HESITATING WITH AND WITHOUT LANGUAGE HERITAGE - PROSODIC ASPECTS OF FILLER PARTICLES IN THE RUEG CORPUS

Marlene Böttcher, Margaret Zellers

ISFAS, Kiel University, Germany
mboettcher@isfas.uni-kiel.de, mzellers@isfas.uni-kiel.de

ABSTRACT

This study investigates the frequency and prosodic form of filler particles (FP, such as *uh* and *uhm*) produced by speakers of English, with and without a heritage language (HL) background. This is done based on semi-spontaneous corpus data. A subset of 24 speakers with Russian, German or no HL background, each producing 2 narrations, was analyzed in this study. There were 304 FPs realized in the 50 min of analyzed speech. Overall English speakers' productions show close similarity in FP form. The analysis shows no effect of HL status for FP frequency and frequency of different FP forms (*uh* vs *uhm*) as well as FP duration and prosodic phrasing. The acoustic analysis of pitch slope, however, showed differences between speaker groups with different and without HL background. Interesting results emerged when including gender into the analysis, which will be addressed in future work.

Keywords: filler particles, hesitation, spontaneous discourse, heritage speakers, bilingualism

1. INTRODUCTION

Spontaneous discourse differs from prepared or read speech in many ways. As part of its unplanned nature, it contains many disfluencies like repetitions, repairs and fillers. Such disfluencies can comprise up to ten percent of words in natural conversations [1]. One type of filler in English are the constituents *uh* and *uhm* (and their variants). Different terms have been used in the literature (e.g. *hesitation, disfluency, error* or *filled pause, discourse particle*) reflecting researchers' approaches on fillers as either symptoms [2, 3], or signals in discourse [4, 5, 1] or stressing their similarity with silent pauses [6, 7]. Following [8] we use the term *filler particle* (FP) to refer to lexically underspecified, non-inflectable entities which are frequently produced in naturally occurring speech, in an effort to be agnostic about the status and function of FPs. Generally, FPs are produced preceding new thought units. Their use is therefore connected to speech planning and language processing as well as discourse organisation [4]. Speech planning is an interesting area in bilingual research, since the availability of two languages and the suppression of one of them depending on the context of use demands a higher cognitive load in bilinguals' speech [9]. Higher cognitive load can then be related to slower speech rate and an increase of hesitation phenomena [9, 10]. Hence, the use of FPs has been found to be an interesting area of research in the study of bilingualism [11, 12] and heritage language use [13].

2. BACKGROUND

2.1. Filler particles

In their frequently cited work, [4] were among the first to classify FPs like *uh* and *uhm* as words "with conventional phonological shapes and meanings [...] governed by the rules of syntax and prosody" [4, p.75]. In English a central vowel is often followed by a nasal[1]. The dichotomy between a vocalic (V, *uh*) and a nasalized variant of FPs (VN, *uhm*) is well established [4, 5, e.g.]. and the use of the two forms has been related to gender-specific use patterns: while female speakers tend to produce more VN variants, male speakers seem to produce more V variants [14]. With regard to their prosody, FPs have been described with reduced fundamental frequency (F0) and a level or gradually falling F0 contour [1, 8]. Their phonological form of a central vowel optionally followed by a nasal and their reduced pitch characteristics leads to FPs having relatively low perceptual prominence [15]. FPs have been interpreted as symptoms of difficulties in speech planning [6] and have been used as an indicator for difficulties in bilingual speech due to higher cognitive load [10]. While FPs are connected to speech planning and processing, their occurrence is rule based and follows specific functions in spontaneous discourse [5]. FPs can be seen as one of many aspects of language learners have to acquire. Since FPs are perceptually not prominent they may pose challenges in language acquisition. That is, phenomena at the edge of our consciousness

are more difficult to learn than more salient aspects of language. This might be especially relevant in bilingual contexts with limited input such as in heritage language acquisition.

2.2. Filler particles in bilingual speech

Investigations of FPs in bilingual speech have shown deviances from monolingual FP use. In a study on English speakers of French, [11] found that bilinguals overly produce FPs along with other non-lexical markers compared to lexical discourse markers. These speakers also overused the vocalic *uh* form compared to *uhm* [11]. Similarly, a study by [16] on English speakers of Dutch found more FPs and silent pauses in their L2 speech compared to L1 [16]. While these findings suggest the acquisition of FP frequency and form to be a challenge for L2 learners other research suggests differently. In a study on English and te reo Māori bilinguals, [17] showed that these bilingual speakers display a language-specific use of FPs in both their languages. The acquisition of an appropriate FP use might therefore be related to the amount of exposure and the language status.

2.3. Bilingual heritage speakers

Heritage speakers (HS) represent a specific case of bilinguals speaking a heritage language and a majority language. The majority language (ML) is typically the dominant language of these bilinguals and used in everyday and most areas of the public sphere, e.g. work and education. The heritage language (HL) as a minority language in the larger society is acquired in specific contexts and typically spoken at home, e.g. with relatives and friends [13]. The use of the HL is therefore limited to certain interlocutors, genres and registers. Prior research has shown HL acquisition to be an interesting area of research regarding language acquisition, language contact phenomena and language change [18].

2.4. Research aim and hypotheses

This study addresses the productions of FPs in the speech of English speakers with Russian, German and no HL background. Specifically, the aspects of FP frequency, FP segmental form (vocalic *uh* (V) vs vocalic-nasal *uhm* (VN)) and the prosodic realisation of FPs are analysed. Following prior findings on FP use in bilingual speech, the hypotheses are:

1. Bilingual speakers of English produce FPs more frequently than monolingual speakers since the cognitive load of monitoring two languages is higher in bilingual speech planning [9].
2. Bilingual speakers deviate from the monolingual preference for a VN form of FPs resulting in a transfer of a V preference from heritage Russian [19].
3. Bilingual speakers deviate from monolingual realisations in duration, prosodic phrasing and slope resulting in transfer of discourse particle intonation from their HL (e.g. rising *nu* in Russian [20]).

3. DATA ANALYSIS

3.1. The corpus

The hypotheses presented above are investigated using the RUEG corpus, a corpus of semi-spontaneous speech including data from both mono- and bilingual speakers with different language backgrounds [21]. The narrations within the RUEG corpus were elicited by means of the Language Situation Method [22]: A video of a car accident was used to prompt participants to explain what happened in two situations. Participants were asked to provide a police report in form of a voice message on the phone, yielding a formal register. Participants were also asked to describe the incident to a friend by means of a voice message, yielding an informal situation. The hypotheses in this study do not make any predictions regarding possible register differences.

3.2. Acoustic analyses

For the acoustic analysis a subset of 24 speakers from the RUEG corpus data was selected (8 monolinguals (E), 8 bilinguals with Russian HL (R), 8 bilinguals with German HL (D); 11 male; mean age 16 y.) Since each speaker produced 2 narrations there were 48 narrations comprising 50 minutes of monologues. In this data, 303 FPs were realised, which were then further analysed for the acoustic parameters duration, pitch level and pitch slope. Both mean pitch and F0 maxima and minima were measured within the segmented FPs in Praat [23]. The speaker's mean pitch over the whole narration was also measured. The slope of the FP was then calculated as the difference between F0 maximum and F0 minimum in semitones (st). Slopes larger than 40 st were excluded as measurement errors considering physiological constraints [24]. Additionally, FPs were annotated for their segmental form and their prosodic phrasing based on universal

phonetic cues [25]. Based on the F0 measurements pitch contour was categorized as *level, falling* and *rising* with excursion size values larger than 1 semitone for the latter two categories. Statistical analyses were carried out in R using the lme4 package [26].

4. RESULTS

4.1. FP frequency

In the data analyzed here, 303 FPs were realised by 21 speakers in 39 narrations (i.e. not all narrations by all speakers contained FP productions). Given the different length of narrations, the frequency of FPs was normalized per 100 words. There were on average 4.6 FPs per 100 words produced in the data set (D: 4.3; R: 4.5; E: 4.6). A linear regression analysis revealed no effect of the speakers' language background or gender (all p-values > 0.30).

4.2. FP form

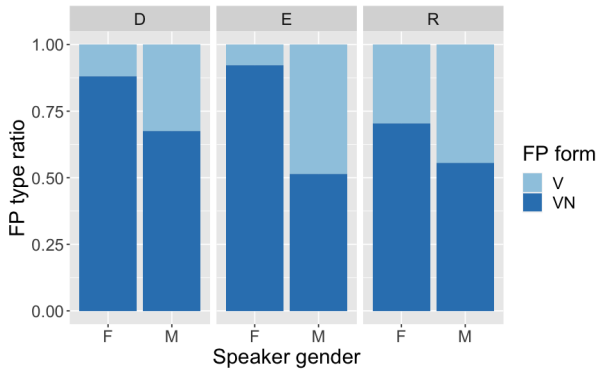

Figure 1: Ratio of vocalic (V) and vocalic-nasal (VN) filler particle (FP) forms across speakers with different language background (D: German heritage, E: monolingual, R: Russian heritage) and gender (F: female, M: male)

Overall the most frequent FP segmental form was vowel + nasal (VN: n=199) followed by vocalic forms (V: n=100). There were only few cases of nasal forms (N: n=3) and one instance of a glottalized variant; the latter two forms will be excluded in the following analysis. The preference for the VN variant is shared by all speaker groups with an average of 79% VN forms of FPs produced by speaker (D: 85%; R: 72%; E: 79% VN forms). Another observation relates to gender differences in the FP form preference. While all speakers prefer the VN form of FPs, female speakers show a higher use of VN forms and rarely V forms. The mean ratio of VN forms produced by speakers is 86% for female and 69% for male speakers. The gender of the speaker had a significant main effect on the mean VN proportion of FPs. A post-hoc Tukey pairwise comparison revealed a higher proportion of VN forms produced by female compared to male speakers (female: β = 17.13, SE = 6.89, df = 13.95, t = 2.48, p < 0.05). There seems to be a tendency for a less robust gender difference within the Russian heritage speaker group (cf. Figure 1). Yet no firm conclusions can be drawn from this small data set with only four female speakers in each group.

4.3. FP duration and phrasing

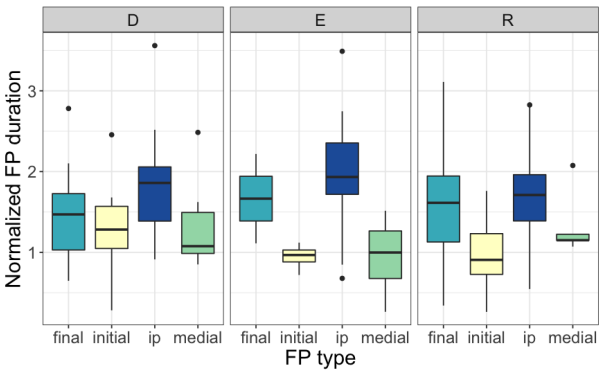

Figure 2: Normalized filler particle (FP) duration related to prosodic phrasing (ip: separately phrased) across speakers with different language background (D: German heritage, E: monolingual, R: Russian heritage)

The mean duration of FPs in the analysed data was 0.34 s with VN forms longer ($\bar{x}$= 0.38 s) than V forms ($\bar{x}$= 0.26 s). To compare FP duration across prosodic phrasing and language background the measured duration was normalized by the speech rate of the speaker. Both FP form (VN vs. V) and prosodic phrasing had a significant main effect on normalized FP duration. Post-hoc Tukey pairwise comparisons revealed a longer duration of FPs produced in final position and phrased separately compared to phrase initial position (final: β = 0.47, SE = 0.09, df = 287.29, t = 4.99, p<.001; initial: β = -0.57, SE = 0.08, df = 288.37, t = -7.51, p<.001) while separately phrase FPs (ip) were also longer than FPs in phrase medial position (ip: β = 0.33, SE = 0.12, df = 280.41, t = 2.83, p<.001). Additionally, a post-hoc Tukey pairwise comparison revealed a shorter duration of V forms compared to VN forms (V: β = -0.46, SE = 0.06, df = 289.81, t ratio = -7.90, p<.001). This is true for all speaker groups as is illustrated in Figure 2.

4.4. Filler pitch

Overall, the pitch level of FPs tends to be lower than the speaker's mean ($\bar{x}$= -1.35 st). The most frequent intonation contour produced on FPs was falling (n= 198; D: n= 72, 71%; E: n= 25, 54%; R: n= 101, 65%) while level (n= 49, D: n= 13, 13%; E: n= 4, 8%; R: n= 32, 21%) and rising contours (n= 56; D: n= 16, 16%; E: n= 17, 4%; R: n= 23, 15%) were comparably frequent. The slope in st per second was overall greater for rising contours ($\bar{x}$= 21 st/s) than for falling contours ($\bar{x}$= 12 st/s). Both language background and contour type are relevant factors for slopes of FPs. The linear mixed model for the slope of the pitch also showed a significant interaction for Russian HL speakers. Post-hoc Tukey pairwise comparisons revealed that FPs produced with a rising contour had a larger slope compared to FPs produced with falling intonation (falling: β = -6.31, SE = 1.82, df = 250.73, t ratio = -3.47, p < 0.01; level: β = -14.33, SE = 2.14, df = 253.65, t ratio = -6.69, p < 0.01) as well as larger slope of FPs produced with falling compared to level intonation (β = 8.02, SE= 1.48, df = 259.72, t ratio = 5.42, p < 0.01).

5. DISCUSSION

The results of this study illustrate FP productions by multilingual English speakers for the case of HSs living in the US with either German or Russian heritage background.

The analysed data does not provide evidence for an increased FP use due to higher cognitive load in bilingual speakers. The first hypothesis therefore has to be rejected contrary to prior findings [9, 10, 11]. This is in line, however, with their status as majority speakers of the analysed language.

Further evidence for this is provided in the analysis of segmental form and duration. Againn no group difference was found. FPs are produced by ML speakers of English similarly to other words, as suggested by [4]. As words they follow prosodic considerations such as pre-boundary lengthening [27] and are lengthened in phrase-final position or if phrased separately. The differences in duration found for V and VN forms can be related to the respective number of segments. The HL background does not influence the use of FPs as words in ML English.

Overall the data analysed in this study replicates the VN preference of FPs in English as well as the gender difference of VN preference in female speakers [14]. This is further evidence for the linguistic and cultural dominance in the ML of HS. That is, HS seem to acquire sociolinguistic and pragmatic aspects of language use in their ML. The second hypothesis, therefore, also has to be rejected. However, the observed difference between genders is less strong in the speakers of English with Russian HL. It would be insightful to also investigate their FP behavior in their HL to see whether this is a trend within this group which translates to the HL as well or whether it results from individual variation.

Prosodically the FPs in the analysed data were produced in a reduced manner, with low F0 and predominantly falling intonation. Unexpectedly, FPs were also frequently produced with rising contours. While in the distribution there was no effect of language background, speakers with Russian HL produced significantly larger slopes, especially with rises. This could be related to the influence of Russian on the intonation of these speakers' English due to typological differences in the two languages [28] or a transfer of intonational patterns from other discourse particles in Russian (e.g. *nu* [20]). The third hypothesis could, therefore, be partially confirmed. Further analysis is necessary to see whether the intonation of FPs reflects yet another area of cross-linguistic influence as shown in prior studies on bilingual intonation (e.g. [29]). The surprisingly frequent occurrence of rising FPs overall could alternatively be connected to discourse functions which are specifically used in these types of narrations.

6. CONCLUSION AND OUTLOOK

This study investigated the productions of FPs in ML English speakers with and without HL background. In the analyzed data speaker groups differed in the prosodic realisation of FPs which could be related to a possible influence of the HL. These findings could also be related to different signalling functions of FPs [30] used to different degrees by different speakers. The similarity in frequency and segmental form supports the native speaker status of HS in their ML [18]. Future research will address the FP use in the HS use of FPs in their HL hoping to provide a broader picture on FP use in bilingual speech.

7. ACKNOWLEDGEMENTS

Thanks to the phonetics group at Kiel University for helpful feedback and to the RUEG group for support and information on using their corpus data.

8. REFERENCES

[1] E. Shriberg, "To 'errrr'is human: ecology and acoustics of speech disfluencies," *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 153–169, 2001.

[2] M. Corley and O. W. Stewart, "Hesitation disfluencies in spontaneous speech: The meaning of um," *Language and Linguistics Compass*, vol. 2, no. 4, pp. 589–602, 2008.

[3] W. J. Levelt, "Monitoring and self-repair in speech," *Cognition*, vol. 14, no. 1, pp. 41–104, 1983.

[4] H. H. Clark and J. E. FoxTree, "Using *uh* and *um* in spontaneous speaking," *Cognition*, vol. 84, pp. 73–111, 2002.

[5] G. Kjellmer, "Hesitation. in defence of ER and ERM," *English Studies*, vol. 84, no. 2, pp. 170–198, 2003.

[6] H. Maclay and C. E. Osgood, "Hesitation phenomena in spontaneous English speech," *Word*, vol. 15, no. 1, pp. 19–44, 1959.

[7] S. R. Rochester, "The significance of pauses in spontaneous speech," *Journal of Psycholinguistic Research*, vol. 2, pp. 51–81, 1973.

[8] M. Belz, *Die Phonetik von äh und ähm*. Springer Berlin Heidelberg, 2021.

[9] J. F. Kroll and T. H. Gollan, "Speech planning in two languages: What bilinguals tell us about language production," *The Oxford Handbook of Language Production*, pp. 165–181, 2013.

[10] N. H. De Jong, "Fluency in second language testing: Insights from different disciplines," *Language Assessment Quarterly*, vol. 15, no. 3, pp. 237–254, 2018.

[11] G. Gilquin, "Hesitation markers among EFL learners: Pragmatic deficiency or difference?" *Pragmatics and corpus linguistics: A mutualistic entente*, no. 2, pp. 119–149, 2008.

[12] M. M. de Boer and W. F. L. Heeren, "Cross-linguistic filled pause realization: The acoustics of "uh" and "um" in native Dutch and non-native English," *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3612–3622, 2020.

[13] M. Polinsky, *Heritage languages and their speakers*. Cambridge University Press, 2018, vol. 159.

[14] E. K. Acton, "On gender differences in the distribution of um and uh," *University of Pennsylvania Working Papers in Linguistics*, vol. 17, no. 2, 2011.

[15] O. Niebuhr and K. Fischer, "Do not hesitate!-unless you do it shortly or nasally: how the phonetics of filled pauses determine their subjective frequency and perceived speaker performance." in *Proc. of INTERSPEECH*, 2019, pp. 544–548.

[16] N. H. de Jong, "Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency," *International Review of Applied Linguistics in Language Teaching*, vol. 54, no. 2, pp. 113–132, 2016.

[17] S. Wong and V. Papp, "Transferability of non-lexical hesitation markers across languages: Evidence from te reo Māori-English Bilinguals," in *Proc. of the 26th IAFPA*, Huddersfield, 2018, pp. 35–36.

[18] H. Wiese, A. Alexiadou, S. Allen *et al.*, "Heritage speakers as part of the native language continuum," *Frontiers in Psychology*, vol. 12, 2022.

[19] S. Stepanova, "Some features of filled hesitation pauses in spontaneous Russian," in *Proceedings of ICPhS*, vol. 16, 2007, pp. 1325–1328.

[20] A. Kuosmanen, "On the relationship between the melodical structure and discourse functions of the particles nu and vot in Spontaneous Russian," in *ESCA Tutorial and Research Workshop (ETRW) on Dialogue and Prosody*, 1999.

[21] H. Wiese, A. Alexiadou, S. Allen *et al.*, "RUEG corpus," Apr. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.3765218

[22] H. Wiese, "Language situations: A method for capturing variation within speakers' repertoires," *Methods in Dialectology XVI*, vol. 59, pp. 105–117, 2020.

[23] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. version 6.2.03," http://www. praat. org/, 2021.

[24] Y. Xu and X. Sun, "Maximum speed of pitch change and how it may relate to speech," *The Journal of the Acoustical Society of America*, vol. 111, no. 3, pp. 1399–1413, 2002.

[25] N. P. Himmelmann, M. Sandler, J. Strunk, and V. Unterladstetter, "On the universality of intonational phrases: A cross-linguistic interrater study," *Phonology*, vol. 35, no. 2, pp. 207–245, 2018.

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "lme4: Linear mixed-effects models using eigen and s4," 2014. [Online]. Available: https://CRAN.R-project.org/package=lme4

[27] A. E. Turk and S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *Journal of Phonetics*, vol. 35, no. 4, pp. 445–472, 2007.

[28] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.

[29] Y. Zuban, T. Rathcke, and S. Zerbian, "Intonation of yes-no questions by heritage speakers of Russian," in *Proc. of 10th International Conference on Speech Prosody*, 2020, pp. 96–100.

[30] T. Pistor, "Prosodic universals in discourse particles," in *Proc. of Speech Prosody*, vol. 178, Boston, USA, 2016, pp. 869–872.