



A preliminary study on filler particles in Bulgarian Judeo-Spanish (Ladino)

Beeke Muhlack^{1,2}, Jürgen Trouvain¹, Mitko Sabev¹ and Bistra Andreeva¹

¹ Language Science and Technology, Saarland University, Saarbrücken (Germany)

trouvain|msabev|andreeva@lst.uni-saarland.de

²Institute for Empirical Linguistics, Goethe University Frankfurt (Germany)

b.muhlack@em.uni-frankfurt.de

Abstract

In this cross-linguistic study we investigate the vowel quality of filler particles (FPs) produced by Ladino-Bulgarian bilingual speakers compared to monolingual speakers of Bulgarian and Spanish, respectively. The FP vowel quality of the bilinguals in both languages is more similar to the FPs of the monolingual Bulgarian speakers, overlapping with realisations of /a/ and /ɤ/, and not with /e/ as produced in FPs by Spanish speakers. Our data suggests that the presence of a schwa-like central vowel in the inventory influences the FP vowel quality.

1. Introduction

In this preliminary study which is a small part of a larger language documentation project [9] we investigate the vowel quality of filler particles in Bulgarian Judeo-Spanish, also known as Ladino, a variety of Spanish spoken by less than 200 speakers today [9]. Judeo-Spanish varieties emerged when Sephardic Jews "exported" their Spanish varieties from Spain to new settlements at the end of the 15th century. Those varieties developed independently from Spanish and were influenced by surrounding languages such as Bulgarian in Bulgaria. The aim of this study is to examine if FPs in Ladino share a similarity with Spanish FPs or if they are influenced by the Bulgarian contact language to the degree that they show no differences to the FPs of Bulgarian monolingual FPs.

FPs are often labelled as hesitations or disfluencies, but they can actually serve different functions, e.g. structuring discourse [11], turn-taking [2] or seeking attention [16]. Segmentally, they consist of similar forms across languages, which can be categorised into three typical forms: (1) a vocalic form (*uh*) consisting of a vowel only, (2) a nasal form (*hm*) consisting of a (bilabial) nasal consonant and (3) a vocalic-nasal type (*um*) that is a combination of the two previous segments. Differences across languages in the use of FPs mainly involve the preferred type of FP and the different vowel quality produced in the FP [2]. Previous research suggests that the preferred type of FP may be related to the most frequent syllable type [12]. For example, closed syllables are more common than open syllables in English [3] and the preferred FP type in English is the vocalic-nasal type *um*, which is a closed syllable itself. In Spanish, however, open syllables are more frequent than closed syllables [6] which may be the reason for the preference of the vocalic FP type *uh*, i.e. an FP type that is an open syllable itself. Speakers that are more familiar with one syllable type seem to prefer this syllable type (possibly unconsciously) in their hesitation device.

In many languages, the vowel quality of FPs takes the form of a central vowel, e.g., in German, English, Arabic, Dutch, Hungarian [4, 7, 12, 13, 14] and possibly many others. In

contrast, Spanish FPs employ an unrounded close-mid front vowel similar to realisations of the lexical vowel /e/ [12]. In comparison to other languages, this vowel quality seems quite unique. A possible reason for this may be the lack of a mid-central vowel in the Spanish vowel inventory and the high frequency of the phoneme /e/ [8]. As a Spanish variety, Ladino has a vowel inventory that is similar to Spanish (/i e a o u/ [6]). Bulgarian has a vowel inventory of 6 vowels /i e a ɤ u/ in stressed position [9, 15]. While there is no schwa-like vowel in Spanish [6], Ladino adopted the full mid-central vowel /ɤ/ from Bulgarian through the incorporation of Bulgarian loan words [1, 10]. It is hypothesised that the vowel inventory, and the existence of a mid-central vowel in particular, influences the vowel quality that is used in FPs.

Investigating the Spanish variety Ladino in the context of Bulgarian seems beneficial for exploring this hypothesis, as this variety represents a language that does not include a mid-central vowel in their original vowel inventory, but which familiarised itself with this vowel quality through the context of Bulgarian. The use of a mid-central vowel quality in Ladino FPs would then support the hypothesis that the existence of schwa in a vowel inventory may, in fact, predict the vowel quality of FPs.

In order to explore this hypothesis, FPs in the Spanish variety Ladino are compared to those of Bulgarian and Spanish monolinguals. Since Ladino is a minority language in Bulgaria and no monolingual speakers of this variety exist anymore, the Ladino speakers are in fact bilingual speakers of Bulgarian and Ladino. This aspect will be addressed in the discussion.

There are three research questions regarding the bilingual speakers (Bulgarian and Ladino):

1. Do speakers differentiate between the FP vowel quality in their two languages?
2. Is the FP vowel quality of the speakers more similar to Bulgarian or more similar to Spanish monolingual speakers?
3. If the FP vowel quality is more similar to Bulgarian monolingual speakers, does it match the Bulgarian mid-central vowel /ɤ/ in stressed syllables?

2. Method

The inspected data contains three sets of speakers. First, a set of four female speakers (80-88 years) who were recorded in narrative interviews. All speakers are bilinguals of Bulgarian and Ladino but with Bulgarian as their dominant language. The total speaking time was 13 minutes per speakers (min/spk) for their Bulgarian data and 14 min/spk for Ladino. The second speaker set contains four monolingual female Bulgarian speakers (79-86 years, recorded also in narrative interviews) with 12 min/spk. The third speaker set are 10 female speakers of Spanish (19-21 years, i.e. not age-matched) from the Diapix-

FL corpus [5] with 8 min/spk. In total, there are four data sets: bilingual Bulgarian, bilingual Ladino, monolingual Bulgarian, monolingual Spanish.

In addition to the orthographic annotation, all speech signals of the data sets were annotated and segmented regarding their FPs as follows: FP type (*uh* or *um* or *hm*), and the vowel section in *uh* and *um*. The first two formants were calculated at the midpoint of the vowel of all FPs and lexical vowels.

3. Results

3.1. Bulgarian and Ladino of the bilingual speakers

Table 1 shows that *uh* is by far the most used FP type in both languages, Bulgarian and Ladino, of the bilingual speakers.

Table 1: Absolute numbers of FPs in both languages of the bilingual speakers.

Type	Bulgarian	Ladino
uh	131	236
um	29	28
hm	19	32

In order to answer the first research question, whether bilinguals differentiate in their FP vowel quality between the two languages (Fig. 1), two linear mixed models (one for F1, another for F2) were calculated: $\text{lmer}(f1 \sim \text{lang} + (1|\text{speaker}), \text{data} = \text{df})$ for the first formant and $\text{lmer}(f2 \sim \text{lang} + (1|\text{speaker}), \text{data} = \text{df})$ for the second formant. A significant effect was found for F1, however, the difference between groups is only 27 Hz (lower in Ladino) which does not represent a relevant difference. The difference of 49 Hz (rise in Ladino) for F2 does not reach statistical significance.

3.2. Bilingual vs. monolingual Bulgarian and Spanish

When comparing the FP vowel quality of the bilingual speakers with those of the Bulgarian and the Spanish monolingual speakers, to answer the second research question, the following picture appears. As can be seen in Fig. 2 the spread of the formants of both monolingual groups is much narrower than the ones by the bilinguals.

There is a slight overlap of the values of the Ladino speakers and the Spanish monolinguals (Pillai score between 0.6-0.8). However, the overlap is much larger between Ladino bilinguals and Bulgarian monolinguals (score: 0.02). The Pillai overlap scores between Bulgarian bilingual FPs and Bulgarian monolingual FPs is 0.06, and between Ladino bilingual FPs and Bulgarian bilingual FPs is 0.03. (All MANOVAs that were calculated result in a p-value below the significance threshold of 0.05.)

3.3. Bilingual FP vowel vs. central vowels in Bulgarian

Regarding research question 3, Figs. 3 and 4 show the F1/F2 scatter of Ladino and bilingual Bulgarian FP vowels, plotted as 2-dimensional kernel density distributions over the F1/F2 dimensions, compared to the monolingual Bulgarian corner vowels /i a u/ and the mid central /s/. All vowels were produced in stressed position. Comparing Bulgarian bilingual FPs with Bulgarian monolingual lexical vowels (Fig. 3) a Pillai score of 0.15 is reached for FP vs. /a/, and 0.29 for FP vs. /s/. Comparing Ladino (bilingual) FPs with Bulgarian monolingual lexical vowels the Pillai scores are 0.17 for FP vs. /a/, and 0.21 for FP

vs. /s/. All values represent a high degree of overlap of the FP-vowels with both central lexical vowels of Bulgarian monolinguals.

4. Discussion

Referring to the three research questions stated at the beginning, we can now give the following answers.

(1) *Do speakers differentiate between the FP vowel quality in their two languages?* No, the Ladino-Bulgarian bilinguals investigated here use the same vowel quality in their FPs in both languages.

(2) *Is the FP vowel quality of the bilingual speakers more similar to Bulgarian or more similar to Spanish monolingual speakers?* The FP vowel quality of the bilinguals is more similar to the Bulgarian FPs, as a central vowel is produced. This supports the hypothesis that the existence of a mid-central vowel in the vowel inventory of a language predicts the vowel quality produced in the FPs. However, the examined speakers are bilingual speakers of Bulgarian and Ladino, and in fact, are more dominant in Bulgarian. The status of Ladino may have become such of a second language as it is not as frequently used. The similarity of the FPs in both varieties may then be explained by the process of transfer from the first language to the less dominant second language.

(3) *If the FP vowel quality is more similar to Bulgarian monolingual speakers, does it match the Bulgarian central vowel?* The FP vowel quality spreads over the vowel space of Bulgarian /a/ and /s/ with a high degree of overlap with both vowels.

In terms of frequency, the bilingual speakers use more FPs in Ladino than in Bulgarian. A possible answer is that the bilingual speakers need more planning effort due to the fact that their Ladino is less dominant than their Bulgarian. This would support the second-language status of Ladino for our speakers.

5. Conclusion

This preliminary study made use of a rather limited data set with just four speakers. Although the findings seem to be rather promising the analysis of a larger Ladino data set is necessary. Nevertheless, the methods applied here could be applied to other languages in comparable contexts.

One question that emerges with this study is whether it is possible to identify one or more features in a language that can predict the FP vowel quality. It shows that the vowel inventory but more so the presence of a central vowel are good candidates for such a feature, but also the most frequent vowel phoneme may be a possible feature. Our data suggests that the presence of a schwa-like central vowel influences the FP vowel quality. Further research into other languages is required, covering languages with schwa as a phoneme and those without. Ideally, there would be no confounding factor of bilingualism or second-language proficiency, as the similarities between FPs by bilingual speakers may be due to transfer from their dominant/first to their less dominant/second language.

Acknowledgements: This research was funded by the DFG (project 491553503, ‘Judeo-Spanish in Bulgaria: a contact language between archaism and innovation’). We are grateful to Elena Kireva and Snezhina Dimitrova for recording the Bulgarian and Ladino sessions. We also thank two anonymous reviewers for their helpful comments.

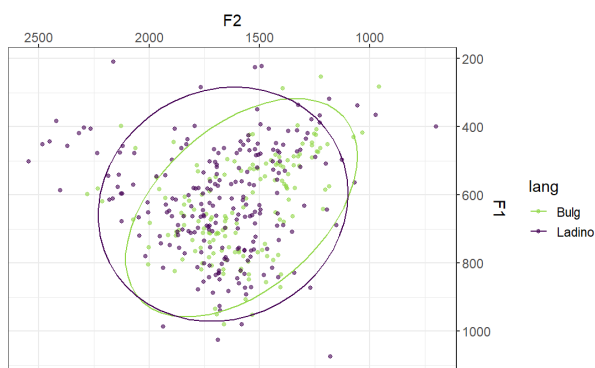


Figure 1: Values for first two formants of the FP vowels in the bilinguals' Bulgarian and Ladino in an F1-F2-plane in Hertz.

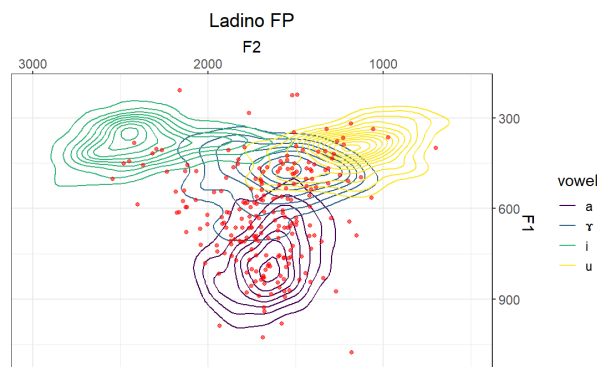


Figure 4: F1 and F2 values of the Ladino FP vowels of the bilinguals (in red) compared to [i a u ʏ] in monolingual Bulgarian.

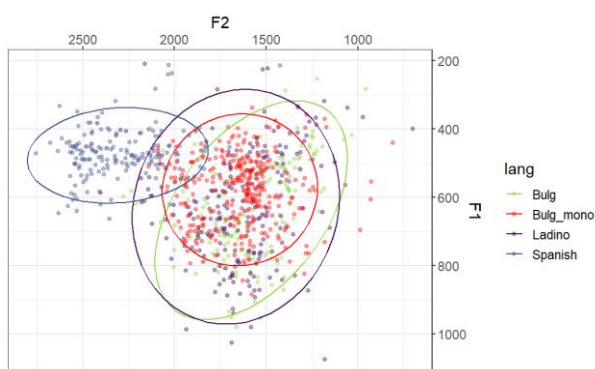


Figure 2: F1 and F2 values of the FP vowels of Bulgarian (green) and Ladino (purple) bilinguals, Spanish (blue) and Bulgarian monolinguals (red).

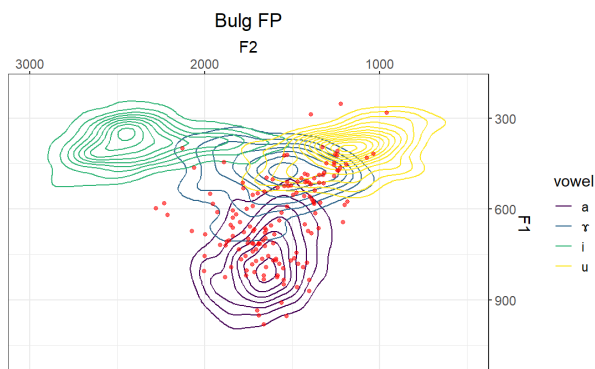


Figure 3: F1 and F2 values of the Bulgarian FP vowels of the bilinguals (in red) compared to [i a u ʏ] in monolingual Bulgarian.

6. References

- [1] Bradley, T. (2022). Judeo-Spanish. In: Gabriel, C., Gess, R., Meisenburg, T. (eds), Manual of Romance phonetics and phonology. De Gruyter, pp. 808–838.
- [2] Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition* 84, pp. 73–111.
- [3] Crystal & House (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88(1), pp. 101–112.
- [4] De Leeuw (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, 19(2), 85–114.
- [5] Lecumberri, MLG, Cooke, M & Wester, M (2017). A bi-directional task-based corpus of learners' conversational speech. *International Journal of Learner Corpus Research* 3(2), pp. 175-195.
- [6] Gabriel, C. (2022). *Phonetik und Phonologie des Spanischen*. In: R. Klabunde, W. Mihatsch, & S. Dipper (Eds.), *Linguistik im Sprachvergleich: Germanistik, Romanistik, Anglistik* (pp. 27–48). J.B. Metzler.
- [7] Gósy, M., Gyarmathy, D. & Beke, A. (2017). Phonetic analysis of filled pauses based on a Hungarian-English learner corpus. *International Journal of Learner Corpus Research*, 3(2), pp. 149–174.
- [8] Guirao, M., & García Jurado, M. A. (1990). Frequency of occurrence of phonemes in American Spanish. *Revue Québécoise de Linguistique*, 19, pp. 135–149.
- [9] Grünke, J., Sabev, M., Gabriel, C., & Andreeva, B. (2023). Vowel reduction in spontaneous Bulgarian Judeo-Spanish. *Proc. International Congress of Phonetic Sciences, Prague*, pp. 2844–2848.
- [10] Kanchev, I. (1975). *Fonética y fonología del judeoespañol de Bulgaria*. Ph. D. dissertation. University of Sofia.
- [11] Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word* 15(1), 19–44.
- [12] Muhlack (2023). Filler particles in English and Spanish L1 and L2 speech. *Proc. International Congress of Phonetic Sciences, Prague*, pp. 2418-2422.
- [13] Muhlack, B., Trouvain, J. & Jessen, M (2023). Distributional and Acoustic Characteristics of Filler Particles in German with Consideration of Forensic-Phonetic Aspects. *Languages*, 8(2), 100.
- [14] Muhlack (forthcoming). Filler particles: Phonetic details, cross-linguistic comparisons, and the recall effect. Ph.D. Thesis. Saarland University.
- [15] Sabev, M. & Andreeva, B. (2024). The acoustics of Contemporary Standard Bulgarian vowels: A corpus study. *Journal of the Acoustical Society of America*, 155(3), pp. 2128–2138.
- [16] Sadanobu, T. & Takubo, Y. (1993). The discourse management function of fillers: A case of "eeto" and "ano(o)". *International Symposium on Spoken Dialogue*, pp. 271–274.