Abstract

While the expression of speaker alignment has been of interest in Conversation Analysis for decades, the prosody of agreement has so far only received little attention. This is the first investigation of the timing of agreement and disagreement turns in conversational Austrian German and the prosody of prefaces to these turns. To assess the timing and prosody, we extracted the duration of the pause before the agreement shape, examined its F0 shape and performed linear and logistic regression analyses to estimate the relationship of these features with agreement. We find that agreement prefaces tend to be short and uttered after a short pause. Prefaces to straightforward disagreements are also produced after a short pause but tend to be long. Prefaces to a sub-category of disagreeing turns (“yes you’re right, but…”) tend to be long and uttered after a long pause. Various F0 shapes occur in all three types of prefaces, but there is a robust relationship between dip-rises and prefaces to straightforward disagreements. While the timing patterns are in line with previous findings in the literature on speaker alignment, we did not find unique F0 patterns for agreement or disagreement.

Index Terms: speaker alignment, agreement & disagreement, prosody, conversational speech, Austrian German

1. Introduction

The expression of agreement and disagreement in conversations has been investigated in the tradition of Conversation Analysis (e.g., [1-3]), but prosody is an aspect that has received little attention in this field. Several phonetic studies, however, have shown that prosody contributes to create meaning beyond the propositional meaning of an utterance. Viní et al. [4], for instance, have shown that the Czech discourse marker jasné may have the pragmatic functions of resignation, reassurance, surprise, or impatience, depending on the word’s prosody.

The present study is the first quantitative investigation of the F0 shape and the timing of “ja” (yes) and “nein” (no) prefaces to agreeing and disagreeing turns in conversational Austrian German. In contrast to more specific expressions such as “genau” (exactly) and “stimmt” (right), which are mostly used to exclusively express agreement, utterances prefaced by “ja” and “nein” may both express either agreement or disagreement. Therefore, these words are particularly suited for this investigation, as the same lexical content can have different communicative functions. Table 1 gives examples of an agreement (AG) prefaced by “nein” and a disagreement (DG) prefaced by “ja”.

Table 1. Examples of agreement (AG), disagreement (DG) and an agree+disagree sequence (AG+DG) from the GRASS corpus [7]. Prefaces are marked in bold.

<table>
<thead>
<tr>
<th>German</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AG</strong></td>
<td></td>
</tr>
<tr>
<td>010M: Da muss sogar die Grafik Mist sein eigentlich So even the graphics are probably garbage</td>
<td></td>
</tr>
<tr>
<td>009M: Nein sie ist eh nicht gut No they’re really not good</td>
<td></td>
</tr>
<tr>
<td><strong>DG</strong></td>
<td></td>
</tr>
<tr>
<td>008M: Aber grad Kaffee ist glaube ich gut gegen Kater But I think especially coffee is good for hangovers</td>
<td></td>
</tr>
<tr>
<td>028F: Nein No</td>
<td></td>
</tr>
<tr>
<td>008M: Ja sicher, die Antioxidanzien Yes sure, the antioxidants</td>
<td></td>
</tr>
<tr>
<td><strong>AG+DG</strong></td>
<td></td>
</tr>
<tr>
<td>004M: Naja er kann schon König werden Well he can become king</td>
<td></td>
</tr>
<tr>
<td>024F: Ja aber er wird es wahrscheinlich nicht Yes but he probably won’t</td>
<td></td>
</tr>
</tbody>
</table>

Ogden [8] investigated the phonetics of (dis)agreements in American and British English and found that straightforward agreements as well as disagreements had ‘upgraded’ phonetics, that is, an expanded pitch span, more dynamic and overall higher pitch and slower tempo. Both agreements and disagreements were also timed without delay. However, the explicit disagreements in his data often consisted of reassertions of a speaker’s prior stance, which could be framed as an argumentation context. Agreement prefaces to disagreements (AG+DG) were uttered with delay and had a ‘downgraded’ prosody, that is, narrower pitch span, faster tempo and decreased loudness.

For German, Koester [9] postulated the importance of high vs. mid vs. low key. He argued that the kind of pitch pattern used depends on whether the stance expressed is epistemic or affective. When expressing an epistemic stance, he argues, there is no need to mitigate a disagreement and speakers use
falling tone, while they use rising and flat tone for polite disagreement. For turn prefaces, he noted the absence of fall-rise at the beginning of a disagreement sequence and a flat intonation in a token agreement before a disagreement (corresponding to AG+DG). In the cited examples, prefaces to agreements were associated with a fall, a fall-rise or a flat tone, and prefaces to disagreements with a fall or a flat tone.

Besides these qualitative investigations, two small quantitative studies investigated the prosody of ‘yeah’ tokens in American English [10] and of ‘ok’ tokens in Swedish [11]. Freeman et al. [10] found that ‘yeah but’ was characterized by a rise-fall, reluctance by a falling-rising F0, and agreements to negative assessments by a lower, flat F0. Forsberg & Abelin [11] found a relationship between agreement and falls on the one hand, and doubt and rises, on the other hand, for ‘ok’ tokens in Swedish.

Since these studies mostly investigated languages other than German ([8,10,11]) or focused on the prosody of whole turns instead of only prefaces ([8-9]), their results are heterogeneous, and they do not allow us to formulate clear hypotheses. Hence, this is predominantly an exploratory investigation.

2. Methods

2.1 Data and annotation

Six one hour-long spontaneous conversations between friends or partners (3 m-m, 3 m-f; 12 speakers) from GRASS [7] were annotated for three (dis)agreement categories in an iterative consensus annotation (4 annotators): prefaces to agreement (AG), agreement prefaces to disagreements (AG+DG) and prefaces to disagreements (DG; cf., Table 1). No task was given and the conversations were read aloud in a casual speaking style across a continuum of Standard Austrian German and various dialects, mostly from the southeast of Austria.

The annotation yielded 347 preface tokens: 224 AG tokens (180 [ja], 44 [nein]), 36 AG+DG tokens (all [ja] and 87 DG tokens (21 [ja], 66 [nein]), [ja] tokens were pronounced as [ja] or [ja] or a variant in-between. Most [nein] tokens were pronounced as some version of [na]; 6 [na] tokens were excluded due to the different syllabic structure. The smaller number of disagreements is in line with other studies (e.g., [8-9]).

2.2 Duration and F0 features

We segmented the target word and measured its duration (Worddur). We annotated three labels to capture the timing to the previous turn: 1) when a silent pause preceded the (dis)agreement (labelled as PAUSE), 2) when no pause duration could be measured because the (dis)agreement did not come directly after the utterance it aligns with (due to laughter, clarification questions, etc.; labelled as COMPLEX), and 3) when prefaces were uttered in overlap with the first turn (labelled as OVERLAP). For PAUSE tokens, pause duration (Pausedur) was measured. We decided against using negative pause durations in tokens labelled as OVERLAP because the turn taking was often too complex to be captured by a duration measure.

F0 was extracted with REAPER [12]. This algorithm was chosen because it tracks F0 more reliably in creak [13], which occurs frequently in our data. Octave jumps were corrected manually. In 28 tokens, F0 could not be corrected, resulting in a data set of 319 tokens for the F0 analysis. F0 was speaker normalized by converting Hertz to semitones based on the individual speaker median during the whole conversation. We calculated six F0 features: F0dur, F0dur, position of F0dur and F0dur, F0range and F0mean.

We categorized all 319 tokens into six shapes, according to the F0 range and the relative position of F0dur and F0dur with respect to the three thirds of the word duration (cf., Figure 1). We categorized a token as flat if the F0range was below 2.8 ST. This is an empirical value that separates perceptually flat tokens from ‘proper’ contours.

![Figure 1. Schematic representation of the six F0 shapes, based on the relative position of the F0dur and F0dur in the initial, medial and final third of the word.](image)

2.3 Statistical methods

We performed the statistical analysis on three datasets: the whole data set for Worddur (N = 347), a subset for F0 (N = 319), and another subset for Pausedur (N = 197; cf., 3.1). A data set for which all measures are available would have been very small (N = 187).

First, we built linear mixed effects regression models in R [14] with lmer [15], in which Pausedur, Worddur and individual F0 features (cf. Section 2.2) were the dependent variables. The models included the independent variables Category (AG, AG+DG, DG) and Token (‘ja’, ‘nein’), their two-way interaction, and the random effect Speaker. We reduced the models with the step() function [16]. To assess differences between the three categories, we performed pairwise comparisons with emmeans() [17].

Second, to assess the relationship between F0 shapes and the three (dis)agreement categories, we built logistic regression models with multinom() [18], with Category as dependent variable and Shape and Worddur as predictor variables, including their interaction. A stepwise backwards regression with the step() function resulted in a model without the interaction of Shape and Worddur. To observe all comparisons, we rotated the reference levels of Category (AG, AG+DG, DG) and Shape (flat, fall, medial peak, medial dip, dip-rise, rise) in these models. Here, we only present the significance levels of the predictors of interest for our analysis. The complete outputs for all models can be found at https://online.uni-graz.at/kfu_online/wbForschungsportal.ebShowPortal?pPersoN=93597&pMode=E. We did not include Token in the logistic models because of the large amount of ‘ja’ in AG (81%) and AG+DG (100%) and ‘nein’ in DG (76%), respectively, which would have taken away explanatory power from the features we are interested in.

3. Results & Discussion

3.1 Timing to previous turn

Table 2 presents the percentage of pauses (PAUSE), overlaps (OVERLAP) and cases in which no pause could be measured (COMPLEX; cf., Section 2.2) in each of the three categories. The
majority of tokens was produced after a pause in all three categories. However, AGs were also relatively often produced in overlap, which is in line with previous studies [1,8] In DG, the percentage of complex timing was relatively frequent, for various reasons. One reason was to delay a disagreement by asking a short clarification question before disagreeing (cf., [1]). Sometimes, speaker 1 also added an increment to a first assessment when an agreement was not immediately forthcoming (cf., [3,6]). Moreover, when laughter occurred before a disagreement, this could be interpreted as mitigating any tension that might be related to disagreeing with one’s interlocutor. AG+DG tokens had the largest percentage of pauses and the smallest percentage of overlaps (cf., [1]).

<table>
<thead>
<tr>
<th>Category</th>
<th>AG</th>
<th>AG+DG</th>
<th>DG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause</td>
<td>128 (57%)</td>
<td>25 (69.5%)</td>
<td>44 (51%)</td>
</tr>
<tr>
<td>Overlap</td>
<td>58 (26%)</td>
<td>3 (8.5%)</td>
<td>14 (16%)</td>
</tr>
<tr>
<td>Complex</td>
<td>38 (17%)</td>
<td>8 (22%)</td>
<td>29 (33%)</td>
</tr>
</tbody>
</table>

The linear regression model for \( \text{Pause}_{\text{Dur}} \) was built only on the tokens labelled as \( \text{Pause} \) (N = 197). The final model (cf., 2.3) contained only \( \text{Category} \) and \( \text{Speaker} \) as predictors of \( \text{Pause}_{\text{Dur}} \). AG+DG tended to have longer pauses than AG (AG vs. AG+DG: \( \beta = -0.346, t = -2.95, p = .0099 \)) and DG (AG+DG vs. DG: \( \beta = 0.324, t = 2.47, p = .037 \)), but the comparison of AG vs. DG was not significant. Figure 2 (left) presents violin and box plots per \( \text{Category} \) for \( \text{Pause}_{\text{Dur}} \).

Figure 2. Violin and box plots of pause duration (left; \( N = 347 \)) and word duration (right; \( N = 197 \)).

3.2 Word duration

The final model for word duration contained only \( \text{Category} \) as predictor of \( \text{Word}_{\text{Dur}} \). AG tended to have a shorter word duration than AG+DG (AG vs. AG+DG: \( \beta = -0.078, t = -3.96, p = .0003 \)) and DG (AG vs. DG: \( \beta = -0.049, t = -3.84, p = .0016 \)), but the comparison of AG+DG vs. DG was not significant. Figure 2 (right) presents violin and box plots per category for word duration. The longer word duration in the two disagreement categories could be due to hesitation lengthening (cf., [19]), which could be interpreted as expressing reluctance, but could also be due to emphasis lengthening, which might be the case in at least some DG tokens. AG tokens might be shorter because they carry less functional load than in AG+DG, in which the \(<ja>\) prefix is often the only indicator for the agreement part (cf., AG+DG in Table 1).

3.3 F0

For individual F0 features (cf., Section 2.2), none of the linear regression models showed any significant differences between the categories. Moreover, we did not find that key, represented by \( \text{F0}_{\text{shoc}} \), was an important factor in our data as suggested by Koester [9], nor did we find an expanded pitch range in agreements, like Ogden [8]. However, these authors studied whole (dis)agreements and not just prefaces.

Figure 3 shows the distribution of shapes across the categories. All shapes occurred in all three categories, except for \( \text{medial peaks} \) (yellow) in AG+DG. In all categories, tokens with a flat F0 (grey) were more frequent than the other shapes. In the disagreeing turns (AG+DG and DG), \( \text{rises} \) (dark blue) and dip-rises (light blue) were more common than in AG. About half of the DG tokens had a final-rising contour (a dip and/or a rise). Dip-rises (light blue) were significantly more likely to be produced in DG than AG, compared to flat F0 (p = .020; grey), falls (p = .085; orange) and medial peaks (p = .022; yellow), which were more likely in AG than DG. Dip-rises were also marginally significantly more likely in DG than AG+DG, compared to flat curves (p = .067), which were more likely in AG+DG than DG. Medial dips (green) were marginally significantly more likely in DG than AG, compared to medial peaks (p = .085), which were more likely in AG than DG. Medial dips were also marginally significantly more likely in AG+DG than AG, compared to falls (p = .075), which were more likely in AG than AG+DG.

The relationship between \( \text{Category} \) and F0 shape was not as clear in our data as in quantitative studies on English [10] or Swedish [11] since in our data, most shapes occurred in all three categories. Our results are only partly in line with the only study of German we are aware of [9]. Like [9], we also found falling and flat contours in agreements and disagreements, among other shapes, as well as a high percentage of flat contours in AG+DG. However, while the most consistent relationship in our data was between DG and a dip-rise, [9] noted that this
contour was absent from disagreement prefaces in his data. The high percentage of dip-rises and other dips in DG are, however, in line with what Freeman et al. [10] found for reluctant ‘yeah’. Thus, dip-rises and other dips in our data, particularly in DG, could be interpreted as expressing reluctance. Alternatively, dip rises could also be interpreted as contradictions, like Goodhue & Wagner [20] found for contradicting prefaces in English. While our results indicate that there is a connection between DG and dip-rise, there is no one-to-one relationship between a specific contour and an agreement or a disagreement.

3.4 Relationship of F0 shapes and durations

To assess whether there might be concord or a trade-off between intonation and timing in the expression of (dis)agreement, we investigate their relationship in this section.

We built a linear regression model with \( \text{Words}_{\text{dur}} \) as dependent variable and Shape and Category, as well as their interaction, as predictors, and Speaker as a random factor. Figure 4 presents boxplots of \( \text{Words}_{\text{dur}} \) per Shape for the three categories. In AG+DG, dip-rises had the longest word duration (significantly longer than falls (\( p = .036 \)) and flat F0 (\( p = .019 \)). In DG, dip-rises also had the longest duration (longer than: flat F0 (\( p = .0001 \)), falls (\( p = .001 \)), medial peaks (\( p = .039 \)), medial dips (\( p = .093 \)), and rises (\( p = .020 \)). Dip-rises in DG and AG+DG were also significantly longer than dip-rises in AG (AG vs. AG+DG \( p = .023 \), AG vs. DG \( p = .004 \)). This indicates that the longer duration of dip-rises in these two categories is not just a consequence of the more complex pitch shape, but one of the strategies that speakers use when they express disagreement, whether in a straightforward way (DG) or in the format of agree+disagree.

It is possible that there is a trade-off between F0 shape and pause duration. For instance, if a disagreement in the format AG+DG is not preceded by a long pause, it might be marked by a specific F0 shape to express reluctance. Therefore, we built a linear regression model predicting \( \text{Pause}_{\text{dur}} \) with an interaction of Shape and Category, with Speaker as a random effect. None of the pairwise comparisons were significant. This may indicate that there is no trade-off between our data shape and pause duration, or that the data set was too small (cf., Section 2.3).

4. Conclusions

In this study, we investigated pause, word duration and F0 features of prefaces to agreements and disagreements in Austrian German conversational speech. Overall, we found clearer tendencies for pause and word durations than for F0. Moreover, we found no correlation between individual F0 features and (dis)agreement.

In the conversations examined in our study, agreement prefaces tended to be short, were often flat or uttered with a fall or a medial peak. They tended to be timed shortly after the previous utterance or often even in overlap. Agreement prefaces to disagreements tended to be long, were often uttered with a flat F0 or a dip and tended to be timed with a longer delay to the previous turn. Disagreement prefaces also tended to be long, were often uttered with a dip-rise, but tended to be produced with a shorter pause or with short turns inserted between the first turn and the disagreement. The prevalence of rising contours in disagreement prefaces points to their forward-looking nature and dips could be interpreted as expressing contrast.

All characteristics in agreement prefaces point to directness/immediacy, and all characteristics in AG+DG (turn design, timing and prevalence of flat contours) point towards reluctance/politeness. The results for disagreement prefaces, however, are not as clear. A short pause indicates directness, but there was also a high number of disagreements in which delaying strategies were employed (e.g., inserting a clarification question or laughter). Long word durations could result from hesitation lengthening, possibly indicating reluctance, or from emphasis of the disagreement preface, which is more in line with being direct. The fact that the agreement prefaces to disagreements are a more consistent group in our data than straightforward disagreement prefaces is not surprising since the former is a subgroup of disagreements. Disagreement prefaces, however, occurred in a variety of disagreement contexts, which could affect timing and intonation. It has, for example, been shown that timing is related to the turn format, for instance if an utterance contains an explanation [21]. Disagreements in response to accusations and other kinds of “other-attributions” (e.g., “you liked it, didn’t you”) often show characteristics of typical agreements (e.g., immediate timing) rather than disagreements [6]. Furthermore, the usual behaviour of agreements and disagreements is typically reversed in argumentation contexts [3,5]. A more detailed analysis of the specific contexts in which disagreements were uttered could reveal other subgroups in the disagreeing turns in our data.

A first glimpse into the distribution of features over speakers showed that no two speakers behave alike, and that no speaker showed a clear preference for one strategy over another. A detailed analysis of speakers could reveal a catalogue of strategies speakers have at their disposal. For instance, even though we did not find any trade-off effects in this study, it could well be that individual speakers use combinations of features as strategies to express agreement and disagreement in a specific situation.

A perception experiment testing prefaces with different timings and prosodic shapes could shed some light on the importance of individual features or their combinations and on how listeners construe the projection of a (dis)agreement.

5. Acknowledgements

The work of A. Kelterer and S. Wepner was funded by the FWF grant P-32700-N. We would like to thank Julian Linke for his help with the extraction of durations.
6. References


