Linguistica Silesiana nr 43, 2022
ISSN 0208-4228
DOI: 10.24425/linsi.2022.141220

VALÉRIA KREPSZ
Hungarian Research Centre for Linguistics
Humboldt-Universität zu Berlin
valeria.krepsz@hu-berlin.de

VIKTÓRIA HORVÁTH
Hungarian Research Centre for Linguistics

ÁGNES HÁMORI
Hungarian Research Centre for Linguistics

DOROTTYA GYARMATHY
Hungarian Research Centre for Linguistics

CSILLA ILONA DÉR
Károli Gáspár University of the Reformed Church in Hungary

BACKCHANNEL RESPONSES IN HUNGARIAN CONVERSATIONS: A CORPUS-BASED STUDY ON THE EFFECT OF THE PARTNER'S AGE AND GENDER *

Backchannel responses (BChs) have been an important research topic in conversation analysis since Yngve (1970) drew a distinction between speech produced by the person holding the turn and talking in the 'main' channel, while the listener occupying the 'back' channel of the communication. However, much less is known about the acoustic-phonetic characteristics of BChs.

The aim of the present study is twofold: 1. it seeks to provide a general analysis of BChs as a formal group in 10 Hungarian conversations, using a formal, phonetic approach. 2. Additionally, another sub-study was conducted concerning the contextual variability of BChs investigating their change, according to the age of the conversation partner in three-party conversations.

The age of the subject did not significantly affect the frequency and duration of the fieldworker1's BCh or the other acoustic parameters, however it did make some difference in pause patterns. In contrast, gender-related changes were detected in the

* The research was supported by the Hungarian National Research, Development and Innovation Office of Hungary [projects No. K-128810]. A sincere thank you to Ms. Elizabeth A. Spiteri for her diligent proofreading of this paper.

fieldworker1's utterance: having produced more BChs when talking to older women. This may be explained by the possible intimacy in female conversations as well as by the influence of politeness. The results might offer new insights of the conversation's structure with regard to their age and gender.

Keywords: backchannel responses, conversation, corpus linguistics, conversational partner's age and gender

1. Introduction

Backchannel responses have been considered an important phenomenon in conversations, and research has developed into a notable area since the end of the 1960s. The terminology is not consistent, but several terms are used for the phenomenon, such as *listener responses* (Deng 2009), *continuers* (Schegloff 1982), *backchannel* (Cutrone 2005), *backchannel responses* (Li et al. 2010), *feedback utterances* (Prévot et. al. 2016). We define backchannel response/ feedback/signal in this study as Ward and Tsukahara (2000: 1177): "the short utterances produced by one participant in a conversation while the other is talking". In other words, as the verbal and nonverbal feedback provided by the listener, without taking the floor, in order to signal attention, understanding or encouragement, for example *ühüm* 'hum', *aha* 'uhhuh', *igen* 'yes', *az tök jó* 'that's cool'. In terms of their form, backchannels are generally short because they are not planned to interfere with, distract, or interrupt the main channel's speaker.

Studies have shown that apart from signalling attention, agreement or understanding, backchannels are used for example expressing judgment, sympathy, but they can also express opposition, boredom, or scepticism (Ward and Tsukahara 2000). Gravano and his colleagues (2007) showed that context plays a central role in identifying the current function of a backchannel and that a single word such as *okay* can perform multiple functions at the same time. Other studies have distinguished different types of BCh responses based on the type of comments they express.

The frequency of BChs can vary along a number of factors, such as the number of participants, the topic, the individual characteristics, and the culture of the speakers, as well as the language. Jurafsky and his colleagues (1997) analyzed 1155 American English conversations, and found that 19% of the utterances were realized as backchannels. According to Maynard (1989), BChs appeared even more frequently in Japanese. It seems that the frequency of backchannel is a culture-dependent, social interaction behaviour.

Due to the many influencing factors, their phonetic realization can often be varied. The main types of BChs, with regard to their formal features, can be considered as prototype categories with graded membership and fuzzy boundary. This means that there are no clear-cut borders between the types and that the neighboring categories may share common parts, and a certain type might have prototypical and peripheral representatives as well (Croft and Cruse 2004). The different types and realizations of backchannel responses can be arranged on a continuum based on their duration, phonetic features (articulation, vocal/verbal character), and grammatical structure. One end of the continuum is represented by non-lexical, non-verbal responses, such as laughter, sniffing, sighing. At the other end of the continuum, there are verbal and lexical responses, which might be simple or complex in terms of their structure (a similar categorization is provided by Wong and Kruger 2018). Simple backchannel responses are made up of a single word, such as *igen* 'yes' or *persze* 'of course'.

Complex backchannel responses are made up of more than one lexical (or semi-verbal) unit, often forming a separate sentence, as in *ja, értem* 'oh, I see' and *ez így van* 'that is right'. The wide space between the two extremes of backchannel responses is occupied by semi-verbal items that display some characteristics of both nonverbal and verbal backchannel responses. The major representatives of this category are the various realizations of humming, which is the most frequent type of backchannel responses. Besides humming, the semi-verbal category includes interjections such as *hú* 'whoa', *aha* 'uhhuh', *ú* 'wow', *ha* 'ha', *jaj* 'ouch', *hajjaj* 'ah, well' etc., having a half lexical-half nonverbal character. In combined cases the lexical element appears together with a nonverbal element, e.g. humming + *igen* 'yes'. Those cases in which a semi-verbal and a lexical element are combined. Backchannel responses vary according to their duration, verbality, and to their grammatical, phonetic and semantic elaboration along this continuum.

Backchannels have been investigated mainly from pragmatic aspects, however, less is known about the phonetic characteristics of the backchannels themselves or their context. According to previous literature results, the occurrence of BCh can be predicted by several acoustic-phonetic parameters. These factors are called backchannel-inviting cues (Ward and Tsukahara 2000). Ward and Tsukahara define a region that can serve as a backchannel-inviting cue if implemented for at least 110 ms and at a low pitch. This is because low pitch often appears at the phrase boundary or after a disfluency. In another study (Gravano and Hirschberg 2009), pitch, intensity, jitter, shimmer, and noise-to-harmonics ratio (HNR) values were examined as backchannel-inviting cues, based on the Columbia Games Corpus. They found that the more assumed backchannel-inviting cues occur simultaneously, the more likely they were to appear in the backchannel.

Cathcart et al. (2003) have proposed a language model to predict the location of backchannels based on pause duration and speech tags. Other research has developed a five-point prediction system for English that includes parameters like coming after at least 700 milliseconds of speech, providing you have not output back-channel feedback within the preceding 800 milliseconds, although the significance of backchannels in conversations is a universal feature, they appear in communications with different frequencies and forms. This is also evidenced by the fact that in the case of Japan, the "predictive" characteristics also differed from the English.

Based on an examination of Ward and Tsukahara's (2000) two-person conversations, it can be concluded that other factors may contribute to the appearance of backchannels. A backchannel may appear after an irregular tone, or more likely to occur after certain lexical elements. Backchannel responses often occurred following vowel lengthening as well (Maynard 1989). However, these elements usually do not appear alone, but together as a pattern. For instance, lexical elements are usually realized together with a given F0 trend. It is paramount to recognize that the occurrence of BChs is not necessarily solely controlled by the listener, but is often cued, encouraged, or allowed by the speaker in both Japanese and English conversations. The prosody of *okay* and *uhhuh* backchannel responses was investigated in a relatively small corpus in English. Data showed that the pitch contour showed a rising pattern on their second syllable in BCh function (Hockey 1993). The frequency and the prosody of English affirmative words were analyzed as well as the prosody of the BChs's context in a 9-hour-long material, regarding their function (Benus et al. 2007). Results showed that despite their high lexical variability, backchannels were prosodically well-defined. Affirmative words in backchannel function were realized with higher pitch and intensity as well as with greater pitch slope than those expressing other pragmatic functions. The duration of the affirmative words was found to be longer in backchanneling function than in other functions. *Okay* had a lower intensity compared to other types. Intonational phrases with lower mean pitch and greater pitch slope were followed mostly by *okays*, while intonational phrases, realized with greater intensity, and were followed by *uhhuh*s in general. Interrelation was found in the durations as well: the duration of the preceding intonational phrase and the backchanneling speaker's latency positively correlated with backchannel duration. Data also showed some gender-related differences: female speakers produced a BCh response more often when speaking to other female speakers than when speaking to males. Furthermore, females continued speaking at the main channel more often following a BCh from a female addressee than from a male.

In conversation, the form (duration, lexicality, etc.) varies according to the actual function of backchanneling. The use and realization of BChs can be influenced by many factors, for example by the environment (e.g. noise circumstances), the current physical and mental state of the speaker, biological and sociolinguistic factors (e.g. age, gender) and the communicative and social goals of the speaker as well as the conversation partner(s). For example, according to many studies, women do more agreeing and showing of support, in both same- and mixed-sex interactions (e.g. Carli 1989). The majority of studies which examined the use of backchannel responses found that women use more of them during conversations (e.g., Edelsky and Adams 1990). Furthermore, both men and women produced more backchannel responses in mixed conversations (with a member of the opposite sex) than they did in single conversations (with a member of their same sex cf. Feke 2003).

The speech and verbal behaviour of the speaker are inseparable from the social context of the speech situation and from the other participants. It is especially important in conversations with older people, where both speech accommodation (usually in the form of elderspeak) (Coupland et al. 1988; Kemper et al. 1995) and politeness phenomena often play a major role. As evidence for elderspeak, earlier research found that young speakers when talking to older people had modified the speech rate, the grammatical complexity, the semantic content, and discourse style (Kemper et al. 1995).

However, there are few research results on the effect of age or of the partner's age on the speakers' BChs responses. Some studies (Gould and Dixon 1993) have found that younger adults produce more BChs than older adults in various circumstances, explicitly indicating the active monitoring of the partner's production. Other results show that the total duration of the conversation and the ratio of overlapping speech were influenced by the age of the experimental person (Bata 2009).

In the present paper, two pieces of sub-research were conducted to identify the backchannel responses in Hungarian triadic conversations, based on two sub-corpora of the Hungarian Spoken Language Database (Neuberger et al. 2014a). In the first research, the general features of backchannel responses were analyzed, concerning the main aspects of phonetic features, form (lexical structure), and use (position and frequency), aiming to explore their main types in the conversations. In the second research, the effect of the subject's identity on the fieldworker's BChs was analyzed. The uses of backchannels as possible markers of elderspeak as well as some additional phonetic patterns of speech were analyzed. Furthermore, the variability of backchannel responses within the conversations was analyzed in both studies: first, changes of BChs in terms of the speaker's role, and, second, changes regarding the position of the BCh in the overall structure of the conversation. Moreover, two rarely explored aspects were investigated: the connection between backchannel responses and the interactional role of the speaker, as well as the change of the backchannels observed in the participants' speech (aspects of age and gender).

It is hypothesized that the usage of BChs varies according to the speaker, in and across conversations depending on the global structure, and even on the partner's age.

The importance of our research is justified by several factors. Firstly, this is the first systematic corpus-based study using Hungarian conversations, and the results may contribute to identify the backchannels used in the Hungarian language. Secondly, regarding Hungarian backchannels, only a few, relatively narrow-scope research projects can be listed in connection with humming and laughter (Markó et al. 2014; Neuberger et al. 2014b), and those are only loosely connected to the pragmatic and interactional theories of verbal interaction. Overall, no comprehensive or complex research has been conducted in this field to date. Data on internal variability may help us understand the dynamics of backchanneling, including their relationship with socio-interpersonal factors and uncover aspects previously not studied in Hungarian. Finally, a unique feature of our research that no previous research has addressed, is to have explored the effects of the speech partner's personality on backchanneling behavior. The results provide a basis for future investigation into the social-interpersonal functions of backchannels.

2. On the Conversational Corpus of the Hungarian Spontaneous Speech Database

The Hungarian Spontaneous Speech Database was developed at the Phonetic Department of the Research Institute for Linguistics (Neuberger et al. 2014a). The database contains approximately 500 records. Each recording contains a total of 8 subtasks, e.g. sentence repetition, text reading, interviews, and conversation. The recordings are conducted in a silent room, using an AT4040 microphone. Recording is made digitally, direct to the computer, with GoldWave sound editing software, with sampling at 44.1 kHz (storage: 16 bits, 86 kbytes/s, mono). The subjects are all monolingual adults from Budapest, not one of them reported any hearing disorders, their ages range between 20 and 90 years. The recordings are anonymized (the speakers are given codes). For each recording, the following data are documented: the subject's age, schooling, job, stature (height), weight, whether s/he is a smoker, and the topics of the spontaneous speech modules.

The protocol of the conversational part is the following. The conversation part is the sixth subtask of the recording. The fieldworker1 (FW1) and the subject (S) participated in the previous parts, then the fieldworker2 (a colleague of the Fw1) joined the others. The three-participated conversations are quasi-spontaneous in the sense of the speech planning: the topic of conversation had been added at the beginning of the recording without any pre-fixed protocol or script to be followed. The participants had no time to plan the speech production, and the outcome depended on the opinion and utterances of others. Topics are selected by the fieldworker1 in accordance with the subject's age, job, and area of interest (based on the previous narrative parts of the recordings), e.g. New Year's Eve, wedding experiences, job hunt, Easter and Christmas holiday, school violence, keeping pets in a flat, bringing up children, cycling as a form of traffic etc.

The conversational sub-corpus of the BEA is developed with the funding of the Hungarian National Research, Development and Innovation Office of Hungary [projects No. K-128810]. This conversational corpus contains 130 conversations from the original database, developing the original annotation with regard to conversational phenomena. The existing six-tier annotation (IPU (inter-pausal units) + pauses (minimal duration of the silent pauses was 30 ms), word level for 3 speakers) was complemented with further tiers by the software Praat (Boersma and Weenink 2018): backchannel responses; overlapping speech; turns; turn-takings; whether a given turn belongs to a narrative or a dialogic (differentiated by text structure organization, cf. Tolcsvai 2001; Hutchby and Wooffitt 2006) section inside the global structure (see *Figure 1*). Like dialogic sections, each narrative section consisted of more turns. In the case of narrative sections (e.g. storytelling), the current speaker keeps the floor longer than the basic rules of turn-takings ordinary allow (Sacks et al. 1974; Norris 2000; Hutchby and Wooffitt 2006). In these cases, the current speaker speaks a little or no interruption from the other participants. We decided to mark a section to be narrative when one of the participants created a longer narrative section by telling a story, describing something or setting out an opinion; while the others only produced backchannels or short comments during the current speaker's narrative section. In addition, one conversational storytelling often leads to a response story, fitted to the topic of the preceding story by one of the other participants (cf. Norris 2000). For instance, the fieldworker1 created a narrative section on the topic, sharing her opinion or experience about it, then she asked the subject's opinion. Then the subject created another narrative on his/her thoughts – the fieldworker2 may create her own one responding to the previous narratives. As opposed, dialogic sections were characterized by more frequent turn-takings with no fixed order or turn length, more overlaps.

3. Experiment 1

Our study aims to offer a multi-dimensional analysis of backchannels in triadic spontaneous conversations, undertaken for the first time in Hungarian. The main aspects of the analysis included the types of realization and the frequency of Hungarian backchannels while, on the other hand, the variability of their usage across speakers and conversations were also analyzed. The dynamic changes of BChs were investigated in terms of the structure of the conversations and the speaker's role. Our main questions were: i) what kind of phonetic patterns can be discovered in various types of BChs in spontaneous conversations conducted in Hungarian; ii) which factors influence the realization (type, frequency, duration) of BChs (the speaker's role, the global structure of the conversation, etc.)?

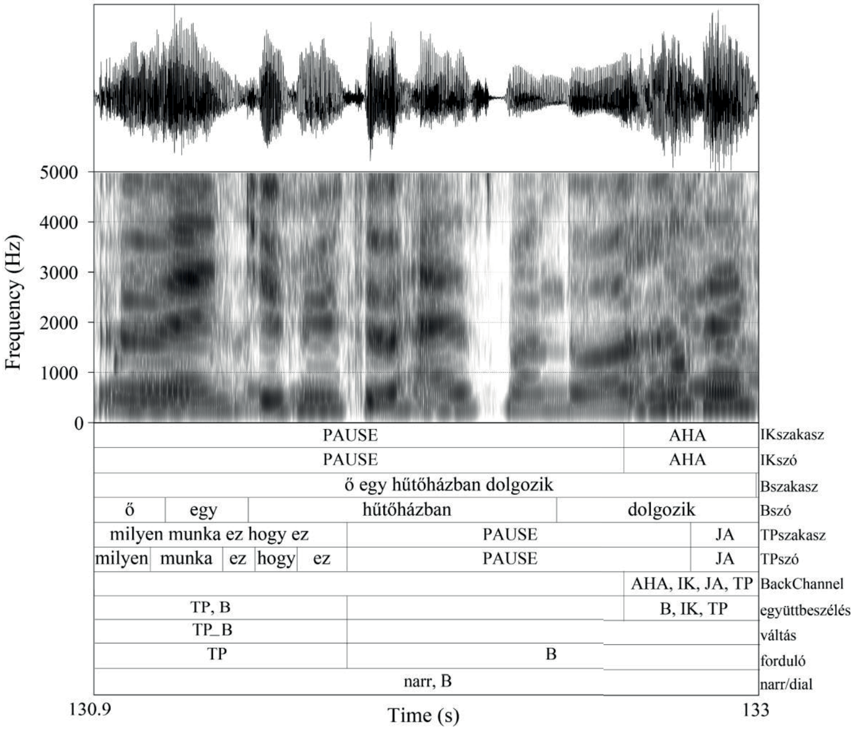

Figure 1. Example annotation (to make the figure more comprehensible, the word level was omitted; Meaning of the inscriptions: 'IK szakasz' = the Fw1's IPUs + silent pauses, 'B szakasz' = S's IPUs + silent pauses, 'TP szakasz' = Fw2's IPUs + silent pauses, 'együttbeszélés' = overlapping speech, 'váltás' = turn transition, 'forduló' = turn, 'narr/dial' = narrative or dialogic speech section; 'AHA' = humming, 'ő egy hűtőházban dolgozik' = he works in a refrigeratory, 'milyen munka ez hogy' = what kind of work is that?,'JA' = okay)

3.1. Data, methods and participants

10 conversations, each containing 3 participants were analyzed from the Hungarian BEA corpus (Neuberger et al. 2014a). The whole dataset is 131 minutes long (mean duration: 13 min, range: 6.9–23.3 minutes).

The two fieldworkers (Fw1 and Fw2) both were the same persons on all recordings: two monolingual, Hungarian female colleagues of the same age, both from Budapest and awarded with the same scholarly degree. They were 28 when the recordings started and are now 35 (the recordings and building the database took almost 10 years to complete). The Subjects (S) were different people in each conversation. For our study, 5 male and 5 female monolingual Hungarians from Budapest were selected, aged 20–35. The participants talked about an everyday topic, prepared by the fieldworker1 in advance. The recordings were not pre-constructed, quasi-spontaneous ones, but the fieldworker1 tried to maintain spontaneous conversations with the participants for as long as possible.

In total, 822 backchannel signals were annotated and the annotations were based on a functional approach of backchannel signals. The annotated backchannels were analyzed according to structure, type, duration and frequency. Each backchannel signal was analyzed by four linguists: two specialized in pragmatics and two phoneticians. A total of 4% of the material had to be omitted from the analysis owing to methodological problems (e.g. it was impossible to decide what was said). The duration of BChs was automatically obtained by a Praat script written by one of the authors.

The analysis was based on a formal quantitative approach (Wong and Kruger 2018), investigating phonetic and structural features of BChs. Four main categories were created, serving a basis for the phonetic and the statistical analysis. In order to examine the frequency of the occurrence of backchannel responses concerning the global structure, the sections of conversations were marked according to whether they were narrative or dialogic parts (both of them consisted of more turns), involving a text-linguistic perspective and combining it with the phonetic analysis. Each narrative and conversational section was then cut up into 5 equally long subsections based on their duration using a Praat script. As the duration of sections differed, the 20-percent long subsections were of equal duration only in a given section, and their duration varied in different sections. For example, a 200-second-long section was cut up into five 40-second long subsections, or a 400-second-long section was segmented into five 80-second long subsections. During this temporal segmentation process neither grammatical nor structural information were considered; it was done automatically. Next, the number of backchannel responses was measured for each subsection. We also noted the type of the section (narrative or dialogic) and who was speaking (Fw1–fieldworker1, S–Subject or Fw2–fieldworker2).

Statistical analyses were carried out by R software (R Core 2018), and a multilevel general linear model was built on our data. Duration and frequency of occurrence were the dependent variables, while the categories of backchannel responses were the independent variables (according to elaboration and structure). The random factor was the subject. The distribution of data was checked with the χ^2 goodness of fit test.

3.2. Results

Backchannel responses were extremely versatile in our corpus. The most frequent one was *ühüm* ‘hum’, which made up 34.4% of all occurrences; followed by laughter (24.1%); while the third place is shared by *igen* ‘yes’

(11.7%) and *aha* ‘uhhuh’ (10.5%). These made up a little more than 80% of all occurrences. All other backchannel responses were rather rare, each occurring in less than 2% of instances in our corpus. These included e.g. *ja*, *persze* ‘ah, of course’, *hűha* ‘wow/blimey’, *azta* ‘wow’, *hát igen* ‘well yes’, *persze* ‘of course’.

During further analysis, backchannel responses were categorized according to different points of view: phonetic form and grammatical-semantic structure. According to phonetic features, we distinguished between four major categories. (1) **Nonverbal vocal signals** included sighs, sniffing, clearing the throat, coughs, etc. (2) **Semi-verbal signals** are backchannel responses that are half-way between nonverbal and vocal signals, such as humming and several versions of *aha* ‘uhhuh’, *ah* ‘ah’. It must be noted that these signals are elaborated on to a different extent. Categorization was based on the perception of transcriptionists and researchers. However, articulatory or acoustic analysis should also be carried out in the future. (3) **Verbal signals** are made up of lexical words, such as *persze* ‘of course’, *semmi* ‘nothing’, *szerintem is* ‘I agree’, *szörnyű* ‘terrible’, *nem is kell* ‘you don’t need to’. (4) **Complex signals** included a combination of the previous three categories (e.g. laughter + *igen* ‘yes’; *ühüm* ‘hum’ *az tök jó* 'that's cool'). Most backchannel responses belonged to the semi-verbal category (46%). In this category, the most frequently occurring signal was *ühüm* ‘hum’ (74%) and *aha* ‘uhhuh’ (22%). The proportion of the verbal (26%) and nonverbal (25%) responses proved to be almost the same in the corpus. The most frequent verbal backchannel response was *igen* ‘yes’ (44%), while laughter (97%) was by far the most frequent nonverbal feedback. The distribution of data was statistically analyzed, which proved that the distribution of signals is not random: χ^2 (3) = 307.411, $p < 0.001$.

We distinguished between **one-member, multi-member** and **multi-member repetitive backchannel responses.** Most of the backchannel responses in the corpus had only one member (87.7%), such as *ühüm* ‘hum’, *ja* ‘ah’, *persze* ‘of course’, *yes* ‘igen’. Multi-member responses made up 10.8% of tokens (e.g. *hát igen* ‘well yes’, *nem is kell* ‘you don’t need to’. Only 1.5% of backchannel responses were repetitive (e.g. *igen igen* ‘yes yes’, *ja ja* ‘ah ah’).

The next step was to examine the frequency of the occurrence of backchannel responses, which proved to be rather diverse in different conversations and with various S speakers (see *Figure 2*). On average, 27 backchannel responses were produced by each speaker per conversation. Not surprisingly, most of these signals were found in the IPUs of the fieldworker1 due to her role in the conversation; on average 45 signals per conversation (ranging from 33 to 76%). The fewest backchannel responses were uttered by the Subject (S), (mean: 16 per conversation, ranging from 5 to 28%). The fieldworker2 produced 21 backchannel signals on average (ranging from 6 to 47%). The frequency distribution of backchannel responses among the 3 speakers presented great variance in different conversations. For example, in Conversation 1, the three participants produced similar numbers of backchannel responses, but more typically, the fieldworker1 uttered most of these signals (e.g. in Conversations 6, 8, 9; see *Figure 2*).

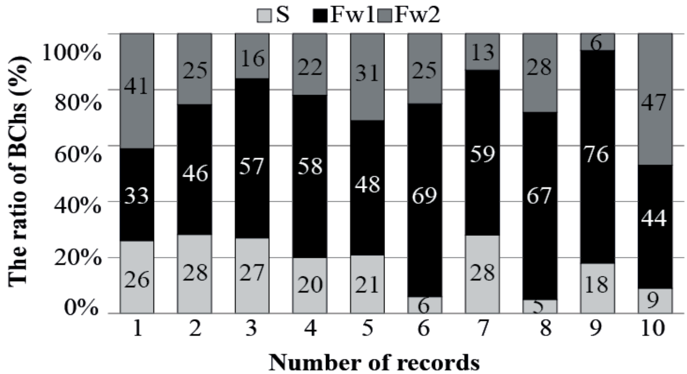

Figure 2. Number of backchannel responses produced by each speaker in the conversations

How often backchannel responses occur does not only depend on whom the speaker is, but it is also influenced by the structure and the dynamical change of the conversation. To explore this factor, we identified narrative and dialogic sections in the recordings and cut them up into 5 equally long subsections. The number of backchannel responses was calculated for each subsection and examined to see whether there was some correlation between the section type (narrative vs. dialogic) and the person who was speaking (Fw1 or Fw2). Irrespective of the section type or the speaker (Fw1, Fw2 or one of the 10 speakers), the frequency of the occurrence of backchannel responses showed the following tendency (see *Figure 3*). Most backchannel signals were recorded in the fourth subsection (i.e. between 61% and 80% of the full duration of a given section), while the second most frequent occurrence time was in the closing subsection. In the first three subsections, the backchannel responses occurred infrequently (on average, 18% of responses occurred in each section, which equals 156). As the statistical test did not show significant differences between the subsections, they only marked a tendency.

Conversely, the frequency of occurrence in backchannel signals differed significantly in narrative and dialogic sections (see *Figure 4*). The number of backchannel responses given in the narrative sections (114 on average) was more than twice as many as the number of those in the dialogic parts (55 on average). The distribution within a given section, however, followed the same tendency: most BChs occurred in subsection 4 (23% of signals on average), a little less in the last subsection (21% on average), and the first three subsections contained 18–19% of backchannel signals each. The difference proved to be significant for the section type (narrative vs. dialogic): $F(847, 9) = 87.351$, $p < 0.001$.

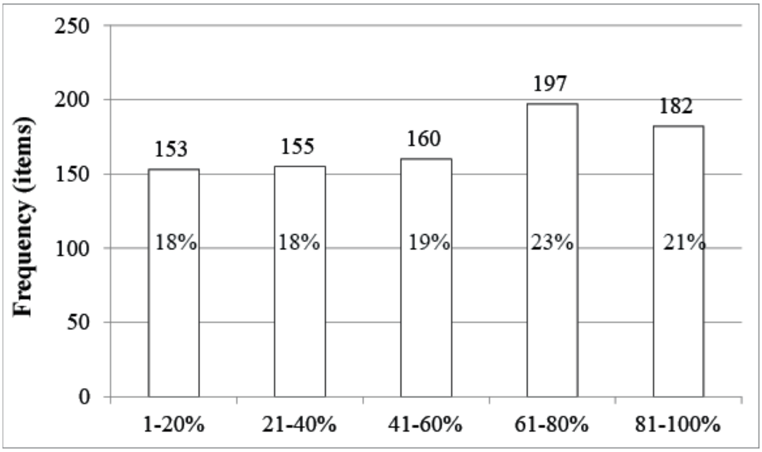

Figure 3. Frequency of occurrence of backchannel responses in 20% subsections (number of items at the top of the columns and in percentage in the middle of the columns)

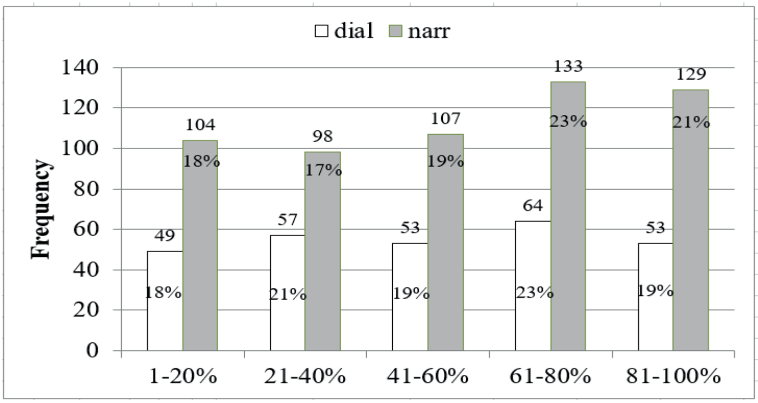

Figure 4. The occurrence of backchannel responses in subsections compared to the type of the section (dialogic vs. narrative)

The tendency witnessed in the distribution of backchannel responses in the subsections was not correlated with the role of the speaker (Fw1 vs. Fw2), see *Figure 5.*

Concerning the phonetic environment, 80% of backchannel responses were realized while the other person was talking, and only 20% occurred during intervals, for example significantly fewer signals were given while the actual speaker paused. Compared to the total number of a given backchannel response, during pauses only 22% of humming 20% of laughter and 17% of *aha* ‘uhhuh’ and *igen* ‘yes’ were found in the recordings.

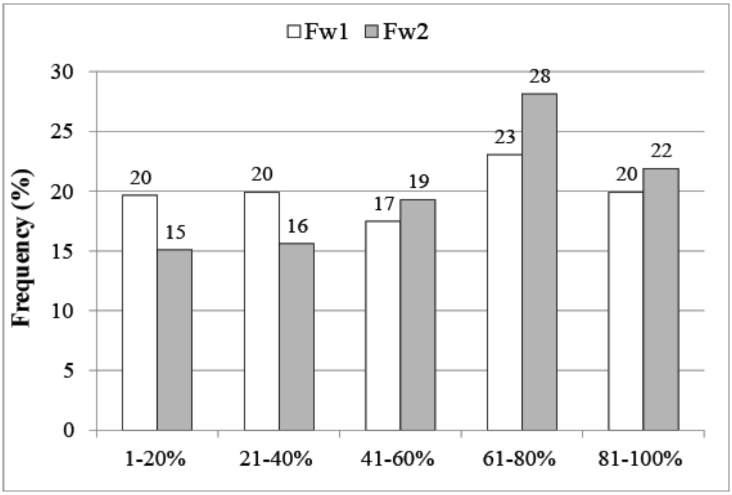

Figure 5. The occurrence of backchannel responses in the subsections compared to the person who is speaking

The duration of the most frequent backchannel signals was also analyzed (see *Figure 6*). The nonverbal type of laughter, the semi-type humming, and the verbal type *igen* 'yes' were examined. We only measured the occurrences in pauses. Laughter lasted for 410 ms on average (min: 160 ms, max. 850 ms), and most often their realization was between 200 and 300 ms long Variability within each speaker was analyzed for Fw1 and Fw2 respectively, as they were the same people in all 10 interviews, while the Subject was a different person in each conversation. Laughter was realized for both Fw1 and Fw2 (mean for Fw1: 430 ms; mean for Fw2: 429 ms; standard deviation for Fw1: 195 ms; standard deviation for Fw2: 81 ms).

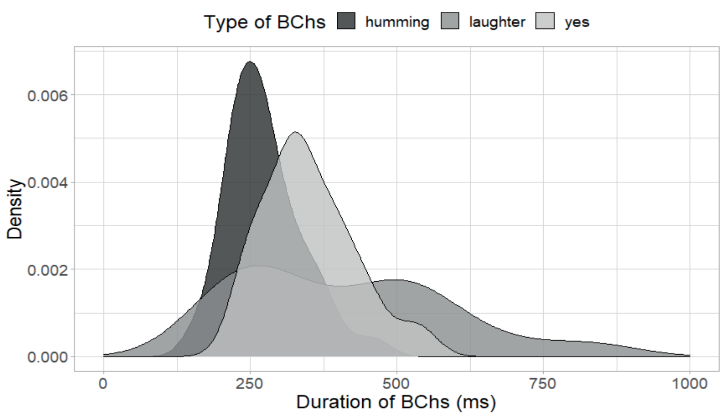

Figure 6. Density of the most frequent BChs

Hummings were shorter than laughs on average (273 ms), and the deviation was also lower (min: 151 ms; max. 472 ms). However, humming most often lasted for 200–300 ms, similar to laughter. The duration of humming was very similar for the two speakers (Fw1, Fw2), (mean for Fw1: 272 ms; mean for Fw2: 259 ms; SD for Fw1: 64 ms; SD for Fw2: 59 ms). The average duration of *igen* 'yes' backchannel responses was 349 ms, (min.: 248 ms; max.: 534 ms); most often they were realized between 300 and 400 ms. These backchannel responses, given by Fw1, were typically longer than those of Fw2, but owing to the number of occurrences, statistical tests were not carried out (mean for Fw1: 390 ms, mean for Fw2: 308 ms; standard deviation for Fw1: 78 ms; standard deviation for Fw2: 76 ms).

3.3. Conclusion

Our results of analyzing the backchannel responses confirmed that several types of backchannel responses exist which do not form discrete categories. Rather, they form a continuum, based on the degree of their verbal elaboration. Nevertheless, we had to create discrete categories for acoustic-phonetic analysis according to objective aspects. Within these dimensions, variability is high. The analysis of more than 800 backchannel signals corroborated that semi-verbal backchannel responses occurred the most often. Although in some conversations the three speakers used a similar number of backchannel signals, in most cases, it was the fieldworker1 who dominated the backchannel.

Frequency of occurrence does not only depend on the person of the speaker (cf. Tottie 1991), but it is also influenced by the kind of roles speakers take. Most of the backchannel responses were recorded as coming from the fieldworker1 in all conversations. The Fw1 led the conversations and initiated new topics. The lowest number of backchannel signals were produced by the Subject (S) in each conversation. According to the protocol, it is the S who should speak and use the main floor of the conversation.

As the occurrence of backchannel responses is also influenced by the inner structure of the conversation and its dynamic change in time. Most backchannel signals were found in subsection 4 (61–80%). In the microstructure of conversations, towards the end of each section, the number of backchannel responses increased.

The number of backchannel responses in the narrative sections was twice as many as in the dialogic parts (mean: 114 and 55, respectively). This suggests that if a speaker delivers a longer narrative speech section, the listener feels the urge to signal their attention by vocal responses. In contrast, in dialogic parts, the partners often change roles, reflecting on each other's statements continuously, which means less backchannel feedback is required.

Most backchannel responses (80%) were realized as simultaneous speech, only 20% occurred during silent pauses. This is in line with the function of backchannel responses: they serve to signal attention and provide support for the speaker, without interrupting the current speaker.

The analysis of the most frequent backchannel signal in our corpus proved that the most common response was the shortest one, for example humming. Markó (2007) recorded very similar means and minimum duration for affirmative humming. Our data showed as well, that the duration of laughs and hums were approximately of the same duration in the speech of Fw1 and Fw2, although the duration of *igen* 'yes' displayed bigger differences between the two speakers. Laughs were the longest phenomena out of the three backchannel responses examined in depth. Laughs produced as backchannel responses were shorter on average than reported by previous studies (cf. Neuberger 2017; however, in this study, the laughs with different functions were not differentiated).

In the present corpus, the minimum duration of hums and laughs was around 150 ms, which is equal to the duration of a Hungarian syllable made up of two sounds (see e.g. Krepsz and Gósy 2017). The unit of perception is the syllable (Lehiste 1972). Nonverbal backchannel responses, however, do not have syllables, but in order to allow them to fulfill their discourse organizing function, they must have the appropriate duration to ensure perceivability.

4. Experiment 2

Based on Experiment 1, the following questions may arise: (i) to what extent does the age of the partner influence backchanneling behavior (ii) can the characteristics of BChs be considered as individual factors? The present study examines only one speaker's BChs patterns addressed to listeners of different ages in order to reduce the number of influencing characteristics and to focus on the age factor as well as on the effect of elderspeak. The study aims to analyze BChs in spontaneous conversations in terms of the participants' age for the very first time in the Hungarian language. Furthermore, we aimed to investigate the realization of BChs as well as the fieldworker1's pausing and F0 as possible evidence for elderspeak, analyzing this topic also for the first time, in Hungarian triadic conversations (iii) to explore the BChs' main phonetic characteristics, iv) to find out if there is a difference in the realization of BChs the fieldworker1 uses, and whether it depends on the partners' age, v) to see if there is evidence for elderspeak in the usage of BCh responses aside from some phonetic parameters (pausing, F0, articulation rate) in the fieldworker1's speech.

We investigated these questions using a sociolinguistic approach, relying on a variationist perspective and the concepts of elderspeak (Coupland et al. 1988;

Kemper et al. 1995). In addition, earlier studies corroborated for Hungarian conversations, that some conversational phenomena were influenced by the age of the experimental person (cf. Bata 2009). Therefore, our hypotheses were the following: i) the frequency of the fieldworker1's backchannels is influenced by the speaker's age: more BChs are expected in the conversations with older speakers due to politeness ii) as well as more frequent backchannels, the fieldworker1 produces more and longer silent pauses, speaks with less articulation rate and a higher fundamental frequency when talking to older speakers because of elderspeak (cf. Coupland et al. 1988; Kemper et al. 1995). In addition, we assumed iii) the frequency of BChs changes across the sections of the conversations.

4.1. Data, method, and participants

20 conversations (about 250 min, average duration: 13,5, min: 7 min., max: 23,5 min.) from the Conversational Corpus (Horváth et al. 2019) were analyzed and annotated using Praat software (Boersma and Weenink 2018). The recordings were quasi-spontaneous as the topic of the conversation had been added at the beginning of the recordings. The participants had no time to plan the speech production, and the outcome depended on the opinion and utterances of other speakers. Three people participated in each conversation: the fieldworker1 (who coordinated the conversations), the fieldworker2 (a colleague of the fieldworker1), and the subject (S). The fieldworker1 (Fw1) and the fieldworker2 (Fw2) were the same people in each case. The speakers (S) were different people in the conversations (5 young males, 5 young females ages between 20–35 years, mean: 28 years, and 5 older males, 5 older females ages between 60–75 years, mean: 69 years, with typical hearing thresholds). In this study, the fieldworker1 was the experimental person. The Fw1 was a monolingual Hungarian female speaker working in the field of linguistics. She was 28 years old when the series of recordings began. The time span between the first and last recordings was approximately 10 years.

The BChs were analyzed in the Fw1's speech in terms of the S's age. On the other hand, the realization of BChs (forms and frequency) was analyzed in respect of the global structure of the conversations using the following method: the duration of the conversations was cut into 5 equal sections using a Praat script. The occurrence of BChs was analyzed in these equal parts. In addition, we analyzed the fieldworker1's first introductory section. In this section, she set out the topic, talked about it herself as well, and then she asked the other's opinion about it (mean duration of these sections in the 20 conversations: 30 s, SD: 16 s, min.: 12 s, max.: 60 s). This section was structured similarly in each conversation so that data could be compared. We analyzed the articulation rate (AR), the fundamental frequency (F0), and the silent pauses (frequency, duration) of the fieldworker1 in terms of the subjects' age. The influence of other factors (emotions, speech accommodation, mutual involvement) was minimal at the beginning of the conversations.

The values of the articulation rate, the fundamental frequency and the silent pauses (frequency and duration) were analyzed by linear mixed models in the R program (R Core Team 2018) with the lme4 package (Bates et al. 2015), and the *p* values were obtained by Satterthwaite approximation (lmerTest package, ANOVA function, Kuznetsova et al. 2015). The independent factors were the ratio of BChs, F0 values, duration, and the ratio of silent pauses, while the dependent factors were the five-partitions and the age groups. For each parameter, a random intercept and a random slope model were prepared (with the speaker as a random factor for each variable) and compared to the two models (with ANOVA in lmerTest, Kuznetsova et al. 2015). There were no significant differences between the models, and because of the lower AIC (Akaike 1973) values, the random slope models were used in the results.

4.2. Results

The ratio of backchannels was also measured in terms of the total speech time of the fieldworker1. In the 10 conversations, the ratio of all backchannels' duration of the total speech time was 9.7% on average. The ratio of BChs was the greatest in the conversations with elderly females, on average (12.6%). Backchannels occurred in the smallest ratio (6.7%) of the Fw1's total speech time while talking to young females. It must be noted that major differences were found between the conversations, irrespectively of the S's age (see *Figure 7*).

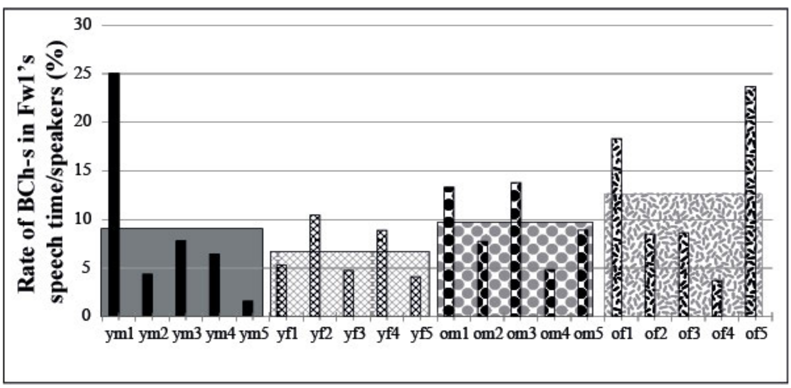

Figure 7. The ratio of backchannels in the speech time of the Fw1

The occurrence of BChs does not only depend on who the partner is, but it is also influenced by the structure and the dynamical change of the conversation. Conversations were cut into 5 equally long subsections (1-20%, 21-40%, etc.) using a Praat script. The number of BCh responses was calculated for each subsection (see Figure *8*). In the group of elderly females, the Fw1 produced the fewest BChs in the last section of the conversations. In the group of elderly males, however, the most BChs from the Fw1 were found in the first and last sections of the conversations, similar to the tendency found in the group consisting of young females. The Fw1 produced fewer BChs at the beginning and the end of the conversations with young male speakers.

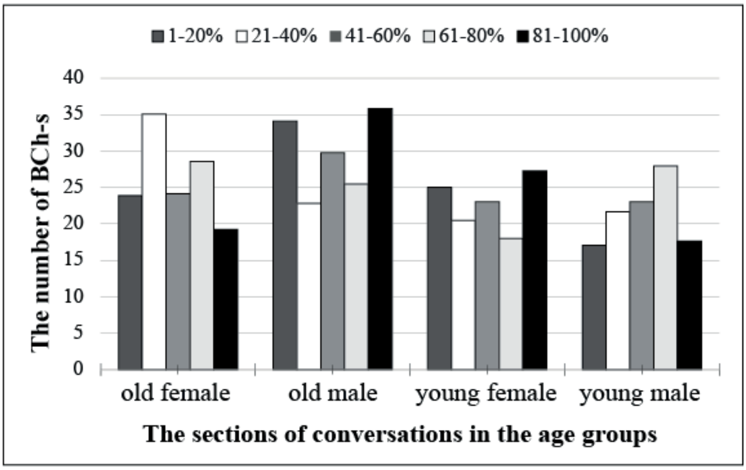

Figure 8. The number of backchannels in the subsections of the conversations

We analyzed the ratio of the main types of BChs. The borderline cases between non-lexical and phrasals (*hm*, *aha*, etc.) occurred in half of the cases (see *Figure 9*). The verbal and nonverbal BChs occurred in very similar proportions, while the mixed BChs were the least frequent (e.g. laughter + *yes*).

We analyzed the correlation between the duration of conversation and the number of BChs. A strong positive linear correlation ($r = 0.723$, $p < 0.001$) was found between the duration of the conversation and the number of BChs. The longer the S's IPUs were, the more BChs the Fw1 produced.

The articulation rate, the silent pauses, and the fundamental frequency were analyzed in each age group regarding the possible phonetic characteristics of elderspeak alongside the backchannel behavior. The articulation rate of the fieldworker1 was analyzed in each IPUs of the introductory section of the conversations regarding the subjects' age. The age of S did not affect the fieldworker1's AR. The mean AR of I was 5.6 syll/sec (SD: 1.3 syll/sec) when speaking to elderly females and 5.7 syll/sec (SD: 1.3 syll/sec) when speaking to

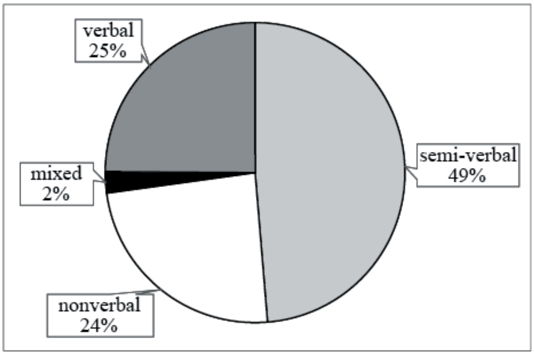

Figure 9. The ratio of the main BCh types

elderly males. The mean AR of I was 5.5 syll/sec (SD: 1.2 syll/sec) when speaking to young females and 5.9 syll/sec (SD: 1.6 syll/sec) when speaking to young males.

The frequency of the Fw1's silent pauses (SPs) showed significant differences across conversations (mean: 18 SPs/minute, SD: 8 SPs/minute). The age of the subjects had not influenced the duration of Fw1's silent pauses. The mean frequency of her pauses was 21 pauses per minute (SD: 9 SPs/min) during the introductory sections of the conversations with young subjects, while it was 16 pauses per minute (SD: 6 SPs/min) talking with older subjects. In general, the fieldworker1's silent pauses were more frequent when speaking to male subjects than when speaking to female subjects (Table 1.). The frequency of silent pauses showed a significant difference between the groups (F(2 19) = 4.939, p = 0.021, η^2 = 0.382]. The Fw1's silent pauses were significantly more frequent (Bonferroni post-hoc: p = 0.014) when talking to young male subjects than when talking to young female subjects. Finally, SPs were more frequent in conversations with young male subjects than in those with elderly females (p = 0.003).

Table 1: The frequency of silent pauses in the fieldworker1's introductory session

	The frequency of silent pauses per minute	
	mean	**SD**
young female	15.8	4.3
young male	27.3	9.3
elderly female	12.9	6.7
elderly male	19.3	4.9

The duration of the fieldworker1's silent pauses was also analyzed. The age of the subjects had not influenced the duration of the Fw1's silent pauses. The mean duration of her pauses was 462 ms (SD: 322 ms) in the introductory sessions of conversations with young subjects, while it was 474 ms (SD: 308 ms) in the conversations with older subjects. However, the duration of silent pauses showed a significant difference between the groups ($F(2\ 196) = 5.912$, $p = 0.003$, $\eta^2 = 0.058$, cf. Figure 10). The Fw1's silent pauses were significantly shorter (Bonferroni post-hoc: $p = 0.046$) when talking to older female subjects (mean: 364 ms, SD: 292 ms) than when talking to older male subjects (mean: 539 ms, SD: 301 ms).

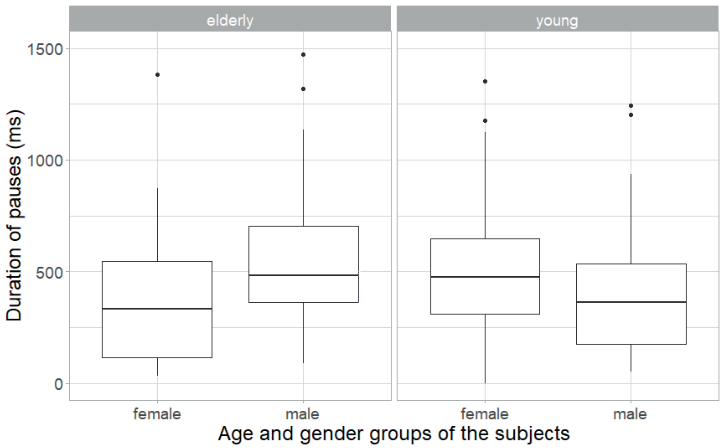

Figure 10. The duration of the fieldworker1's silent pauses with the regard to the subject's age and gender

We analyzed the fundamental frequency of the fieldworker1 in her introductory section (see *Figure 11*). The age of the partner had not influenced the Fw1's F0 in the introductory section of the conversations. However, the partner's gender had a statistically significant effect on the fieldworker1's F0 in the conversations with younger subjects. The fieldworker1 spoke to the young males with lower fundamental frequency than to the young females ($F(1, 2945) = 36.422$, $p < 0.001$).

4.3. Conclusions

The present study has offered an analysis of some aspects of backchanneling behavior, mainly concerning the partner's age. Contrary to our hypothesis, the partner's age did not significantly influence the frequency and the duration of the fieldworker1's BChs. However, the most frequent uses of BChs on the

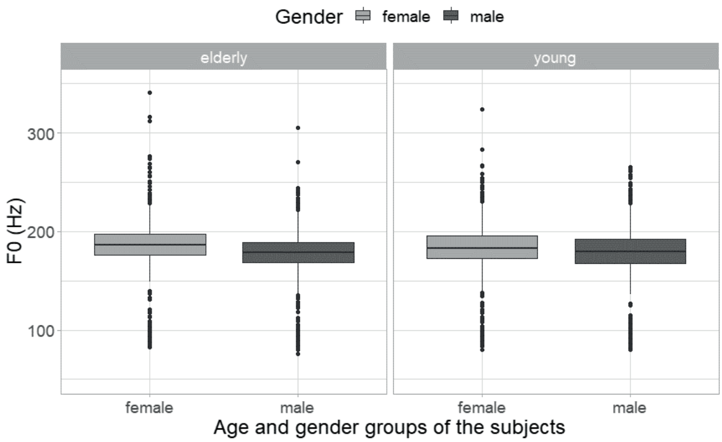

Figure 11. The F0 of the fieldworker1 with the regard to the subject's age and gender

fieldworker1's part were detected in conversations with older women, which may be explained by the presence of intimacy between female speakers, and in the way, supportiveness is expressed (Marche 1988). Such findings may also be interconnected with politeness strategies, but this needs further analysis.

Significant differences were found in the frequency of BChs between individual conversations; the age groups were not homogeneous in terms of the backchannel behaviour of the fieldworker1.

The dynamic changes in the frequency of backchannel responses were analyzed. The sections of conversations (the total time of each conversation were cut into five equal sections) contained various numbers of backchannels. We found that the frequency of BChs showed dynamic changes across conversations.

The most frequent BChs are semi-verbals (hummings, *aha*, etc.) (cf. Kjellmer 2009). Semi-verbals are the basic category of BChs (see Experiment 1), and their usage was not influenced by the age of the partner (nor by socio-cultural norms which are common vis-a-vis elderly people). The multifunctionality (serving many functions: acknowledgment, surprise, acceptance, etc., cf. Coulthard et al. 1981) results in their frequency.

The interrelation between the duration of the conversations and the total number of BChs were analyzed. A positive correlation was found between the duration of the conversation and the number of BChs. The longer the S's IPU was, the more BChs were produced by the fieldworker1. Longer speech time and more backchannels may mark mutual involvement. Besides, the interviewer wants to signal her support and attention while the subject is speaking.

Some phonetic characteristics of the fieldworker1's speech were analyzed in terms of elderspeak, using the first introductory narrative part of the conversations. The articulation rate of the fieldworker1 had not been affected by the age of the subjects. Our hypothesis that the fieldworker1 would speak at a slower speech rate to older subjects was not confirmed. Furthermore, more and longer silent pauses were expected in her speech with older people than with younger ones. Substantial differences were found in the frequency of silent pauses between the conversations (from 1 silent pause per minute to 39 silent pauses per minute). However, the age of the subject did not have a significant effect, contrary to our hypothesis. The topic or the fieldworker1's involvement may have affected the pausing strategies to a greater extent than the partner's age. However, not so much age-related, but some gender-differences were found: the fieldworker1's silent pauses were more frequent when talking to male subjects than when talking to female subjects, regardless of the age group. In addition, the fieldworker1 spoke to the young male subjects with the most frequent silent pauses, in the introductory part of the conversations. However, as pausing can be related to many phenomena in a complex way (breathing, speech planning and speech managing, politeness, cultural differences), we cannot yet account for gender differences.

The F0 of the interviewer was also analyzed as a possible marker of elderspeak. Similarly to the pausing strategies, the age of the partner did not influence the F0 values of Fw1 in the introductory part. However, some gender-related differences were found again, instead of age-related ones: the fieldworker1 spoke to the young males with lower fundamental frequency than she did to young females. The relationship between the changes in fundamental frequency (F0) and the attractiveness or dominance of the interlocutor belonging to the opposite sex has been investigated in several studies. According to several research studies, the higher frequency voices of women and the lower pitch of men are more attractive for the opposite sex (e.g. Collins and Missing 2003; Feinberg et al. 2008). However, other studies suggest that this phenomenon is more complex in nature. Although male students found a higher voice more attractive in a female interlocutor, the relationship was not linear: female voices above 280 Hz were rated as less attractive, presumably because very high female voices were associated with sexual immaturity (Borkowska and Pawlowski 2011, the same tendency for men: cf. Saxton et al. 2016). Instead of an increase in F0, two other studies found the opposite trend: both sexes used a lower-pitched voice when speaking to a more attractive subject of the opposite sex (Hughes et al. 2010). Pisanski et al. (2018) found that women spoke in a higher and more variable voice pitch to men they had selected as potential mates but lowered their pitch with the most attractive men. Thus, it was concluded that individual- and group-level mate preferences may differ from each other and this is also reflected in pitch. This, however, does not change the fact that a deeper pitch is associated with a higher degree of dominance for both sexes (Fraccaro et al. 2013). In our case, the significantly lower pitch of the female fieldworker1 when talking to young men can be explained by dominance: as a woman, she wanted to radiate an air of competence and seriousness to men of the same age. This shows the presence of the active discursive construction in the preassigned interaction role (interview leader) and the related identity traits.

5. Discussion

Hungarian backchannels occur with various realizations along a continuum from non-lexical types to multiple phrasal expressions. The frequency of backchannels was influenced by the speaker's role with the fieldworker1 producing responses the most often. On the other hand, the frequency was affected by the global structure of the conversations: narrative sections contained more BChs. Besides, the number of backchannel responses increases towards the end of each section. Backchanneling has a role in closing down parts of the conversations, in organizing discourse structure.

The usage of BChs was also analyzed concerning the partner's age. Data suggest that the backchanneling behavior of a speaker is not modified by the speaker's age itself. Contrary to our hypothesis, the usage of BChs did not show significant elderspeak-specific differences. Furthermore, no elderspeak differences were found in the parameters of the fieldworker1. Also, contrary to our hypothesis, the fieldworker1 did not use a higher pitch register when talking to older subjects (similar to Kemper et al. 1995). With regard to the articulation rate, no evidence was found for elderspeak in the introductory part of the conversations, contrary to our hypothesis which was based on an earlier study (Kemper et al. 1995). The groups of elderly speakers were not homogeneous as to the fieldworker1's speech parameters: substantial differences were found between each conversation in one age group.

Based on the material, we found gender-related differences rather than evidence for elderspeak. Furthermore, the group of young males differed the most from the other groups concerning the speech parameters of the fieldworker1. The fieldworker1 spoke to the young males with the lowest pitch and with the greatest number of silent pauses. The articulation rate was the least affected parameter by the subject's age and gender. In a previous study, the participants were asked to simulate a ‘‘sexy voice’’. Results showed that both men and women decreased the pitch of their voices in order to sound sexier. Females lowered the F0 of their voices even more than did the males (Tuomi and Fisher 1979). In another experimental research, both men and women equally lowered their pitch of their voices while talking to an attractive target (Hughes et al. 2010). According to other earlier studies, individuals speak to males and females differently, others rate vocal samples directed to men as sounding more dominant and formal than samples directed towards women (Hall and Braunwald 1981). In addition, an experimental study corroborated that in mixed conversations, females accommodate to the backchanneling style of the males producing more overlapping BCs than in single-sex conversations (Feke 2003). In sum, the attributes of the partners may affect not only the prosodic parameters but the communication strategies of the participants as well.

The present study offers corpus-based analyses of this phenomenon, from various perspectives, which provide pivotal findings from both a theoretical and applied viewpoint. The results of backchannel research are paramount in many applied fields, such as foreign language teaching, intercultural training courses, translation and interpretation and in artificial intelligence applications, especially, in automatic speech processing or the development of chatbots and social chat robots (cf. Jurafsky et al. 1997; Lala et al. 2017). Results raised further research questions on backchanneling behavior influenced by the topic itself, the number of the participants, the connection between the participants or the speech setting (online vs. offline conversations).

References:

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Caski (eds.), *Second International Symposium on Information Theory*, 267-281. Budapest: Akadémiai Kiadó.

Bata, S. 2009. A társalgás fonetikai jellemzőinek alakulása a beszédpartnerek életkorának függvényében [The phonetic patterns of conversations with regard to the partner's age.] In T. Váradi (ed.), *III. Alkalmazott Nyelvészeti Doktorandusz Konferencia*, 3-13. Budapest: MTA Nyelvtudományi Intézet.

Bates, D., M. Mächler, B. Bolker and S. Walker 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 671: 1-48.

Benus, S., A. Gravano and J. Hirschberg 2007. The prosody of backchannels in American English. In J. Trouvain (ed.), *Proceedings of 16th International Conference of Phonetic Sciences*, 1065-1068. Saarbrücken: Univ. des Saarlandes. https://www.academia.edu/15042478/THE_PROSODY_OF_BACKCHANNELS_IN_AMERICAN_ENGLISH

Boersma, P., and D. Weenink 2018. *Praat: doing phonetics by computer*. http://www.fon.hum.uva.nl/praat/download_win.html. [Retrieved: 21 November 2018]

Borkowska, B., and B. Pawlowski 2011. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour* 82(1): 55-59.

Carli, L.L. 1989. Gender differences in interaction style and influence. *Journal of Personality and Social Psychology* 56: 565-576.

Cathcart, N., J. Carletta and E. Klein 2003. A shallow model of back-channel continuers in spoken dialogue. In *EACL 03: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics – Volume* 1: 51-58.

Collins, S.A., and C. Missing 2003. Vocal and visual attractiveness are related in women. *Animal Behaviour* 655: 997-1004.

Coulthard, M., M. Montgomery and D. Brazil 1981. Developing a description of exchange structure. In M. Coulthard and M. Montgomery (eds.), *Studies in Discourse Analysis*, 1-51. London: Routledge.

Coupland, N., J. Coupland, G. Howard, K. Henwood and J. Wiemann 1988. Elderly self-disclosure: interactional and intergroup issues. *Language and Communication* 8: 100-131.

Croft, W., and A.D. Cruse 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cutrone, P. 2005. A case study examining backchannels in conversations between Japanese–British dyads. *Multilingua – Journal of Cross-Cultural and Interlanguage Communication* 243: 237-274.

Deng, X. 2009. Listener response. In S.D. Hondt, J.O. Ostman and J. Verschueren (eds.), *The Pragmatics of Interaction*, 104-125. Amsterdam - Philadelphia: John Benjamins.

Edelsky, C., and K. Adams 1990. Creating inequality: Breaking the rules in debates. *Journal of Language and Social Psychology* 9: 171-190.

Feinberg D.R., L.M. DeBruine, B.C. Jones and D.I. Perrett 2008. The role of femininity and averageness of voice pitch in aesthetic judgments of women voices. *Perception* 37: 615-623.

Feke, M.S. 2003. Effects of native-language and sex on back-channel behavior. In S. Lotfi (ed.), *Selected Proceedings of the First Workshop on Spanish Sociolinguistics*, 96-106. Somerville, MA: Cascadilla Proceedings Project.

Fraccaro, P.J., J.J. O' Connor, D.E. Re, B.C. Jones, L.M. DeBruine and D.R. Feinberg 2013. Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour* 85: 127-136.

Gould, O.N., and R.A. Dixon 1993. How we spent our vacation: Collaborative storytelling by young and old adults. *Psychology and Aging* 81: 10-17.

Gravano, A., S. Benus, H. Chavez, J. Hirschberg and L. Wilcox 2007. On the role of context and prosody in the interpretation of okay. In C. Biemann, V. Seretan and E. Riloff (eds.), *Proceedings of the ACL 2007 Student Research Workshop*, 800-807. Prague: Association for Computational Linguistics.

Gravano, A., and J. Hirschberg 2009. Backchannel-inviting cues in task-oriented dialogue. In M. Uther, R. Moore and S. Cox (eds.), *Proceedings of INTERSPEECH 2009*, 1019-1022. Brighton: ISCA.

Hall, J.A., and K.G. Braunwald 1981. Gender cues in conversations. *Journal of Personality and Social Psychology* 40: 99-110.

Hockey, B.A. 1993. Prosody and the role of okay and uh-huh in discourse. In M. Bernstein (ed.), *Proc. ESCOL '92*, 128-136. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.2337&rep=rep1&type=pdf

Horváth, V., V. Krepsz, D. Gyarmathy, Á. Hámori, J. Bóna, Cs. I. Dér and Zs. Weidl 2019. Háromfős társalgások annotálása a BEA-adatbázisban: elvek és kihívások. [The principles and challenges of annotating the tryadic conversations in the

Hungarian Spontaneous Speech Database BEA]. *Nyelvtudományi Közlemények* 115: 255-274.

Hughes S.M., S.D. Farley and B.C. Rhodes 2010. Vocal and physiological changes in response to the physical attractiveness of conversational partners. *Journal of Nonverbal Behaviour* 34: 155-167.

Hutchby, I., and R. Woofitt 2006. *Conversation Analysis: Principles, Practices and Applications.* Cambridge: Polity Press.

Jurafsky, D., R. Bates, N. Coccaro and R. Martin et al. 1997. Automatic detection of discourse structure for speech recognition and understanding. In *Proceedings of the 1997 IEEE Workshop on Speech Recognition and Understanding,* 88-95. Santa Barbara: IEEE.

Kemper, S., D. Vandeputte, K. Rice, H. Cheung and J. Gubarchuk 1995. Spontaneous adoption of elderspeak during referential communication tasks. *Journal of Language and Social Psychology* 14: 40-59.

Kjellmer, G. 2009. Where do we backchannel? On the use of *mm, mhm, uh, huh* and *such like*. *International Journal of Corpus Linguistics* 141: 81-112.

Krepsz, V., and M. Gósy 2017. Igekötős szavak időszerkezete [Timing patterns of verbs with prefix]. In M. Gósy and V. Krepsz (eds.), *Morfémák időzítési mintázatai a beszédben*, 77-98. Budapest: MTA Nyelvtudományi Intézet.

Kuznetsova, A., P.B. Brockhoff and R.H.B. Christensen 2015. LmerTest: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1-26. https://doi.org/10.18637/jss.v082.i13

Lala, D., P. Milhorat, K. Inoue, M. Ishida, K. Takanashi and T. Kawahara 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In K. Jokinen, M. Stede, D. DeVault and A. Louis (eds.), *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Association for Computational Linguistics*, 127-136. Saarbrücken: Association for Computational Linguistics.

Lehiste, I. 1972. The units of speech perception. In G.H. Gilbert (ed.), *Speech and Cortical Functioning.* 187-237. New York – London: Academic Press.

Li, H.Z., Y. Cui and Z. Wang 2010. Backchannel responses. Backchannel Responses and Enjoyment of the Conversation: The More Does Not Necessarily Mean the Better. *International Journal of Psychological Studies* 21. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.612.2846andrep=rep1andtype=pdf. [Retrieved: 12 June 2019]

Marche, T.A. 1988. *The developmental and gender related use of listener responsiveness and interruption behavior.* MSc Thesis. Newfoundland: Department of Psychology Memorial University of Newfoundland.

Markó, A. 2007. A special conversational device: humming in Hungarian. *The Phonetician* 95: 28-35.

Markó, A., M. Gósy and T. Neuberger 2014. Prosody patterns of feedback expressions. In Hungarian spontaneous speech. In N. Campbell, D. Gibbon and D. Hirst (eds.), *Social and Linguistic Speech Prosody: Proceedings of the 7th* *international conference on Speech Prosody*, 482-486. Dublin: Science Foundation Ireland.

Maynard, S.K. 1989. *Japanese Conversation: Self-Contextualization through Structure and Interactional Management.* Norwood, NJ: Ablex.

Neuberger, T. 2017. Nonverbális hangjelenségek a beszédben. [Nonverbal expressions in speech.]. In J. Bóna (ed.), *Fonetikai olvasókönyv*, 75-90. Budapest: ELTE Fonetikai tanszék.

Neuberger T., D. Gyarmathy, T.E. Gráczi, V. Horváth, M. Gósy, A. Beke 2014a. Development of a large spontaneous speech database of agglutinative Hungarian language. In P. Sojka, A. Horák, I. Kopeček, K. Pala (eds.), *Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science*, vol 8655. Springer, Cham. https://doi.org/10.1007/978-3-319-10816-2_51

Neuberger, T., A. Beke and M. Gósy 2014b. Acoustic analysis and automatic detection of laughter in Hungarian spontaneous speech. In S. Fuchs, M. Grice, A. Hermes, L. Lancia and D. Mücke (eds.), *Proceedings of the 10th International Seminar on Speech Production ISSP*, 285-287. *Cologne*: Univ.

Norris, N.R. 2000. *Conversational Narrative: Storytelling in Everyday Talk*. Amsterdam-Philadelphia: John Benjamins.

Pisanski, K., A. Oleszkiewicz, J. Plachetka, M. Gmiterek and D. Reby 2018. Voice pitch modulation in human mate choice. *Proceedings of the Royal Society B.: Biological Sciences* 285: 20181634. https://royalsocietypublishing.org/doi/10.1098/rspb.2018.1634

Preston, D.R., and N.A. Niedzelski (eds.) 2010. *A Reader in Sociophonetics*. Mouton De Gruyter.

Prévot, L., J. Gorisch and R. Bertrand 2016. A CUP of CoFee: A Large Collection of Feedback Utterances Provided with Communicative Function Annotations. In Calzolari et al. (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC16*, 3180-3185. European Language Resources Association ELRA. https://www.aclweb.org/anthology/L16-1507.pdf

R Core Team 2018. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna. https://www.R-project.org

Sacks, H., E. Schegloff and G. Jefferson 1974. A simplest systematic for the organization of turn-taking for conversation. *Language* 50: 696-735.

Saxton, T.K., L.L. Mackey, K. McCarty and N. Neave 2016. A lover or a fighter? Opposing sexual selection pressures on men's vocal pitch and facial hair. *Behavioral Ecology* 272: 512-519.

Schegloff, E.A. 1982. Discourse as an interactional achievement: Some uses of uh huh and other things that come between sentences. In D. Tannen (ed.), *Analyzing Discourse: Text and Talk*, 71-93. Georgetown: University Press.

Tottie, G. 1991. Conversational style in British and American English: The case of backchannels. In K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik*, 254-271. London: Longman.

Tolcsvai Nagy, G. 2001. *A magyar nyelv szövegtana.* [Text Grammar of Hungarian.] Budapest: Nemzeti Tankönyvkiadó.

Tuomi, S., and J. Fisher 1979. Characteristics of Simulated Sexy Voice. *Folia Phoniatrica* 31: 242-249.

Ward, N., and W. Tsukahara 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics* 32: 1177-1207.

Wong, D., and H. Kruger 2018. Yeah, yeah yeah or yeah no thats right: A multifactorial analysis of the selection of backchannel structures in British English. In V. Brezina, R. Love and K. Aijmer (eds.), *Corpus Approaches to Contemporary British Speech. Sociolinguistic Studies of the Spoken BNC2014*, 120-156. New York - London: Routledge.