

Acoustic-Phonetic Analysis of two Words on the Way to Becoming Fillers

Gósy, Mária – Horváth, Viktória

Research Institute for Linguistics, Hungarian Academy of Sciences

E-mail: gosity@nytud.hu – horviki@nytud.hu

Abstract

Two words in present-day Hungarian (a conjunction and a pronoun) seem to be undergoing a functional change, acquiring the function of fillers, in addition to maintaining their lexical function. Seventeen narratives from the BEA Hungarian Spontaneous Speech Corpus were used to analyze the acoustic-phonetic patterns of these words. Significant differences were found in word duration and second formant values depending on function. Our data confirmed that these words are indeed undergoing a functional change, manifested in changes of articulation.

1 Introduction

Spontaneous speech contains various kinds of disfluencies, including the use of fillers. Like all dis-fluencies, fillers can be studied from different angles and with different goals [2, 8]. Generally, fillers are sounds or words that are spoken to fill up gaps in utterances. They are verbal elements which are not generally recognized as purposeful or containing any definite meaning. They are usually indicative of problems in the processes of speech planning or execution. They may suggest that the speaker does not know how to continue, or is unable to recall a particular word. It has been suggested that fillers may also have the function of self-monitoring and may serve as a floor-holder when a speaker needs some additional time to plan the utterance but does not want to cede the floor to the interlocutor, or they might signal mental stress.

Some of the fillers used in spontaneous dialogues are considered discourse markers, i.e. words or phrases that mark a boundary in discourse [4, 5]. If there is no grammatical or semantic explanation for the occurrence of a word in a given context, disfluency analysis will regard it as a filler. Fillers, coming from various word classes, are language specific.

The original function of these words gradually changes as they lose both their primary semantics and their original role in the utterance. People tend

to overuse them for several reasons. Depending on the speech situation, the same word form might have different functions even with the same speaker [5]. Words of high frequency undergo more adjustments and register the effects of sound change more rapidly than low-frequency words [1]. Holistic gestural and acoustic templates are associated with word meanings [6].

In present-day Hungarian there are two words undergoing such development, with an increasing number of speakers actually using them as fillers. One of them is the conjunction *tehát* ('consequently', 'therefore') while the other one is the pronoun *ilyen* ('like this') which can also have determiner function. The filler function of these words is not listed even in the latest dictionary of Hungarian. They do not occur as fillers in the spontaneous speech of the older generations, who use other words as fillers. It may be hypothesized that change in function is reflected in the pronunciation and the acoustic structure of such words, with speakers signaling unconsciously their different role in the utterance. It was this hypothesis that the present study was designed to test.

2 Subjects, materials, methods

Seventeen spontaneous narratives from the BEA Hungarian Spontaneous Speech Corpus were selected for the research. The speakers were native adult speakers of Hungarian (6 females and 11 males) from Budapest (ages ranged from 22 to 30). The recorded narratives, with a total duration of 459 minutes (7.65 hours) were submitted to analysis. They contained 461 words of *tehát* and 154 words of *ilyen*. Two main functions were identified for *tehát*: the filler function (317 tokens) and the conjunction function (144 tokens). The tokens classed as fillers were in turn divided into two groups: those marking simple hesitation (275 tokens) and those regarded as a 'delay device' [7] used to indicate that they want to keep the floor (42 tokens). The conjunction *tehát* may have two meanings, 'consequently' and 'that is', and on this basis the 144 tokens having this function were

subdivided into two groups (40 tokens for the former meaning and 104 for the latter). In a similar way, two functions were identified for *ilyen*, too, that of filler (107 tokens) and that of pronoun (47 tokens). The function was determined for each token on the basis of context.

The digital recordings were submitted to acoustic-phonetic analysis (Praat 4.04) using a 44.1 kHz sampling rate. The total duration of the words and the frequency values of the first two formants of the vowels ([ε, a:, i] were analyzed. The duration of the words for *tehát* was measured as the interval between the closure of the initial voiceless stop and the attenuation of the final stop (Fig. 1), while that of *ilyen* was measured from the onset of the second formant of [i] to the end of the nasal sound. The corresponding spectrographic, intensity and waveform displays were consulted and auditory perception was considered.

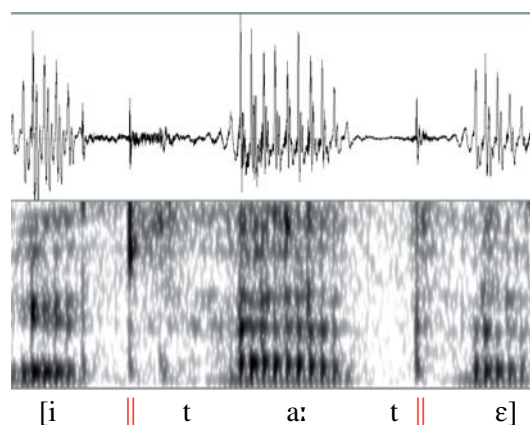


Figure 1: The filler *tát* in “intervocalic” position (oscillogram and spectrogram)

The formant values were measured at the centre of the total vowel duration (using both automatic and manual measurements). To test statistical significance, an Independent Sample Test was used (SPSS, version 8.0). In all cases, the confidence level was set at the conventional 95%.

3 Results

Subjects produced 0.92 words per minute that were analyzed as fillers, 0.69 *tehát* and 0.23 *ilyen*. The mean occurrence of the former was 25.5 words per subject (min.: 7, max.: 93), while for the latter the mean occurrence was 15.5 words per subject (min.: 4, max.: 32).

3.1 Analysis of *tehát*

In the majority of cases, the original phonological form of *tehát* (/teha:t/) appears in

spontaneous speech as a single closed syllable (85.47% of all cases). This form retains the initial and final stops, while the articulation of the vowel in between varies: it may be pronounced as [tat], [ta:t], [tət] or sometimes [tət], and shows some variation across functions and subjects. The (original) bisyllabic form occurs mostly in the conjunction function (19.44% of all the tokens in this function) and also when it is used to keep the floor (28.57% of all such tokens), while it occurs only in 9.8% of all hesitations.

Durational data showed significant differences depending on the function of the words. The articulation of both variants of the word *tehát* ([teha:t] and [tat]) in the hesitation function was completed in a shorter time than the articulation of *tehát* in the conjunction functions ($F(2, 417) = 11.542$, $p < 0.001$). As a filler, *tehát* also occurred at the very beginning of the utterance (similarly to *well* in English), indicating that at the given moment the speaker has not yet committed themselves as to the content of the given utterance and still has a number of choices available, but is definitely signaling that they will speak. Thus, the function of these fillers is not only to express hesitation, but also to signal that the speaker wants to keep the floor. Such fillers will be referred to in this paper as ‘delay devices’.

The duration of the words in this function turned out to be significantly longer than the duration of the same word in the filler function and in the conjunction function ($F(2, 315) = 34.193$, $p < 0.000$ and $F(2, 184) = 11.228$, $p < 0.001$). The tokens of *tehát* in the conjunction function, irrespective of meaning, showed no significant differences in duration (see Fig. 2).

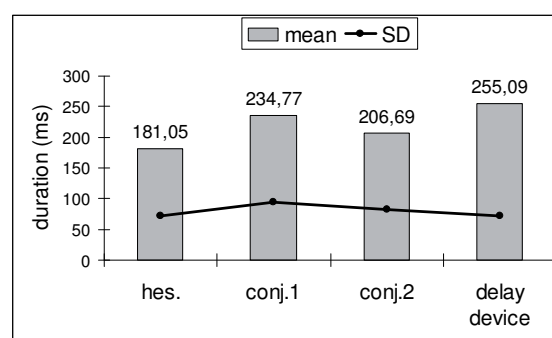


Figure 2: Duration of *tát* across functions (means and standard deviations)

The frequency values for the first formants of the vowels realizing the phoneme /a:/ in the word *tát* showed no significant differences (Table 1). However, it was possible to statistically confirm that there is a difference in the second formants of the

vowel in *tát* between the hesitation function and the conjunction function, on the one hand (for males: $F(2, 305) = 25.214, p < 0.001$; for females: $F(2, 92) = 8.268, p < 0.005$), and on the other hand, between the function of 'delay device' and the conjunction meaning 'that is' (for males: $F(2, 101) = 9.64, p < 0.001$; for females: $F(2, 31) = 5.918, p < 0.021$).

Table 1. *F1-values of [a:] depending on function (conj.=conjunction).*

Functions of <i>tehát</i>	F1 frequency values (Hz)			
	males		females	
	mean	SD	mean	SD
Hesitation	602.4	66.9	684.2	121.7
Conj.1	651.1	68.6	684.2	121.7
Conj.2	623.4	62.3	672.1	112.8
Delay device	617.1	54.6	694.2	132.5

No F2 frequency differences were found between hesitations and the function of 'delay device'. Additionally, no frequency difference in the second formants could be demonstrated in the pronunciation of the word *tát* in the two conjunction function ('consequently' or 'that is'). The actual vowel quality of the /ta:t/ monosyllables is distinguished by the second formant values (see Fig. 3). The vowels in hesitation and in the delay device function tend to be [ə] vowels while those in the conjunction function approach the formant structure of Hungarian vowel [a:].

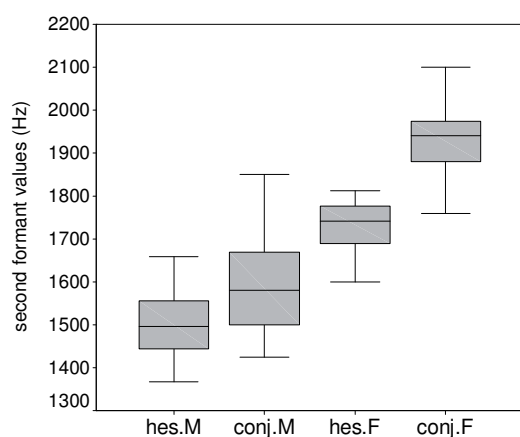


Figure 3: *F2-values for the vowels in tát in various functions (M=male, F=female, hes.=hesitation, conj.=conjunction; median and range)*

Variable articulation of the vowel in *tát* yields a vowel quality which is more reduced but not more peripheral. The reduction is so strong that the vowel [a] loses its original patterns and turns

into the neutral vowel (which, however, is neither a phoneme nor a variant in Hungarian).

3.2 Analysis of *ilyen*

Two different functions were identified for the word *ilyen* ('like this'), that of filler and that of pronoun. The filler function of this word seems to be a new phenomenon in spontaneous speech and occurs only in the speech of the younger generation. The duration of the word *ilyen* ([ijɛn]) shows significant differences depending upon its function (see Fig. 4): it is longer in hesitation ($F(2, 152) = 4.900, p < 0.028$).

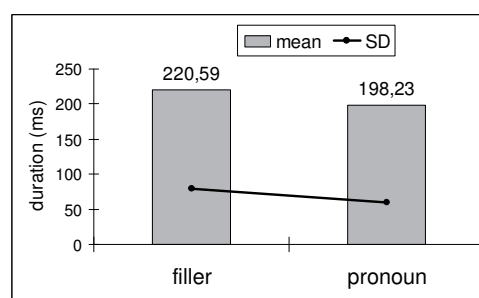


Figure 4: *Duration of the word ilyen across functions (means and standard deviations)*

These data support our assumption that a synchronous change is taking place in the function of *ilyen*, and this functional change is inducing a change in the temporal patterns associated with the word.

The formants of the stressed vowel ([i]) were alike in all cases (with no significant differences; mean F1s in fillers: 375.98/414.58 Hz, in pronouns: 348.35/413.14 Hz and mean F2s in fillers: 2101.75/2298.16 Hz, in pronouns: 2121.44/2303.42 Hz). The females' formants of the unstressed vowel ([ɛ]) showed no significant differences, either. This means that stressed vowel articulation by all subjects and unstressed vowel articulation by females do not reflect the words' actual function. However, the pronunciation of the unstressed vowel [ɛ] by males showed variations depending on function. The altered articulation is confirmed by the significant change in the second formants of the vowel [ɛ] ($F(2, 101) = 4.106, p < 0.045$) while the mean values for their first formants remained unchanged (Table 2).

Figure 5 presents the males' second formant values of [ɛ] in both functions. The unstressed vowel [ɛ] is pronounced closer to neutral [ə] in the filler function while in the pronoun function it

conforms to the characteristic second formant values of the Hungarian vowel [ɛ].

Table 2. *F1-values of [ɛ] depending on function (in the case of males).*

Functions of <i>ilyen</i>	F1 frequency values (Hz)	
	mean	SD
Filler	598.88	54.95
Pronoun	586.29	59.79

The second formant values of the reduced vowel in *tát* hesitations were compared with those of the vowel [ɛ] in the word *ilyen* in its filler function.

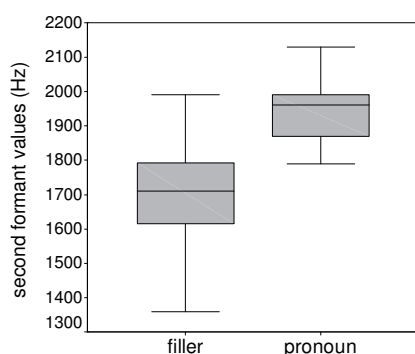


Figure 5: *The F2-values for the unstressed vowels [ɛ] in the word ilyen in its two functions (median and range)*

The reduction of [a:] in *tát* hesitations differs significantly from the reduced vowel quality of [ɛ], which is confirmed by comparison of their second formant values ($F(2, 271) = 26.933, p < 0.001$).

4 Conclusions

The BEA Hungarian Spontaneous Speech Corpus made it possible to conduct a study on a shift from lexical function to filler function in two words. The functional shift was supported by the finding that there was a significant difference in word duration and second formant values depending on function.

Hesitations result in shorter words in the case of *tehát* and longer words in the case of *ilyen* compared to their counterparts with lexical functions. This apparent contradiction can be explained by two factors. (i) The word [tət] (from *tehát*) as hesitation is close to the common hesitation sound in Hungarian which is the neutral vowel (schwa) with a wide range of both its F1 and F2 formants but sounding always an [ə] or [ø]-like sound. (ii) The word *ilyen* is a pronoun and a determiner. When it

is followed by a content word (a noun), most of the speech planning is focused on the latter. In that case articulation of the unstressed pronoun takes a short time. When it takes longer, it signals speech planning trouble.

A similar tendency was found in the articulation of English “the” [3]: normally it is pronounced with a reduced schwa vowel, but when there is a problem in speech planning, it is lengthened to a non reduced vowel (“thee”).

The final question is why speakers seem to need new forms to fill gaps in their speech production process when they have available a number of well-known, conventionally used fillers. Part of the answer may be that schwa, the most common hesitation form in Hungarian, is stigmatized, therefore speakers try to avoid and replace it with a real word that is not so conspicuous.

We do not consider the words studied here discourse markers since their contextual analysis did not identify the function of interaction between the interpretation of the preceding and following segments. Their function is quite different: they help the speaker to cope with problems in speech planning. The above analysis confirmed our hypothesis on function-dependent articulation of the two words under study, leading to the conclusion that acoustic-phonetic analysis may shed light on synchronous changes taking place in a language.

References

- [1] J. Bybee. *Phonology and language use*. Cambridge University Press, New York, 2001.
- [2] H. H. Clark, J. E. Fox Tree. Using *uh* and *um* in spontaneous speaking, *Cognition* 84: 73-111, 2002.
- [3] J. E. Fox Tree, H. H. Clark. Pronouncing “the” as “thee” to signal problems in speaking, *Cognition* 62: 151-167, 1997.
- [4] J. E. Fox Tree, J. C. Schrock. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34: 727-747, 2002.
- [5] A. H. Jucker. The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics* 19: 435-452, 1993.
- [6] J. Pierrehumbert. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, P. Hopper (eds.) *Frequency effects and the emergence of lexical structure*. Benjamins, Amsterdam, 137-157, 2001.
- [7] S. Schachter, N. Christenfeld, B. Ravina, F. Bilous. Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology* 60: 362-367, 1991.
- [8] E. Shriberg. To ‘errrr’ is human: Ecology and acoustics of speech disfluencies. *Journal of the IPA* 31: 153-169, 2001.