How consistent are non-native speakers in their usage of filler particles when talking to native speakers?

Malte Belz, Miriam Müller, Christine Mooshammer

1Humboldt-Universität zu Berlin, Germany

Abstract

This paper investigates the intra-individual consistency in the production of filler particle frequencies and forms. The data consist of five non-native (L2) mid to high proficient speakers of German with English as first language. Each L2 speaker conversed with five different German native (L1) speakers in task-based dialogues. The results show that the L2 speakers are rather consistent in the use of FP frequencies and forms across the five dialogues, with no obvious alignment for static measures such as mean FP frequency or the proportion of vocalic-nasal and nasal FP forms. When tracing their relative FP frequencies over the course of the dialogue, some L2 speakers show convergence to L1, while others remain persistent. The data thus partly undergird the speaker-specificity of filler particles. It remains open whether observed convergence between L1 and L2 speakers in this data is due to alignment or happening coincidentally.

Index Terms: filler particles, filled pauses, phonetic convergence, alignment, intra-individual variation, L2 German, L1 German

1. Introduction

By filler particle (FP) we understand “phonetic exponents which are segmentally structured, semantically empty, syntactically unconstrained, and do not show an interjunctional function” [1]. We will not make any assertions about the pragmatic function of FPs. Inter-individually, speakers vary in their produced FP forms and frequencies [2, 3, 4, 5, 6]. Here, we focus on the intra-individual variation of L2 speakers talking to different L1 interlocutors using the same task.

For German, a large variety of FP forms have been observed, with V and VN forms being most frequent [7, 1]. Here we categorize the varieties in the following way. Vocalized FPs are phonetic exponents containing a perceivable structure (e.g., [r̃m], [mc], [E]) which can also be glottalized (e.g., [r]) [1]. Glottal FPs are forms that do not carry a perceivable (vocalic) structure but are merely glottal (such as [??]) [8]. Within vocalized FPs, speakers utter vocalic (V) forms, i.e., forms that are realized mainly with a vowel, vocalic-nasal (VN) forms, i.e., forms that are realized mainly with a vowel followed by a bilabial nasal, and nasal (N) forms, regardless of leading or trailing glottalization.

Speakers show individual preferences both for vocalized and glottal forms [1] as well as for the phonetic realizations of FPs [9], especially in their frequency of V and VN forms both within [2, 5] and across speaking styles [4, 3]. Speaker-specificity can be quantified by showing a low correlation between V and VN forms across speakers [5]. Another measure is the proportion of V and VN forms. In German, male speakers produce 1.1 and female speakers 1.3 times as many VN forms than V forms [7] across different speaking styles. For FP variation in German and English, [10] showed that both German and English speakers prefer VN over V forms, yet German speakers show less consistency, while English speakers tend to avoid V forms.

Potentially, these speaker-specific characteristics could be modulated by alignment or phonetic convergence [11, 12]. As has been found in several studies interlocutors adapt their speech to each other, therefore sounding more similar. Evidence for alignment has been found based on perceptual rating studies [11] and for acoustic parameters such as speech rate, fundamental frequency, and loudness (cf. [13] for an overview). Up to now very few studies look into the alignment of the usage of filler particles. One is the case study by Silber-Varod et al. [14] who found for a single dyad that speakers adapted the frequency of FPs by comparing the accumulative number of FPs over the course of two task. Pardo et al. [15] also found some evidence for convergence of FP frequency in several Map Task conversation but it strongly depended on the initial role assignment: Speakers assigned as givers in the first task increased the number of FP when assigned to the receiver role, therefore showing a path dependency for the occurrence of FP. For English, de Boer et al. [16] focus on the acoustic characteristics, showing that speakers remain consistent in their FP vowel forms over time. To the best of our knowledge, there are no studies on within-speaker alignment of FP frequency and form in L2 German.

In this study, the following research questions will be investigated:

1. How persistent is the within-speaker usage of filler particles regarding frequency and form for the same task?
2. Do speakers adapt towards their interlocutor regarding FP frequency or form?

2. Method

2.1. Corpus

We use the Corpus of Non-Native Addressee Register (CoN-NAR) [17]. CoN-NAR consists of 20 L1 German and four L2 German (L1 English) speakers of differing proficiencies. Each L2 speaker was recorded at five different sessions talking to five different L1 participants, thus establish four rounds. The corpus consists of read word lists, a free conversation (8 min), and two task-based conversations (8 min each) in which the interlocutors had to find 12 differences in modified pictures [18] of the Diapix picture task [19]. The sessions took place via a video-call connection (Zoom). Speakers were placed in the phonetics labs and an adjacent office and recorded via separate microphones and not via the Zoom software.

Here we use a subset of the corpus consisting only of the first Diapix.
2.2. Participants

20 L1 German speakers (10 male, 10 female) between 20 and 38 years of age and four L2 German (L1 English) speakers (2 male, 2 female) without hearing or speech impairment were recorded (cf. Table 1). L1 speakers are labeled as p01–p20, L2 speakers as c02, c04, c06, and c08, according to the corpus data. The intermediate to advanced proficiency levels of the L2 speakers according to the Common European Framework of Reference for Languages are B1, B2, and C1 [20].

Table 1: L2 speakers’ metadata.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Age</th>
<th>Sex</th>
<th>Proficiency</th>
<th>Education</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>c02</td>
<td>23</td>
<td>f</td>
<td>C1</td>
<td>baccalaureate</td>
<td>US</td>
</tr>
<tr>
<td>c04</td>
<td>23</td>
<td>m</td>
<td>C1</td>
<td>baccalaureate</td>
<td>UK</td>
</tr>
<tr>
<td>c06</td>
<td>24</td>
<td>f</td>
<td>B2</td>
<td>university</td>
<td>US</td>
</tr>
<tr>
<td>c08</td>
<td>27</td>
<td>m</td>
<td>B1</td>
<td>university</td>
<td>UK</td>
</tr>
</tbody>
</table>

2.3. Annotation

FPs were annotated on three tiers in Praat version 6.3 [21]. Two annotators were trained with the annotation guideline in [1]. On FP, filler particles were marked with FP, or, if an interjectional function could not be excluded, with FPC. On fp, vocalized FPs were marked with fv, glottal FPs with fg. On MAU, the phonetic segments and boundaries were corrected for glottal stops, vowels, and consonants.

2.4. Measurements

We built an EMU database [22] using emuR [23] in R [24]. The proportion of VN over V is quantified by \( \frac{V_N}{V} \). The relation of VN forms with regard to V forms is quantified by \( \frac{V_N}{V} \). To trace FP frequency over the conservation, we follow Silber-Varod et al. [14] and calculate \( \text{accum}_{FP} \) as the accumulated sum of filler particles per speaker, \( \text{accum}_{word} \) as the accumulated tokens per speaker (including FPs, but excluding silences, coughs, and unidentifiable passages), the relative accumulated frequency of FPs as \( \text{FP}_{rel} = \frac{\text{accum}_{FP}}{\text{accum}_{word}} \), and \( \text{FP}_{gap} \) as the differences of the median of \( \text{FP}_{rel} \) between the two interlocutors.

3. Results

In total, 764 instances are annotated as FPCs (n = 23, 3%) and FPs (n = 741, 97%). FPCs as well as nasal forms (n = 22) are excluded from the analysis, resulting in 719 instances. Glottal forms (fg) are less frequent (n = 50) than vocalized (fv) forms (n = 669), which split into 349 vocalic and 320 vocalic-nasal forms.

Table 2 summarizes the production of vocalized and glottal FPs as well as vocal and vocalic-nasal forms of the L2 speakers. FP frequencies are higher for less proficient speakers (i.e. c06 and c08), while c02 and c04 produce a comparable amount of FPs with a range of 1.8–2.2 FPs per hundred words.

Figure 1 shows the frequencies of vocalic filler particles per L1 and L2 speakers. The L2 speakers show some intra-individual variation but they do not seem to align their frequencies to those of L1 speakers. In most cases, the lines connecting the FP frequency of the L1 and L2 interlocutors exhibit a steep slope. For example, c08 (round 4) produced more FPs for interlocutors with a low occurrence of FPs. Furthermore, the lines are often intersecting, but for alignment non-intersecting, parallel lines would be expected. For example, c08 produced more FPs for interlocutors with a low occurrence of FPs. Only c06 (panel 3) shows evidence for alignment with four of the five interlocutors.

A closer inspection of the vocalized forms (fv) reveal major differences in the distribution of V and VN forms across the L2 speakers (cf. Table 3) which cannot be attributed to their proficiency levels of German. While c2 and c06 produce less V than VN forms (with a V/NV ratio of 0.9), c04 utters 5.1 times as many VN than V forms, with c08 ranging in between.

Figure 2 shows that there is no correlation between the use of V and VN forms across all speakers. Tests of Pearson’s product-moment correlation revealed no significant correlation, neither for the L1 group \((r = -0.66, p = 0.8)\) nor the L2 group \((r = 0.74, p = 0.26)\). Thus, we may conclude that our data show speaker-specific patterns of FP forms.

Figure 3 shows whether L2 speakers change their VN/V proportions in dialogues with different L1 interlocutors. Solid lines indicate interlocutors in the same session.

Figure 4, the relative accumulated frequencies of FPs (FP_{rel}), including both vocalized and glottal FPs, are plotted.
Table 3: FP mean and standard variation (sd) of vocalic (V) and vocalic-nasal (VN) FP forms, proportion of VN of V+VN (VN%), and VN-V-ratio instances (n), frequencies per hundred words (phw), instances of glottal FPs (fg), and instances of vocalic FPs (fv) of L2 speakers.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>¯V</th>
<th>sd</th>
<th>¯VN</th>
<th>sd</th>
<th>VN%</th>
<th>VN/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>c02</td>
<td>6.6</td>
<td>1.8</td>
<td>6.2</td>
<td>3.0</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>c04</td>
<td>1.8</td>
<td>1.1</td>
<td>9.2</td>
<td>3.4</td>
<td>0.8</td>
<td>5.1</td>
</tr>
<tr>
<td>c06</td>
<td>8.0</td>
<td>2.5</td>
<td>6.4</td>
<td>2.3</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>c08</td>
<td>13.0</td>
<td>2.7</td>
<td>21.4</td>
<td>7.5</td>
<td>0.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Figure 2: Vocalic-nasal (VN) and vocalic (V) instances per L1 interlocutors (absolute values) and L2 speakers (averaged). For all 20 dialogues on a normalized time scale in order to visualize convergence patterns that may emerge during the dialogue. Interestingly, many L2 speakers produce FPs very early in the dialogue, as evidenced by the high relative frequencies in the first 10–20% of the time scale. This pattern is not used frequently by L1 speakers.

From this visual interpretation, on the one hand, we see that c02 shows overlapping or close FP_rel trajectories with p02 and p04, c04 with p08 and p10, c06 with p11, p12, and p15, and c08 with p19 and p20. On the other hand, trajectories with visible gaps show rather divergent FP frequencies. Thus, the picture is rather inconsistent, as L2 speakers converge to the native speakers in some but not all of their dialogues. Convergence is quantified by low values of FP_gap (cf. [14]). However, there is no threshold determining convergence. Therefore, FP_gap is arranged in ascending order in Table 4. After checking with Figure 4, the left column could be considered as showing convergent traces of FP_rel over time, whereas sessions in the right column are rather dissimilar.

### 4. Discussion and Conclusion

We investigated four mid to high proficient L2 speakers of German, each conversing with five different German L1 interlocutors. Both their frequencies and form proportions are rather idiosyncratic. Over the course of the dialogue the relative FP frequency of the L2 speakers reflect their mean FP frequency production pattern. Nevertheless, in some cases we observe similar patterns between L1 and L2 speakers, but it remains unclear whether this is alignment. In terms of alignment between L1 and L2 speakers, Wagner et al. [25] found that Dutch native speakers are capable of aligning with recordings of a non-native speaker (L1 Serbo-Croatian). The level of convergence was affected by the perceived accentedness, with a stronger non-native accent leading to a decreased inclination to converge. As the L2 speakers in our study are mid to high proficient German speakers, no strong accentedness is expected. If the results of [25] can be transferred to FP frequency, we should have seen more alignment. Further, there are contradictory observations regarding the level of convergence in the context of the linguistic distance between two speakers. On the one hand, Kim et al. [26] found that speakers converge phonetically more to speakers with the same English dialect than to speakers with a different dialect or...
Figure 4: Relative accumulated frequencies of filler particles per dialogue session of L1 (p01–p20) and L2 speakers (c02–c08).

The alignment which has been observed could be due to the fact that the L1 speakers’ relative frequencies are coincidentally similar to those of the L2 speakers. To overcome this limitation, we are currently annotating the Diapix tasks of the L1 participants with the L1 confederates, which allows us to quantify the consistency of the L1 participants. If they behave consistently in the two sessions (L1–L1 and L1–L2), then changes of the L2 speakers towards the L1 speakers can be attributed to alignment. If the L1 speakers show consistent changes when conversing with the L2 speakers, this could be attributed to a register change, the so-called foreigner-directed speech [29] or non-native addressee register.

5. Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334. Many thanks to our annotator Torben Schilling and to Megumi Terade for comments.

6. References


[5] ——, “Profiling fluency: An analysis of individual variation in


24. R Core Team, “R: A language and environment for statistical computing,” Wien, 2023; R version 4.2.3.