Universität zu Köln

Philosophische Fakultät

Institut für Linguistik – Phonetik

Bachelorarbeit zum Thema:

The intonation of feedback signals in the first and second language

Von:
Eduardo Möking

emoeking@smail.uni-koeln.de

Matrikelnummer: 7335126
Fachsemester: 8
2-Fach-Bachelor: Linguistik und Phonetik / Philosophie

Abgabe am: 30.09.2021

Erstgutachterin: Prof. Dr. Martine Grice

Köln, SoSe 2021

Contents

Abstract

The present explorative study is concerned with the intonation of minimal utterances, such as backchannels and short answers, in learners of a second language (L2). Backchannels are lexical and non-lexical utterances ('mm-hm', 'yes') signalling the speaker's comprehension and acknowledgement of what his interlocutor is saying. In this study, the performance of Italian learners of German and native German (L1) speakers was analysed and compared to identify differences and similarities regarding both group's minimal utterance production. Participants were recorded during a collaborative goal-oriented task. In the analysis, two sub-classes of minimal utterances were distinguished: backchannels and short answers to yes-no- and tag questions (response tokens). In addition, backchannels were further separated into acknowledgment tokens marking passive recipiency (PR) and acknowledgement tokens marking incipient speakership (IS). The minimal utterance production of both groups was analysed to determine the overall rate of backchannels and response tokens and their intonation patterns in relation to different pragmatic functions. The backchanneling behaviour of the L1 and L2 groups was compared to assess to what extent learners approach the target language across two proficiency levels (beginners and advanced). The more significant differences between the two groups were found with regard to the choice of tokens, while intonation patterns across different pragmatic functions were mostly similar. Proficiency did not have a noticeable effect on token choice or minimal utterance intonation. Results suggest that individual variability is a more significant factor.

1 Introduction

One of the central issues in the study of second language acquisition (SLA) has been the question of how to define and measure the various aspects of language proficiency, while taking into account individual variability of speakers and environmental factors impacting L2 performance in the learning process. Looking at oral performance in particular, fluency has been widely recognized as one of the central aspects in the assessment of L2 proficiency, e.g. by the Common European Framework of Reference for Languages. Despite the increased attention on fluency

in SLA research in recent years, the question remains about how to investigate fluency. One problem raised by van Os et al. (2020) and Tavakoli (2016) among others, is that research has mainly focused on fluency in monologues rather than as part of interactions, despite the fact that natural communication mostly takes place in the form of conversations. Fluency in dialogues is a complex phenomenon, to which two interlocutors contribute in a system of alternating turns. This complexity might be the reason for the emphasis on fluency in monologues in research, but it is also the reason why the parameters used to determine fluency in monologues are insufficient for the assessment of fluency in dialogues. Since two participants are involved in managing a successful conversation, turn-taking has to be taken into account when trying to understand fluency in dialogues. But the question of how co-created fluency can and should be operationalized, is an area within SLA research that is yet to be comprehensively explored.

A crucial aspect in the system of turn-taking in dialogues is the production of intonational cues, which can signal to the interlocutor the speaker's intention to carry on with his speech or that his turn is coming to an end. Intonation allows for the interlocutor to predict the end of the speaker's turn and thus simultaneously plan the beginning of his own turn, which results in a relatively short gap between the interlocutors' turns (Levinson & Torreira 2015) that has shown to have cross-cultural mean length of around 200ms (Stivers et al. 2009).

One way for a speaker to prosodically signal his intention to take a turn or pass the opportunity on to the interlocutor is through the use of feedback signals, including backchannels. Previous studies have found that backchannels contribute to understanding and facilitate the conveyance of information while at the same being potentially misleading, especially in cross-cultural contexts where differences between culturally shaped turn-taking systems may come into play (Berry 1994; Ha et al. 2016; Li 2006). There are language-specific norms concerning length of backchannel utterances, intonation and frequency of use, and it can be assumed that, if properly followed, these utterances may ultimately result in greater conversational fluency. If, for instance, particular intonational forms in backchannels are linked to specific conversational moves, such as a speaker's intention to begin a turn as opposed to just acknowledging the interlocutor's utterance, then it is crucial to understand the properties and functions of these intonational forms in order to prevent potential miscommunication in cross-cultural contexts.

Therefore, in the present work I intend to compare the backchannel productions of L1 German speakers to those of Italian learners of German. In particular, I will analyse the backchannel intonation contours produced by both groups and look to determine potential differences and similarities. The aim of this analysis is to investigate whether different patterns emerge between the two groups which are specific to the learners' interlanguage or the possible outcome of a transfer phenomenon. The scope of the present study, however, does not include the development of specific predictions about transfer effects, since the Italian speakers' L1 performance will not be examined. Nevertheless, I will provide an outline for potential future research based on the findings of this study.

Since backchannel intonation is not explicitly taught in L2 classroom settings, there are two likely scenarios to be expected. One the one hand, it is possible to assume that there is a transfer of backchanneling behaviour, including intonation patterns, from the L1 to the L2, due to the marginal attention backchannels and their intonation receive in foreign language teaching. On the other hand, backchannels might be regarded by learners of an L2 as an important means to know whether they are being understood during a conversation. As such, backchannels and their pragmatically relevant intonation patterns would receive an appropriate level of attention, which would favour an adaptation to target language patterns in learners of an L2. In the latter case a more target-like backchannel behaviour should be observed among advanced L2 speakers.

1.1 Background

Backchannels are commonly referred to as minimal utterances used by a listener to demonstrate understanding and acknowledgement of the interlocutor's speech (Clancy et al. 1996, Schegloff 1982). These utterances include lexical and non-lexical tokens like 'mm-hm', 'yes' and 'okay'. The term 'backchannel communication' was first proposed by Yngve (1970), who referred to all kinds of short utterances serving the purpose of signalling acknowledgement and understanding. Schegloff (1982) later noted, however, that tokens such as 'uh huh' were used mainly in contexts where "an extended unit of speech is underway" by the primary speaker, while the secondary speaker signalled with the use of backchannels not on-

ly understanding of what is being said, but also his wish for the primary speaker to carry on with his turn. Schegloff therefore termed utterances used in such contexts 'continuers' (1982: 81). A standard example of a scenario in which continuers apply, is a conversation in which one person gives instructions to another. The person receiving the instructions might then use continuers to acknowledge the instructions and signal that he wants the instruction giver to *continue*.

Jefferson (1984) later introduced the term 'acknowledgement tokens' to refer to utterances fulfilling a similar function. And Drummond and Hopper (1993) further distinguished acknowledgement tokens marking 'passive recipiency' (as in the case of continuers) from short utterances signalling a listener's intention to start a turn of his own, which they called acknowledge tokens marking 'incipient speakership'.

Around the same time research began to consider variation in the use of backchannels across languages and cultures, with the focus shifting toward the question of how this variation can potentially impede successful communication. The role of backchannels in cross-cultural communication has been investigated in a number of studies, with authors pointing out that turn-taking systems vary across languages, which may lead to miscommunication and misunderstandings in cross-cultural settings. For instance, Berry (1994) found that in a comparison of turn-taking styles in Spanish and North-American English not only the frequency of backchannels was higher among Spanish speakers, but also the backchannel utterances were longer, resulting in longer stretches of overlapping speech. Another study, conducted with Canadian and Chinese Speakers, suggested that backchannel responses can be misleading and cause miscommunication in inter-cultural conversations (Li 2006). According to the study's results, backchannels helped facilitate communication among speakers of the same language, while in two inter-cultural groups, where Canadian speakers were paired with Chinese speakers, the opposite effect was observed.

In a study comparing backchannel intonation patterns of Vietnamese and German speakers (Ha et al. 2016) it was found that in Vietnamese the majority of continuers are flat or falling in intonation, whereas in German equivalent tokens tend to have rising intonation. A perception experiment linked to this study revealed that Vietnamese speakers interpreted rising intonation as impolite. German

speakers on the other hand, had difficulties interpreting the pragmatic meaning of continuers with flat or falling intonation or even perceived them as impolite.

These findings have significant implications for the role backchannels play in communication in general, but also in particular for communication in cross-cultural settings and the relevance backchannels should be assigned to in language-teaching environments. Firstly, as suggested by the studies presented above, there is evidence that backchannels positively contribute to communication, for instance by facilitating floor transfer processes. Secondly, backchannel use reflects language-specific features (e.g. intonation patterns, frequency of use, duration and token choice) which may negatively impact cross-cultural communication. Also, in order to better understand the mechanism behind cross-cultural backchannel 'misbehaviour', it is important to shed light on how learners of a second language potentially transfer the culture-specific backchannel behaviour of their native language to the L2.

So far, only a relatively small subset of studies has considered backchannels in the L2, potential transfer effects and their communicative consequences (Cutrone 2011; Shelley and Gonzalez 2013; Wehrle and Grice 2019). With regard to these aspects, there seems to be a consensus that, even though backchannels are a universal phenomenon, particular backchanneling behaviour varies across languages (with regard to pragmatic function, frequency, intonation and token choice). And there is evidence that this behaviour carries over to the L2. In an experiment comparing the function and prosodic form of backchannels in L1 and L2 German, Vietnamese speakers of German produced twice as many non-lexical backchannels (e.g. 'mm-hm') with a *flat* intonation contour than native German speakers, resembling the general rule for flat or falling backchannels in Vietnamese (Wehrle and Grice 2019).

Another study by Castello & Gesuato (2019) investigated the frequency and lexical types of backchannels in Chinese, Indian and Italian learners of English in a language examination setting. They defined backchannels as 'expressions of convergence' and found that Chinese learners used the most backchannels, while L1 Indian speakers used them the least, with Italian L1 students coming in between the two other groups. Galaczi (2014) compared the frequency of backchannels and ex-

pressions of confirmation ('yeah', 'exactly') among language learners with proficiency levels ranging from B1 (intermediate) to C2 (high proficient). Results suggested that intermediate learners provided the least feedback (and mostly in the form of backchannels), while highly proficient learners' "ability to act as supportive listeners through backchanneling and confirmations of comprehension was found to be more fully developed" (Galaczi 2014: 570).

1.2 Goal

The definition of the term 'backchannel' varies considerably in the previous literature. For clarity and the purpose of this exploratory study, I will use the term 'minimal utterances' to refer to all utterances used by the secondary speaker to acknowledge, agree with, and react to the interlocutor's speech with the aim of signalling attention and understanding and responding to questions. In order to distinguish backchannels from short answers, with both being forms of minimal utterances, the term 'backchannel' will be used to refer to continuers and acknowledge tokens marking passive recipiency (PR), as well as acknowledge tokens marking incipient speakership (IS). Minimal utterances produced in replies to direct questions (yes/no or tag) will be referred to as 'response tokens' (RT). Backchannels and response tokens are not regarded as turns on their own, but it is reasonable to operationally distinguish them, as backchannels are unsolicited feedback signals while response tokens are answers to questions posed by the primary speaker, who thereby passes the floor to the secondary speaker.

In German, minimal utterances such as 'mm-hm', 'okay', 'ja' and 'genau' can be used as backchannels in so far as they are uttered as acknowledge PR and IS tokens. However, some of these tokens are used not only as backchannels, but also as responses to direct questions posed by the primary speaker. This is particularly the case with the lexical tokens 'ja' and 'genau'. Therefore, it is crucial that the contexts in which backchannels and response tokens are uttered are correctly distinguished in order to carry out a detailed pragmatic and intonational analysis of backchannels.

To gain a more precise overview of minimal utterances and their prosodic realization, four classes of minimal utterances were distinguished, based on conversational contexts and turn-taking function. These classes are 1) responses to yes/no questions, 2) responses to tag questions 3) acknowledge tokens in turn-initial position (incipient speakership) and 4) continuers, i.e. acknowledge tokens marking passive recipiency. A similar approach was chosen by Savino (2010, 2014), who investigated the intonation of backchannels in task-oriented dialogues among speakers of the Bari Italian variety. Having looked at the prosodic realization of backchannels in different pragmatic conditions, the studies suggested that backchannels produced in the PR condition were predominantly characterized by a rising intonation contour, while those in the IS condition showed a falling contour in most cases, with exceptions being explained by additional pragmatic or paralinguistic factors (Savino 2014: 11).

I will carry out a descriptive analysis of the performance of Italian learners of German in relation to that of a native German control group to identify similarities and differences between groups and provide a tentative explanation for potential differences related to the learners' L1.
To provide a comprehensive overview of backchannels in L1 and L2 German, I will focus on several relevant factors: Frequency of minimal utterances, choice of tokens in relation to pragmatic and contextual factors, intonation and proficiency effects.

In a first step, I will compare the overall frequency of minimal utterances across both groups and proficiency levels to determine whether they are used more frequently by either learners or native speakers, and whether proficiency is related to how much feedback speakers give throughout a dialogue. As for the choice of tokens, I will investigate how the two groups compare in terms of their selection of backchannel and response token types and whether the token choice is correlated to conversational moves (y/n questions and tag questions) and turn-taking function (PR acknowledge tokens, IS acknowledge tokens). The aim of the intonation analysis is to determine whether these functions have an influence on the backchannel's intonation contours, and if so, whether both speaker groups produce the same prosodic cues to mark the same pragmatic functions. Lastly, the L2 group's perfor-

mance will be analysed for proficiency effects. Learners are divided into the subgroups *beginners* and *advanced*, allowing for a more detailed look at possible proficiency effects on L2 backchannel performance.

A further development could be the assessment of the nature of the observed differences between the two groups through a subsequent comparison of the results with an L1 Italian group, which would reveal whether the native backchannel and turn-taking behaviour is transferred by Italian speakers to their L2.

2 Procedure

2.1 Participants

The corpus used in this study consists of six dyads of Italian learners of German performing the task in German (three beginner and three advanced dyads) and five L1 German dyads. The L2 German recordings took place at the Goethe Institute of Naples as part of a broader dissertation project. At the time of the recordings, L2 learners were studying German as a second language either at the Goethe Institute or at the faculty of Foreign Languages and Literatures of the University of Naples. Their proficiency levels were established on the basis of the courses they were attending, which corresponded to the levels proposed by the CEFR. The sample of participants analysed here had either an A2 level (beginner) or a C1 level (advanced). All participants chosen were originally from the dialectal area of Naples, to rule out variation in their L2 resulting from their native linguistic substratum.

The L1 German participants were recorded at the Phonetics Lab of the University of Cologne from. All participants were university students aged between 20 and 30 who had grown up in North Rhine-Westphalia.

Data was collected using a collaborative goal-oriented map task (Anderson et al. 1991). Each participant received a headset with a microphone and was recorded on a separate channel. The instructions for the map task were presented through an instructional video explaining the rules and aim of the task. Participants were instructed to choose the role of instruction giver or follower before receiving the maps. The instruction giver then received a map with landmarks and a route drawn

across them from a starting to an arrival point. The instruction follower received a map where the route was missing. The goal was for the instruction follower to reproduce the route on his map by collaborating with the instruction giver and verbally exchanging information about the position of the landmarks. But the maps contained mismatches, as some of the landmarks were either not identical or placed on a different location on the map. Participants were not informed of the mismatches prior to the task. The misunderstandings arising from these mismatches were intended to be solved by the participants themselves through interaction, as they were not allowed to ask questions to the experimenter during the task. A visual barrier was placed between the participants, not allowing them to see each other. The aim was to incentivize verbal feedback.

2.2 Annotation

The corpus of dialogues was annotated using the Praat software (Boersma & Weenink 2013) and minimal utterances were labelled in accordance with their occurrence in overlaps, turn-initial position, or as responses to yes-no or tag questions. The audio recordings consisted of two channels, one for each speaker. The annotation required three tiers for each speaker: 1) Interpausal units (utterances between silent gaps of a minimum length of 200ms), 2) Backchannels or response tokens (where the lexical or non-lexical token and class were annotated) and 3) Backchannel function (where either 'PR' or 'IS' was annotated, corresponding to function of the respective backchannel utterance). The utterances that were annotated were the non-lexical items such as 'mm*'*, 'hm*'* or 'mm-hm', as well as lexical items such as 'ja', 'genau' and 'okay'. All tokens were subsumed under the four minimal utterance types 'mm-hm', 'ja', 'genau' and 'okay'. When these minimal utterances are used to signal understanding, agreement or alignment without being explicitly invoked by means of a question, they pertain to the category of backchannels, according to the definition used in this study. Backchannels can be produced individually or at the beginning of a turn. When a backchannel was uttered without being followed by a turn on behalf of the same speaker, the token was annotated in the backchannel (BC) tier of the respective speaker. These backchannels pertained to the category of continuers or acknowledgement tokens marking *passive recipiency* (PR), and received the label "PR" in the BC function tier. When a backchannel was uttered in

turn-initial position, i.e. when it was followed by a turn on behalf of the same speaker, the token was annotated with the addition of the label *_t-in* in the BC tier (e.g. *okay_t-in*). Turn-initial backchannels were also annotated as acknowledgement tokens marking *incipient speakership* (IS) and labelled as "IS" in the BC function tier.

Acknowledgement tokens marking passive recipiency and those marking incipient speakership were considered in this study to be the standard case of backchannels, from which short answers to tag or yes/no questions were distinguished as 'response tokens'.

In German, tag questions typically consist of a declarative statement complemented by a question marker such as *oder*, *stimmt's*, or *richtig,* as in the following example taken from one of the German L1 dyads in the corpus:

S1: "Jetzt müsstest du aber über dem See sein, richtig?" ("*Now you should be above the lake, right?*")
S2: "Genau." ("*Exactly.*")

The question tag in this case is the word *richtig,* which fulfils the function of requesting confirmation from the interlocutor. Backchannels on the other hand are uttered as unwarranted signals of understanding and acknowledgement that are not produced in response to an explicit call for feedback. Thus, responses to tag questions were distinguished in the annotation by the addition of the label "*_tag*" (e.g. *genau_tag*).

A similar distinction was made with regard to responses to yes/no questions. For the sake of this analysis, yes/no questions were defined as questions that can be and are intended by the speaker to be answered with a *yes* or a *no*, as in the following example:

S1: "Hast du das 3-Sterne-Hotel?" ("*Do you have the 3-star hotel?*")
S2: "Ja." ("*Yes.*")

Since yes/no and tag questions explicitly call for a response, the feedback they induce is categorized differently from acknowledgement tokens (backchannels) in the annotation. Responses to yes/no questions thus received the label *_yn* (e.g. *ja_yn*) in order to separate them from similar tokens uttered in unsolicited feedback.

Repetitions of parts of the speaker's statement by the interlocutor used as backchannels were also separately annotated. They can appear as in the following example:

S1: "Jetzt rechts." ("Now [turn] to the right.")
S2: "Rechts." ("To the right.")

Similar to other backchannels, repetitions function as a way of confirming and demonstrating understanding of what has been said. However, as repetitions are highly individual, specifically repeating parts of the previous utterance, these tokens do not provide a suitable basis for a comparative analysis, as opposed to tokens like 'mm-hm', 'ja' and 'okay', which are used at a high rate across speakers and dyads.
Backchannels that were not identifiable due to laughing, noise or shortness, were labeled as "_X" and excluded from further analysis, as they would have impeded acoustic analysis.

2.3 Data extraction

In order to extract intonational information from the annotated backchannels, a Praat script was used first to separate the two audio channels for each speaker and move the starting and end point of each annotated backchannel to a zero-crossing point in the signal. This was done to avoid discontinuities in the waveform that may result in click or pop sounds in the audio. The annotated backchannels were then extracted by creating a single audio file for each individual token, except for the ones, which were previously excluded from the annotation, as explained above. In addition, the corresponding parts of the textgrid files, containing the label information, were extracted.

The pitch contour was pre-processed through smoothing and manual correction of pitch points (Cangemi 2015). This made it possible to correct error values of octave jumps, creaky voice and sharp peaks corresponding to articulatory phenomena such as strong bursts. Lastly, intonation contour information was extracted from the audio files by means of a semitone step analysis. To achieve this, pitch points were taken from two time points in the audio files, one at the beginning and one at the end of the signal. A rise of more than one semitone step between the first and last pitch point taken was defined for the purpose of this analysis as *rise* in intonation. A pitch movement of no more than +/- 1 semitone step was defined as a *level* contour, while a downward movement greater than -1 semitone step resulted in a *fall* in intonation.

It is important to note that due to the shortness of the tokens, it is in some cases difficult to extract reliable f0 information, since particularly short backchannels may not have enough periodic energy. As a result, intonation contours can be misrepresented (when the wrong pitch points are taken) or unable to be detected (when not enough pitch points are available).

3 Results

The total amount of BC tokens produced by both groups was 718, from which 574 (80%) were used in the prosodic analysis. The remaining 144 tokens could not be processed, presumably either due to the shortness of the utterance or lack of periodic energy (creaky voice, etc.), and thus did not undergo further prosodic analysis.

The group of L1 speakers, consisting of 5 dyads, produced 390 tokens, 293 of which underwent prosodic analysis. The group of L2 speakers consisted of 6 dyads, 3 of which were beginners and 3 advanced speakers. They produced a total of 328 tokens, with 281 of those having been used in the prosodic analysis. It should be noted, however, that one participant from one of the L2 dyads (FF) produced very few backchannels during the entire task (only four backchannels from which two underwent further analysis). It can only be speculated as to why this particular participant produced such a low output. Low proficiency, idiosyncratic or motivational factors could have played a role. Since I have no Italian L1 data from this speaker

to compare the L2 output with, and otherwise the setting and recording conditions were not different from the experimental settings in other recordings, I decided not to exclude the data from this dyad from the analysis.

3.1 Minimal utterance rate per minute of dialogue

Overall, the L1 group produced roughly 20% more tokens than the L2 group (390 tokens produced by the native speakers, compared to 328 by the learners). Looking at the rate of response tokens per minute (1), it seems that advanced L2 speakers approached the target language values, while the beginner group had a lower output. On average, the L1 German speakers had an output of 7.6 tokens per minute, compared to 7.3 for the advanced group and 4.2 for the beginners.

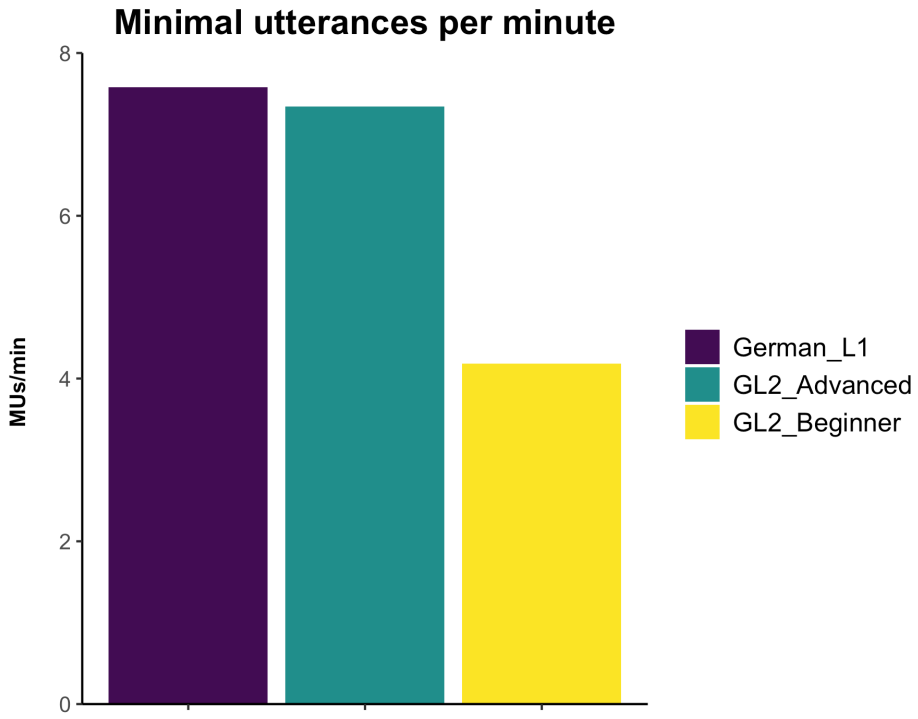

Fig. (1): Minimal utterances per minute by group and proficiency

However, looking at MUs per minute by dyad provides a different picture, suggesting that individual variability plays a crucial role. Fig. (2) shows that two of the L2 beginner dyads indeed have the lowest MU rates among the L2 dyads. But at the same time, one beginner dyad has produced more minimal utterances than the other beginner dyads and even the three advanced dyads. It can be hypothesized that proficiency is not the defining factor for the output of verbal feedback signals. Idiosyncratic factors, such as individual preferences, or L1 patterns could have played a role. Comprehension problems might as well have had a negative impact on the output of the two low proficient dyads with the least MUs per minute, since German was spoken throughout the recording sessions as well as in the instructions of the map task. Despite the relatively small corpus size used for this exploratory study, the results in (2) nevertheless indicate that averaged-out values have to be taken with care. Valuable insights could be concealed by lumping dyads together into proficiency groups, as this example shows.

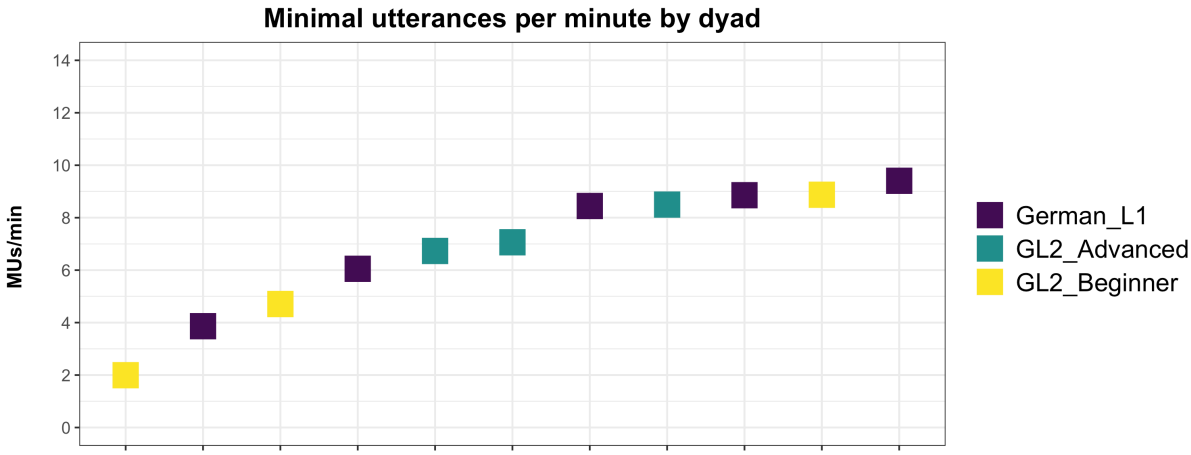

Fig. (2): The rate of minimal utterances produced per minute by dyad

3.2 Minimal utterance type, function and class

A categorical analysis of minimal utterance token choice (Fig.3) reveals more apparent differences between the learner groups and the native speakers. While there is a more even distribution in the choice of tokens among the L2 beginner and advanced speakers, there is a much higher proportion of 'ja' produced by the L1 German group. For the latter group, 'ja' was chosen at a rate of about 45%.

Among the advanced L2 speakers ‘okay’ was the most frequent type (38%). For the L2 beginners it was ‘ja’ (34%) with ‘okay’ being uttered in 32% of the cases. Overall, non-lexical types such as ‘mm-hm’ were produced more frequently by L2 dyads than by the native German speakers.

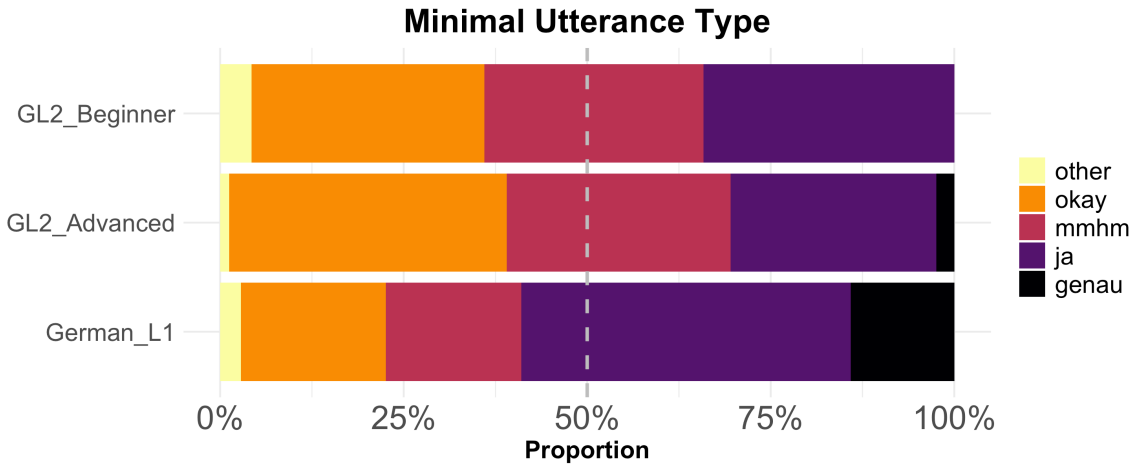

Fig. (3): The choice of minimal utterance type by group.

Taking the choice of minimal utterance types *by function* into consideration, further differences in the output of both groups become visible. Two functions of acknowledgment tokens were taken into consideration: marking passive recipiency (PR), not taking the floor, and incipient speakership (IS), taking the floor. ‘OT’ stands for other conditions such as responses to yes/no questions and tag questions. Minimal utterances categorized under ‘OT’ thereby refer to *response tokens*. Acknowledge tokens produced with the function of marking PR and IS will be referred to as *backchannels*.

Figure (4) shows that the most frequent type used in all three conditions is ‘ja’ in German L1. In the German L2 group on the other hand, there is a more diverse output preference. The backchannel types ‘mm-hm’ (38.8%) and ‘okay’ (38.4%) were the most frequently used acknowledge PR types by L2 speakers. In the context of incipient speakership (IS), the L2 group produced ‘okay’ almost twice as often (61%) than the L1 group (33%).

Another significant difference between both groups is the use of the type ‘genau’. L1 speakers use it in all three conditions and it is most frequent as a response token (32%). By the L2 group it is rarely used, showing a rate of 4% in replies to questions and 0.8% as a backchannel.

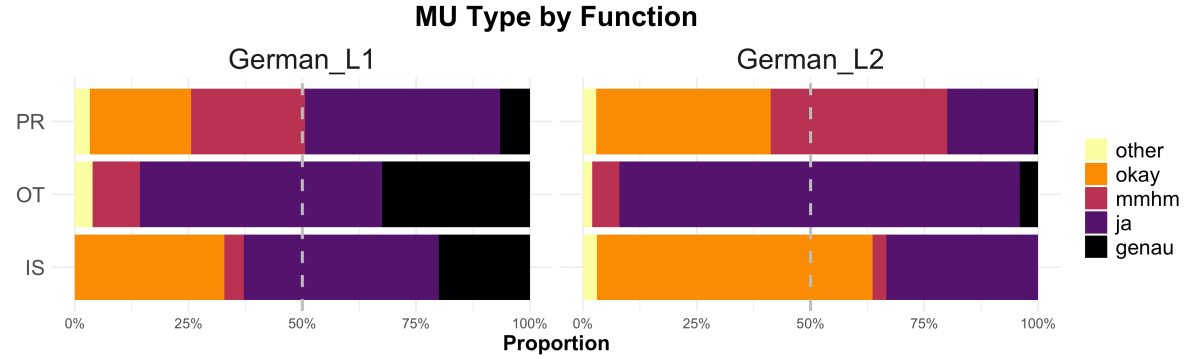

Fig. (4): Minimal utterance types produced in relation to function by group

The following figure (5) provides a more detailed overview of how the minimal utterances are distributed across the different functions. Here, responses to yes/no (yn) and tag questions are shown as individual classes rather than being grouped together under ‘OT’. Additionally, the ‘t-in’ (turn-initial) bar shows the proportion of minimal utterances (other than backchannels) that are followed by a turn of the same speaker.

One of the more salient differences between both groups is the more diverse token choice after yes/no questions in the L1 group. Apart from the lexical types ‘ja’ (51%) and ‘genau’ (32%), the German native speakers used the non-lexical ‘mm-hm’ as a response to yes/no questions in 16% of cases. The L2 group never used the latter token in this context, but predominantly answered with ‘ja’ (96%).

In responses to tag questions, however, it is the L1 group that uses ‘ja’ in the vast majority of cases (77%). And while native speakers very rarely produce ‘genau’ (8%) in this context, the learner group shows the highest output of ‘genau’ tokens (17%) in tag-question responses compared to other conditions.

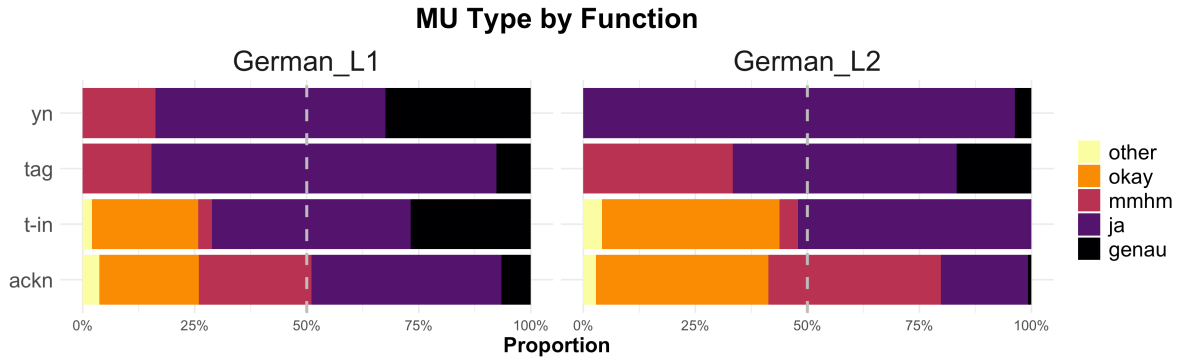

Fig. (5): Minimal utterance types by function across classes, including y/n questions, tag questions, and turn-initial position

A comparison of type choice by *dyad* (Fig. 6) again indicates that individual variability might play an important factor, particularly in the L2 group. In this figure, every bar corresponds to the output of one individual dyad. The dyads are listed in alphabetical order on the y-axis. While there seems to be a relatively stable pattern of distribution in the L1 group, the opposite is the case in the learner group. The L2 dyads BS and CV for instance choose 'okay' in a majority of cases and use 'mm-hm' the least. FF and IF on the other hand have a high output of 'mm-hm', while 'okay' makes up a relatively small proportion. Furthermore, dyads BS and FF, who show particularly opposing performances regarding token choice, are both advanced learner dyads, making it difficult to argue that proficiency is the reason for these discrepancies. Rather individual preferences may be considered as an explanation, as in the case of overall rate of minimal utterances. This remains to be investigated in a comparison with the learners' L1 performance.

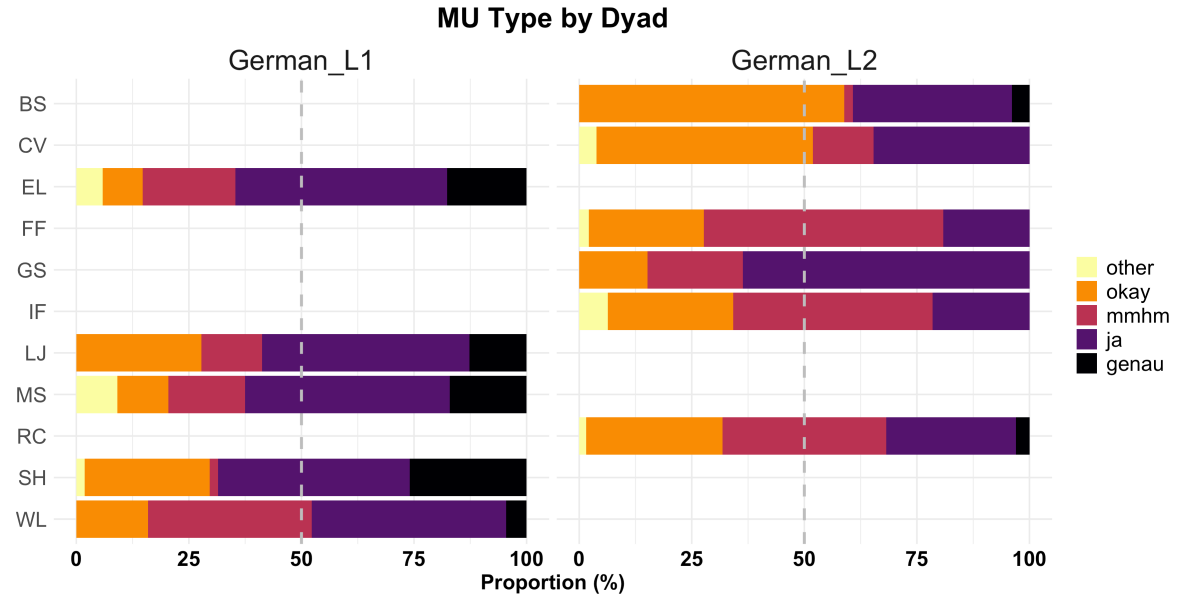

Fig. (6): Token choice by dyads. Each abbreviation on the left stands for one dyad.

3.3 Prosodic analysis

In the following I will present the results of the prosodic analysis. It is important to note that from around 20% of the extracted tokens that were used in the previous part of the analysis no pitch contour information could be extracted. Token length and the intensity with which the tokens were produced are likely to explain the loss of data. In some cases minimal utterances might have been too short (resulting in insufficient periodic energy) in order for an automatic detection of pitch values to be successful. Often backchannels were produced rather quietly or with creaky voice, which negatively affected a reliable extraction of f0 contour information. In the remaining data that did undergo prosodic analysis, creaky voice in the onset or offset of some utterances can explain why some tokens exhibit very large semitone (ST) jumps or falls (+/- 10 ST or more).

Comparing the prosodic realization of minimal utterance types across the two groups (Fig. 7), it becomes apparent that the learner group performed fairly similar to the native speakers in terms of the overall intonational patterns for each type. Both ‘mm-hm’ and ‘ja’ were produced with rising intonation in the majority of cases by both groups. The same is true for ‘genau’, even though there was a significantly lower output of this type in the L2 group. The only categorical difference is that ‘okay’ is produced more often with a rising contour by the learner group. In

30% of cases ‘okay’ shows a rising intonation contour in the L1 group, while L2 speakers produced it with a rising contour 47% of the time.

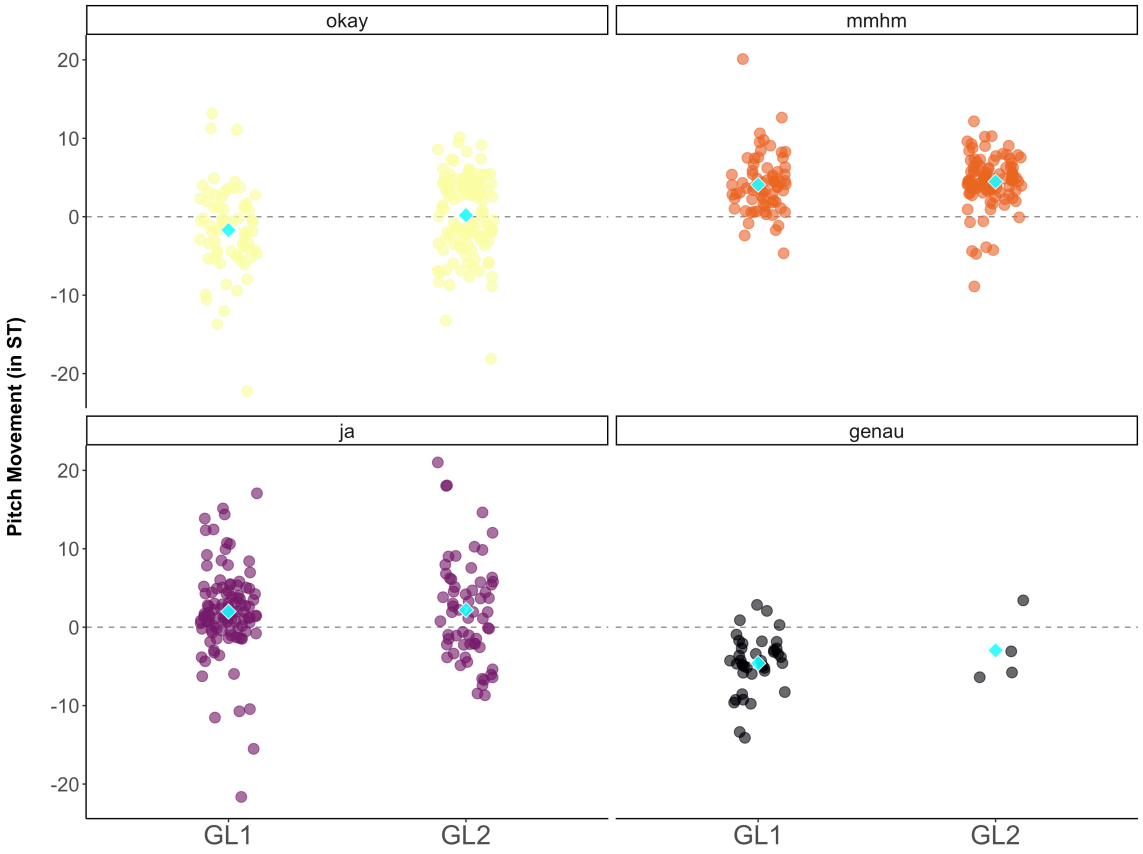

Fig. (7): Pitch movement by group across minimal utterance types.

The next figure (8) shows the pitch movement in relation to the different functions for both speaker groups combined. Comparing in particular the intonation across the functions IS and PR, there appears to be only a subtle difference. Nevertheless, the results shown here confirm that there is indeed a slight tendency for backchannels marking incipient speakership to fall in intonation, while the majority of those marking passive recipiency have a rising pitch contour. Thus the trend observed by Savino (2010, 2014) for native speakers of Bari Italian can be supported in the present exploratory study with data from L1 and L2 German speakers, suggesting that this correlation may be a cross-linguistic one.

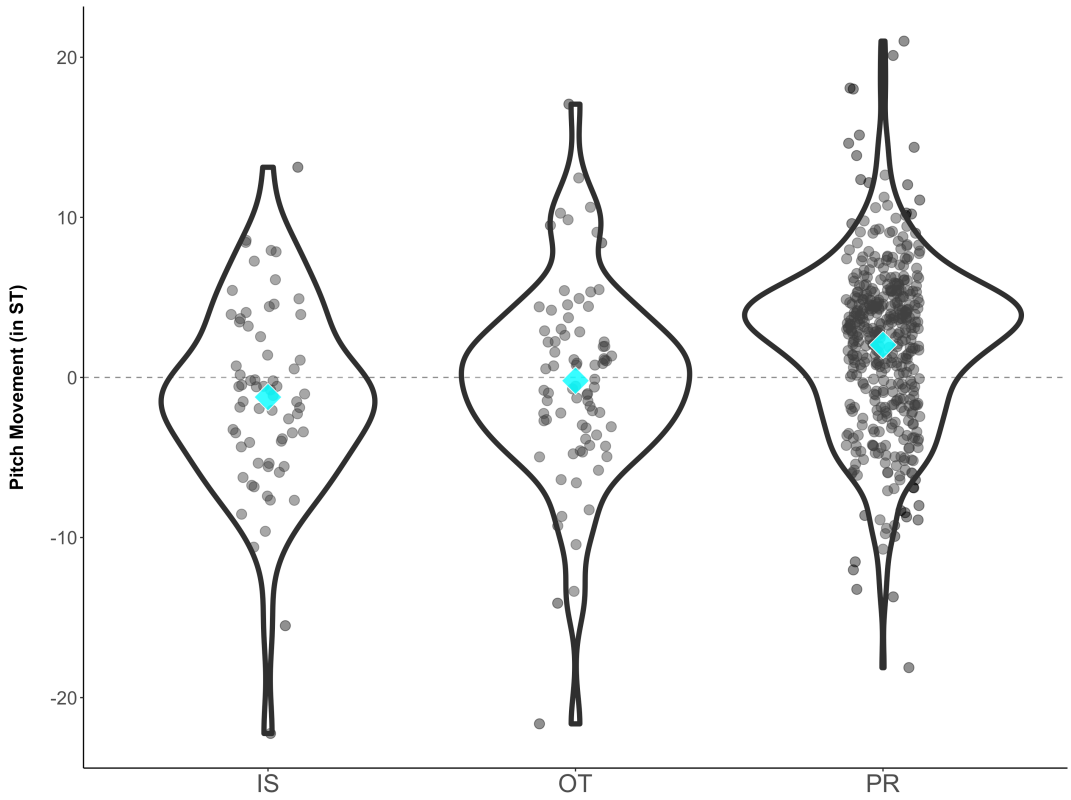

Fig. (8): Pitch movement across the functions IS and PR. OT refers to yes/no and tag questions.

However, there are indications that the pragmatic function is insufficient for predicting backchannel intonation. Figure (9) shows pitch movement by function across backchannel types, suggesting that the token choice overrides the trend observed when only looking at pitch movement in relation to PR and IS.

Minimal utterances of the non-lexical type 'mm-hm' have a rising intonation across all function. In the cases in which it is used as a backchannel, i.e. acknowledge token marking PR and IS, 'mm-hm' tokens show a predominantly rising pitch contour. Even though only a small number of 'mm-hm' tokens are produced as incipient speakership markers, all of them have a rising pitch contour. This contradicts the trend suggested in figure 8. When 'mm-hm' tokens are uttered as response tokens, the same pattern can be observed. In none of these cases 'mm-hm' is produced with a falling intonation.

Looking at 'genau' tokens, the opposite pattern emerges. Across all three functions, 'genau' is predominantly produced with falling intonation.

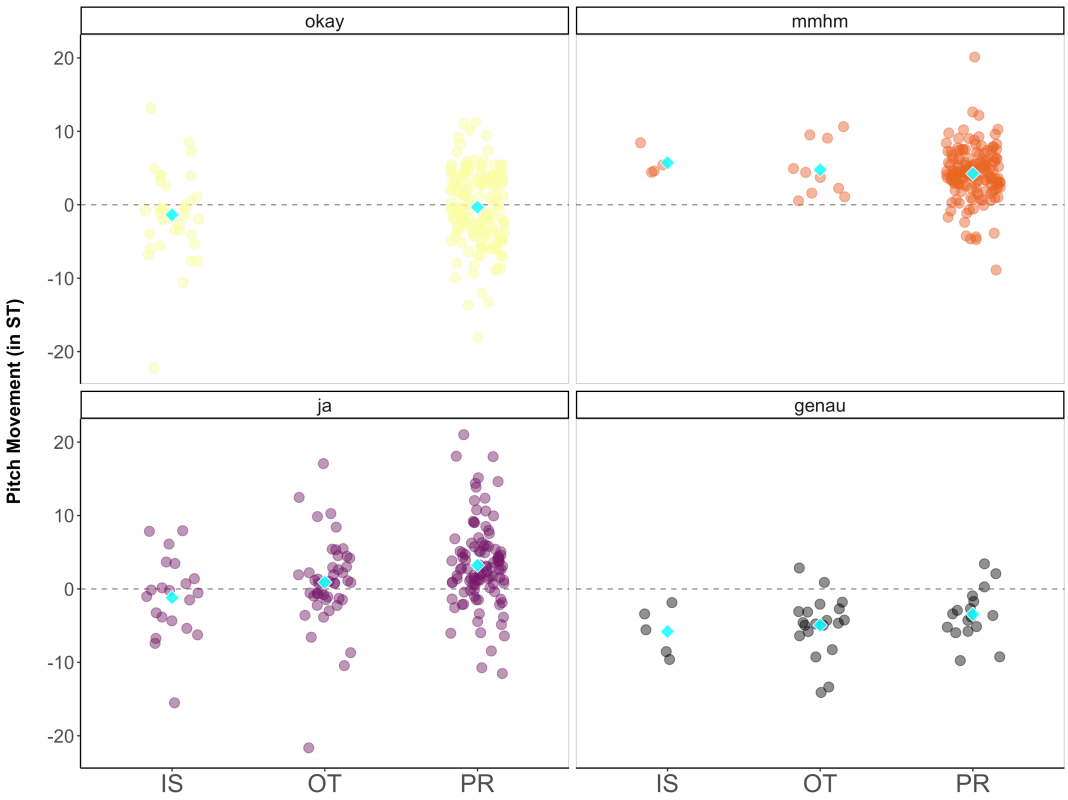

Fig. (9): Pitch movement by function across types.

Moving on to backchannels of the type 'okay', the overall pattern observed is less conclusive compared to the former cases. Indeed there appears to be a tendency for 'okay' tokens to be produced with a falling pitch contour in the incipient speakership condition, conforming to the pattern seen in figure 8. But there is an even distribution with only a marginal trend for either falling or rising intonation for 'okay' tokens used as acknowledgement markers in the condition of passive recipiency. Furthermore, in neither speaker group 'okay' is used in responses to yes/no or tag questions.

Of all types, 'ja' is the only one whose utterances seem to conform to the pattern described in figure 8. Most tokens in IS are falling, while in PR there is a clear tendency for rising contours. In yes/no and tag questions 'ja' is more often produced with a rising pitch movement than with a falling contour.

Considering the output of both speaker groups separately (10 and 11), overall the same patterns can be observed. The most significant differences, however, are seen with respect to 'okay' and 'genau'. Since the learner group opted for 'genau' in only four occasions, there is not enough speech material for reliable conclusions to be

drawn. Nevertheless, three out of four ‘genau’ utterances are produced with a falling pitch contour, which is in line with the L1 pattern for the same utterance type.

In contrast to the mostly falling pitch contours in ‘okay’ utterances regardless of function by the L1 group, learners tended to produce ‘okay’ with rising intonation in the passive recipiency function. Marking incipient speakership, this backchannel type is rising in intonation slightly more often in the L2 group compared to the L1 group.

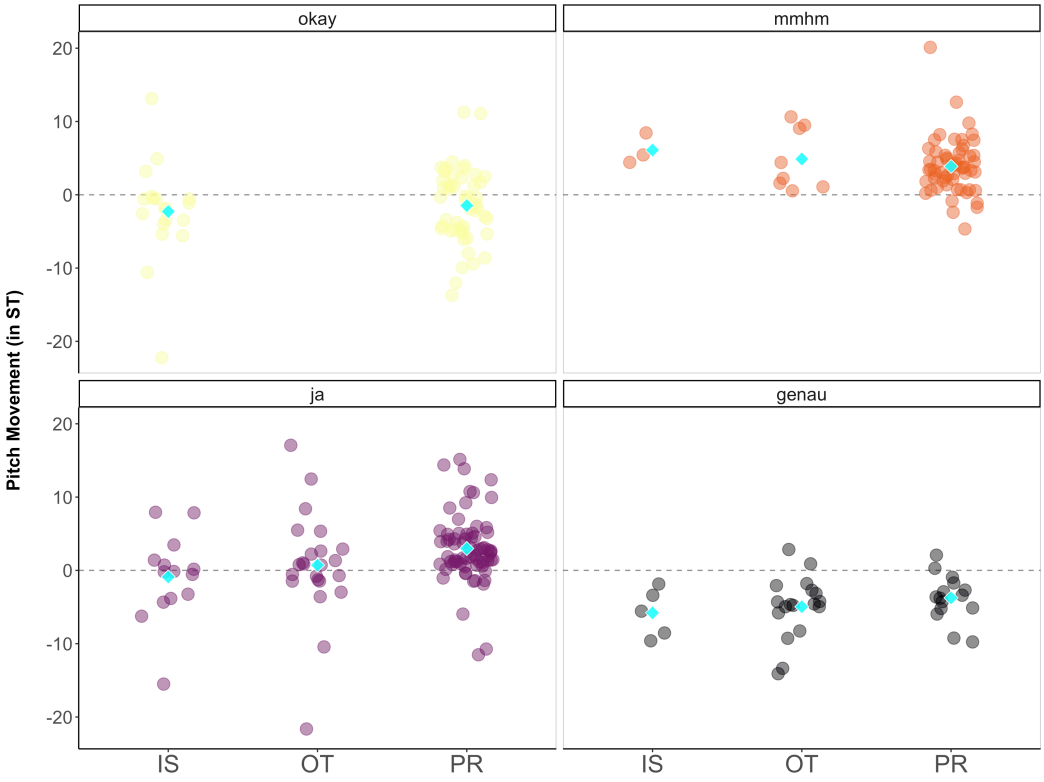

Fig. (10): Pitch movement by function across types in German L1.

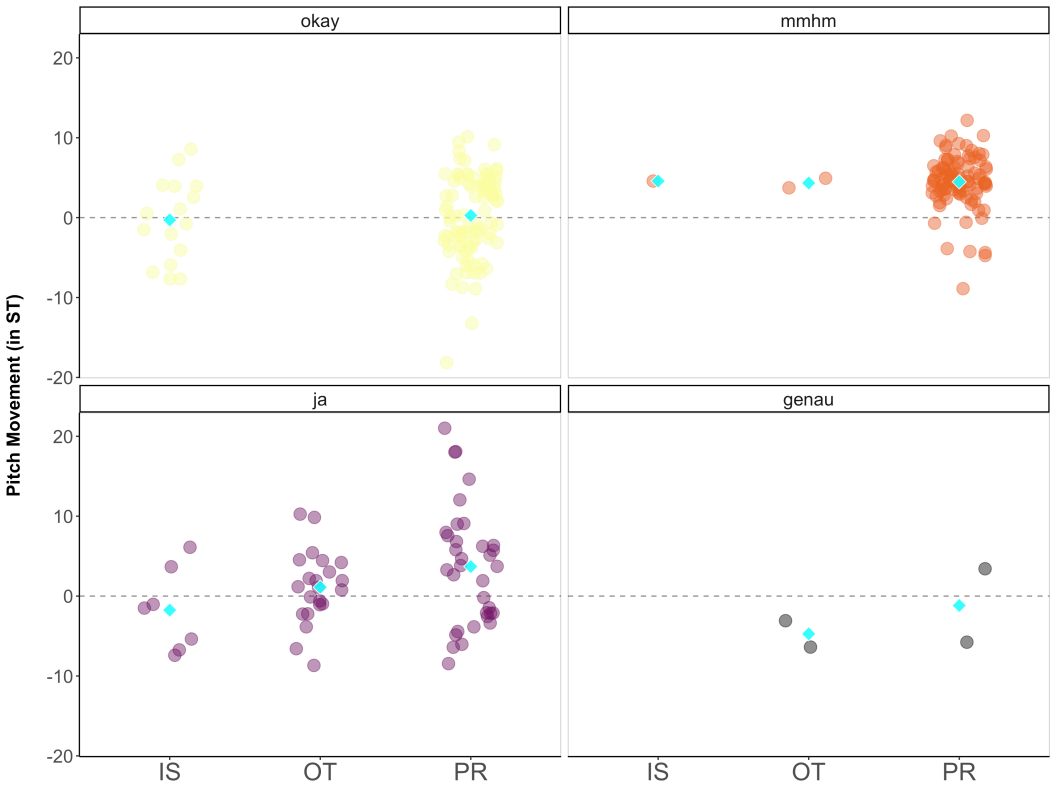

Fig. (11): Pitch movement by function across types in German L2.

4 Discussion

Having considered both token choice and intonation contours of minimal utterances in L1 and L2 German speakers, some minor and major differences between the two groups will be discussed in the following. Firstly, the results of this exploratory study show that L1 and L2 speakers differed mainly with regard to backchannel type and response token choice (figures 3-6). While L2 learners overall displayed a greater tendency to produce non-lexical tokens (e.g. 'mm-hm'), German natives speakers opted for 'ja' in the majority of cases. In addition, there was greater minimal utterance type variation within the L1 group, with 'genau' having been used at a much higher frequency by native speakers than by learners.

Considering token choice in relation to function, more differences appear between both groups. 'Mm-hm' and 'okay' are chosen more frequently as acknowledge tokens marking passive recipiency by the L2 group. The learner group also produces 'okay' significantly more often as acknowledge tokens marking incipient speakership compared to the native speakers. When it comes to the token choice of re-

sponse tokens, learners produced ‘ja’ almost all of the time, while the L1 group used ‘genau’ in around a third of cases. However, as the by-dyad analysis has shown, individual variability should be taken into account when looking at token choice. Especially within the L2 group, some dyads showed vastly different outputs in terms of token choice, which could be concealed by only looking at averaged-out values.
Proficiency did not seem to be a reliable factor to determine native-like performance, as some beginner dyads have shown to produce a more native-like output than some advanced learner dyads.

Overall, it can be hypothesized that token choice in the L2 might be related to a certain extent to language-specific token choice patterns in the speakers’ L1. In order to determine whether or not this is the case, a further more extensive analysis will have to be carried out comparing how learners’ L2 token choice compares to the token choice in their Italian L1. For instance, future research could look into whether Italian learners of German use ‘mm-hm’ tokens in their native language with the same frequency and in the same contexts as in the L2. Another question would be whether more or less literal translations of lexical backchannels (‘ja’ and ‘sì’, or ‘genau’ and ‘esatto’) function in the same way in both languages, i.e. whether they are produced with a similar function-intonation relation in the first and second language.

The intonation analysis of backchannels and response tokens in the context of different pragmatic functions and conversational moves (IS, PR and positive replies to yes/no or tag questions) revealed that there is a tendency for acknowledge tokens marking passive recipiency to be produced with rising intonation. On the other hand, acknowledge tokens marking incipient speakership more commonly show a falling intonation contour. Despite the marginal difference between the two categories, these results nonetheless confirm the observation made by Savino (2010, 2014) for native speakers of Bari Italian, suggesting that this pattern may be a cross-linguistic phenomenon.

Analysing the intonation of minimal utterance types individually has shown, however, that 'mm-hm' backchannels are produced predominantly with a rising intonation contour regardless of the turn-taking function. Tokens of this backchannel type rise in intonation both in PR and IS, according to the results of the present study. A similar function-independent pattern was observed in the case of 'genau', which is produced with a falling pitch movement across all conditions. This indicates that in order to understand how backchannel intonation is related to different functions regarding turn-taking moves, it is crucial to look at backchannel types individually, as the types themselves appear to be produced with type-specific intonation patterns.

A mixed – or function-dependent – pattern was observed for 'ja' in both groups and to a slight extent for 'okay' in the L2 Group. However, in both cases the tendency toward a particular intonation contour in relation to the functions IS and PR was marginal, as can be seen in (10) and (11). In the case of 'ja' backchannels there is a clear pattern in the passive recipiency condition. When speakers signalled their intention to pass the opportunity to take the turn on to the primary speaker, both groups produced 'ja' with a pitch rise in the majority of cases, as the figures show. 'Ja' backchannels marking *incipient speakership* display a more even distribution of rises and falls, despite a subtle trend toward falls in both groups. A similar pattern emerged in the case of 'ja' used in replies to yes/no and tag questions. A probable explanation for this outcome could be that there are further contextual factors influencing intonation patterns that were not picked up by the variables investigated in this analysis. A more fine-grained analysis might be required that takes into account more pragmatic functions, such as the status of the speakers' assumed shared background knowledge, disagreement, acknowledgement of the interlocutor's previous acknowledgement, or the end of a set of instructions for completing a part of the task, as proposed by Savino (2014: 11). Taking these pragmatic functions into account could be essential for further differentiating the distribution of intonation contours observed for 'ja' and 'okay' backchannels across the rather coarse set of functions used in this analysis, and might reveal more differences or similarities between L1 and L2 performances.

Considering, for instance, the speaker's acknowledgement of the interlocutor's previous acknowledgement, it can be observed that in such cases 'okay' is often used as a backchannel. In the following example, taken from an L1 dyad, the primary speaker says 'okay' to acknowledge the utterance of his interlocutor, who just replied to his confirmation request:

S1: "Ja?" ("*Yes?*" – In the sense of "*did you understand?*")
S2: "Ja." (Positive reply, falling pitch contour)
S1: "**Okay**." (*Falling* pitch contour)

The falling intonation might indicate that a certain section of the dialogue is over, or that a particular (sub)topic has been dealt with. The same applies when the primary speaker, who in this scenario is the instruction giver, asks the instruction follower for clarification:

S1: "Was kommt denn bei dir über dem Park?" ("*What do you have above the park?*")
S2: "Äh die Arena (.) also die *neue* Arena, es gibt ja zwei." ("*uh the arena* (.) *the <u>new</u> arena, there are two.*")
S1: "**Okay**." (*Falling* pitch contour)

What both examples have in common is the fact that the instruction giver, who in the map task scenario is usually the one speaking for longer stretches of time, asks the instruction follower a question which is immediately resolved after a short utterance or explanation. There are no further enquiries on behalf of the primary speaker and no expectation that the secondary speaker's turn might continue. The falling intonation marks the end of a segment that had a particular purpose which has been accomplished.

In the following example the same token is used, but this time by the secondary speaker to signal understanding of an instruction that the primary speaker is giving:

S1: "Also wenn du jetzt nach… du gehst jetzt gleich nach oben (…)" (Final rise)
"*So when you… you'll have to go up next (...)*"
S2: "Ja." (Rising intonation)
"*Yes.*"
S1: "Du gehst nicht in 'ner graden Linie nach oben (…)" (Final rise)
"*You don't go up in a straight line (...)*"
S2: "**Okay**." (Rising intonation)
S1 proceeds with instruction.

Here, the secondary speaker is listening to the instruction knowing that it will continue, which she correctly infers from the fact that the instruction giver ends every segment of his instruction with a final rise. Both backchannels produced by the secondary speaker ('ja', 'okay') have a rising pitch contour, signalling understanding of the fact that the instruction giver may continue.

In these three instances, 'okay' backchannels were uttered in different pragmatic contexts but yet all three pertain to the category of acknowledge tokens marking passive recipiency in so far as they were not (immediately) followed by a turn on behalf of the same speaker, but rather they fulfilled the function of signalling to the speaker that he may go on. This shows the need for a further, more detailed pragmatic analysis, mainly in the case of 'ja' and 'okay' backchannels. Particularly because for these two types there is no clear rise or fall tendency as in the case of 'mm-hm' and 'genau'.

However, in contrast to 'okay', 'ja' serves not only as an acknowledge token but also mainly as a response token in positive replies to yes/no and tag questions. Interestingly, again there is no clear pattern with regard to intonation other than a slight tendency for rising pitch movements in both groups. Similar to the case of 'okay' described above, only a more in-depth pragmatic analysis taking into account the context of each utterance might reveal under which conditions 'ja' tokens in positive replies are produced with a rising and when with a falling intonation contour. For reasons of space, the pragmatic analysis that would be required to yield any fruitful insights in this matter cannot be done within the scope of this work.

5 Conclusion and outlook

This exploratory study has investigated the use of minimal utterances, including backchannels and response tokens, by L1 and L2 speakers of German. First, the rate of minimal utterances and the token choice of the two groups were compared. In the next step, the aim has been to analyse how the intonation of backchannels is related to specific turn-taking functions; more specifically, how intonation is used to either signal the speaker's intention to let the interlocutor carry on with his turn (passive recipiency), or indicate the speaker's intention to begin a turn of his own following the backchannel utterance (incipient speakership). Subsequently, the four backchannel types 'mm-hm', 'ja', 'okay' and 'genau' were analysed individually across functions to reveal whether there are type-specific intonation patterns. Backchannels and response tokens were distinguished to investigate potential differences between the two classes of minimal utterances in terms of intonation and token choice. Furthermore, the performance of the learner group, which consisted of beginner and advanced learners, has been analysed for proficiency effects throughout the analysis.

The results of this study have shown that the two groups differ from each other especially with regard to token choice. Overall, the learner group has shown a higher preference for 'okay' and 'mm-hm' as acknowledge tokens marking passive recipiency (PR), and used 'okay' more frequently as acknowledge tokens marking incipient speakership (IS), while also avoiding the use of 'genau' almost entirely. The native speaker group opts for 'ja' in the majority of cases across both functions PR and IS as well as in replies to yes/no and tag questions.

In line with previous observations made by Savino (2010, 2014) for native Italian speakers from the Bari region, the intonation analysis has revealed that there is a tendency for backchannels in the PR condition to be produced with a rising pitch movement, while those in the IS condition have a predominantly falling intonation contour across both speaker groups. Further analysis has shown, however, that the vast majority of non-lexical backchannels of the type 'mm-hm' have a rising pitch contour regardless of the function. Similarly, 'genau' backchannels are almost exclusively produced with falling intonation contours across all functions. This indicates that backchannel intonation to a certain extent is rather type-specific than being related to the functions of 'passive recipiency' and 'incipient speakership'.

While the backchannel type ‘okay’ conforms to the proposed tendency of intonation rises in PR and falls in IS, albeit to a marginal degree, this only applies when the results of both speaker groups are combined. In the L1 group ‘okay’ tokens are produced with predominantly falling pitch contours, whereas in the learner group there are slightly more rising than falling ‘okay’ tokens. Both groups performed fairly similar with respect to the intonation of ‘ja’ backchannels across the different functions. However, the intonation analysis across functions for ‘ja’ tokens has shown a high degree of variability, with no clear-cut tendency toward rises or falls being observed across functions. It has therefore been suggested that more pragmatic and contextual parameters should be taken into account to gain a more fundamental understanding of the factors to determine intonation rises and falls in ‘okay’ and ‘ja’ backchannels.

Proficiency has not shown to be a substantial factor in the choice of tokens or intonation of backchannels. Especially when looking at the performance of the L2 group for each of the dyads individually, it can be noted that there is a high degree of individual variability, making it difficult to rely on averaged-out data. There appeared to be no indication that higher proficiency led to a more native-like output, regarding both token choice and backchannel intonation. As stated above, further research into the learners’ L1 is needed, however, to provide a more in-depth explanation of the L2 patterns observed here.

Acknowledgements

I would like to thank Simon Wehrle (University of Cologne) for providing the Praat and R scripts used for the data extraction and creation of the plots. I would also like to thank Simona Sbranna (University of Cologne) for adapting the scripts to the particular requirements of the analysis.

References

Anderson, A. H., Bader, M., Bard, E.G., Boyle, E.H., Doherty, G.M., Garrod, S.C., (...) & Weinert, R. (1991). The HCRC map task corpus. Language and speech, 34(4), 351-366.

Benus, S., Gravano, A., Hirschberg, J. (2007). The prosody of backchannels in American English. In Proc. of ICPhS 2007, 1065-10-68, Saarbruecken, Germany.

Berry, Anne (1994). Spanish and American turn-takingstyles: A comparative study. In L.F. Bouton (ed.), *Pragmatics and Language Learning Monograph Series*, Vol. 5 (pp. 180-190). Urbana: University of Illinois.

Boersma, P., & Weenink, D. (2013). *Praat: doing phonetics by computer* (Version 6.0.40) [http://www.praat.org]

Castello, E. and Gesuato, S. (2019). Holding up one's end of the conversation in spoken English Lexical backchannels in L2 examination discourse. In *International Journal of Learner Corpus Research*, Volume 5, Issue 2, 231-251. Doi: https://doi.org/10.1075/ijlcr.17020.cas

Cangemi, F. (2015). *mausmooth*. [http://phonetik.philfak.uni-koeln.de/fcangemi.htlm]

Clancy, P.M., Thompson, S.A., Suzuki, R. & Taho, H. (1996). The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics, 26*(3), 355-387.

Cutrone, P. (2011). Politeness and face theory: Implications for the backchannel style of Japanese L1/L2 speakers. *Language Studies Working Papers, 3*, 51-57.

Drummond, Kent and Robert Hopper, (1993). Back channels revisited: Acknowledgement tokens and speakership incipiency. *Research on Language and Social Interaction 26*: 157-177.

Galaczi, Evelina D. (2014). Interactional Competence across Proficiency Levels: How do Learners Manage Interaction in Paired Speaking Tests?. In *Applied Linguistics*, Volume 35, Issue 5, 553–574, https://doi.org/10.1093/applin/amt017

Ha, K.P., Ebner, S. & Grice, M. (2016). Speech prosody and possible misunderstandings in intercultural talk - A study of listener behaviour in Vietnamese and German dialogues. Proc. Speech Prosody 8, Boston, 801-805.

Jefferson, Gad. (1984). Notes on a systematic deployment of the acknowledgement tokens "yeah" and "Mm hm'. *Papers in Linguistics 17*: 197-216.

Levinson, S.C., and Torreira F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*. 6:731. doi: 10.3389/fpsyg.2015.00731

Li, H.Z. (2006). Backchannel responses as misleading feedback in intercultural discourse. *Journal of Intercultural Communication Research, 35*(2), 99-116.

Roberts, S. G., Torreira, F., and Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Front. Psychol*. 6:509. doi: 10.3389/fpsyg.2015.00509

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Savino, Michelina (2010). Intonational strategies for backchanneling in Italian Map Task dialogues. In: *ISCA Tutorial and Research Workshop on Experimental Linguistics*, Athens, 157–160.

Savino, Michelina (2014). The Intonation of Backchannel Tokens in Italian Collaborative Dialogues. In: Vetulani Z., Mariani J. (eds) Human Language Technology Challenges for Computer Science and Linguistics. LTC 2011. Lecture Notes in

Computer Science, vol 8387. Springer, Cham. https://doi.org/10.1007/978-3-319-08958-4_3

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*, Georgetown University Press, Washington, DC (1982), pp. 71-93

Shelley, L., & Gonzalez, F. (2013). Back Channeling: Function of Back Channeling and L1 Effects on Back Channeling in L2. *Linguistic Portfolios, 2*(1), 9.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. Proc. Natl. Acad. Sci. U.S.A. 106, 10587–10592. doi: 10.1073/pnas.0903616106

Tavakoli, P. 2016. Fluency in monologic and dialogic task performance: challenges in defining and measuring L2 fluency. In: *International Review of Applied Linguistics in Language Teaching, 54*(2), 133-150.

van Os, M., de Jong, N.H. and Bosker, H.R. (2020). Fluency in Dialogue: Turn-Taking Behavior Shapes Perceived Fluency in Native and Nonnative Speech. *Language Learning*, 70: 1183-1217. https://doi.org/10.1111/lang.12416

Wehrle, S., and Grice, M. (2019). Function and Prosodic Form of Backchannels in L1 and L2 German. Poster at *Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019*, Seoul, South Korea.

Yngve, Victor. (1970). On getting a word in edgewise. In: *Papers from the Sixth Regional Meeting, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society. 567-577.

Eidesstattliche Versicherung:

Hiermit versichere ich an Eides statt, dass ich diese Bachelorarbeit selbstständig verfasst und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die Stellen meiner Arbeit, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche unter Angabe der Quelle kenntlich gemacht.
Diese Arbeit habe ich in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, _____ Unterschrift: _____