disfluencies. As expected, hesitation phenomena surpassed speech errors in both speech styles as had been shown and explained by the monitoring processes of spontaneous speech in a number of previous studies (Levelt 1989; Postma 2000). Fillers, filled pauses and word repetitions occupy the largest ratios among disfluency types. Narratives are characterized particularly by filled pauses while conversations are characterized by both fillers and word repetitions (see Lease et al. 2006). Speech styles significantly influence the occurrences of disfluencies in spontaneous speech which is the consequence of the different speaking tasks in narrative and in conversation (e.g., Shriberg 2005). Gender differences were found in the occurrences of disfluencies. Males used more disfluencies than females did, particularly those indicating uncertainty in their speech (filled pause, filler, word repetition and restart). The occurrence of filled pauses was significantly higher in males than in females. These findings suggest that males seem to be more careful about what to say in order to avoid any misunderstanding on the part of the listener(s).

5. Discussion

Speech style-specific information of certain types of disfluencies seems to be a decisive factor, this information characterizes the analyzed speech style, narratives and conversation. Results of the automatic classification of the speech styles based on disfluency occurrences confirm this statement. We can conclude that (i) speakers' strategies to overcome their speech planning and execution difficulties seem to follow a universal scheme that is reflected by some types of disfluencies but (ii) disfluencies are not characteristic of the speakers themselves.

6. Acknowledgements

This work was supported by OTKA 108762.

7. References

Shriberg, E.E., J. Bear, and J. Dowding (1992). „Automatic detection and correction of repairs in human-computer dialog". Proceeding of DARPA Speech and Natural Language Workshop. Harriman, NY, pp. 419–424,

Nooteboom, Sibout G., (1973). „The tongue slips into patterns". In: Victoria A. Fromkin (ed.), Speech errors as linguistic evidence, 145-156. New York: Academic Press.

Clark, H. H., and T. Wasow, (1998). „Repeating words in spontaneous speech". In: Cognitive Psychology,37(3), pp. 201–242.

Levelt, W. J. M. (1989). "*Speaking: From intention to articulation*." Cambridge, Massachusetts. A Bradford Book, 1989.

Hudson Kam, Carla L. and E. A. Nicole (2008). „The use of uh and um by 3- and 4-year-old native English-speaking children: Not quite right but not completely wrong". In: First Language 28, pp. 313–327.

Shriberg, E. E. (2001). „To "Errrr" is Human: Ecology and Acoustics of Speech Disfluencies". In: Journal of the International Phonetic Association 31(1), Cambridge University Press. pp. 153-169,

Bortfeld, Heather, S. D. Leon, J. E. Bloom, M. F. Schober and S. E. Brennan, 2001. „Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender". In: Language and Speech 44, pp. 123–147.

Mooshammer, C., P. Perrier, and S. Fuchs (2008). "Speaker-specific patterns of token-to-token variability". In: Journal of the Acoustical Society of America, 123, pp. 3076–3076.

Gósy, M. (2012). "BEA - A multifunctional Hungarian spoken language database", The Phonetician, 105/106, pp. 50–61.

Eklund, R., (2004). "*Disfluency in Swedish human–human and human–machine travel booking dialogues*". Linköping Studies in Science and Technology, Unitryck, Linköping, 2004, (dissertation)

Horváth, V., (2010). "Filled pauses in Hungarian: their phonetic form and function". Acta Linguistica Hungarica 57, pp. 288–306, 2010.

Levelt, W. J. M.. (1989). "*Speaking: From intention to articulation*", Cambridge, Massachusetts. Bradford Book.

Postma, A. (2000) "Detection of errors during speech production: a review of speech monitoring models". In: Cognition, 77, pp. 97–131.

Lease, M., Johnson, M. and Charniak, E. (2006). "Recognizing disfluencies in conversational speech". In: IEEE Transactions on Audio, Speech and Language processing 14, pp. 1566–1573.

Shriberg, E., (2005). „Spontaneous speech: How people really talk and why engineers should care". In: Proceedings of Interspeech, pp. 1781–1784.