Commitments in German Tag Questions: An Experimental Study

Yulia Clausen and **Tatjana Scheffler**
Department of Linguistics, SFB 1287
University of Potsdam
Germany
{yulia.clausen|tatjana.scheffler}@uni-potsdam.de

Abstract

While plain assertions commit the speaker to the expressed proposition, there are other speech act types that involve various degrees of tentative or projected commitment by both speaker and hearer, for example, tag questions. In this paper, we systematically test the behavior of German tag questions in contexts with different kinds of speaker/hearer commitments, and compare them to their English counterparts. Our experimental results paint a detailed picture of the felicity of German tag questions. We show that individual question tags vary not only on the commitment dimension, but also wrt. speaker certainty and the target of confirmation the speaker intends. Together these categories illustrate the variation of German tagged utterances and reveal aspects that need additional modeling.

1 Introduction

The discourse model by (Farkas and Bruce 2010) was developed for default assertions and questions. These constructions are associated with either direct commitment publicizing (assertions) or lack of commitments (questions). However, there exist "intermediate" constructions, such as tag questions, that involve tentative commitments.

It has been argued that these constructions modify the illocutionary force of an utterance (e.g., Reese and Asher 2007; Malamud and Stephenson 2014; Krifka 2015), and therefore affect which commitments, if any, are effected by the utterance. It has been noted that such modifiers (e.g., same polarity vs. reverse polarity tags in English) differ in meaning and function, and these semantic-pragmatic differences in turn arise from the varied relationships between speakers and hearers in the discourse situation, as well as the prejacent proposition (cf. Malamud and Stephenson 2014: 276).

To accommodate these constructions, Malamud and Stephenson (2014) proposed a modification to the conversational scoreboard model (Farkas and Bruce 2010) by introducing projected commitments, which allowed to distinguish between public and tentative commitments of participants. The extended model accounts for the differences between reverse polarity (RP) and same polarity (SP) tag questions in English. But would it also explain the differences one might encounter with other types of similar constructions? The German language has been shown to have a large inventory of invariant tag questions that on the one hand are largely interchangeable and on the other hand are subject to certain context restrictions, as we have shown in our recent corpus study (Clausen and Scheffler 2020).

We systematically test the commitments contributed by four German question tags in an online experiment, by varying the semantic and pragmatic context around the utterances along the dimensions proposed in previous theoretical and corpus-based literature. We translate the examples from (Malamud and Stephenson 2014) and (Wiltschko et al. 2018) to German, slightly adapting them where necessary. Using these as well as our own examples, we collect felicity judgments from native speakers to find out how suitable the tag variants are in certain contexts. We group the results based on the notions of previous *knowledge*, effected *commitments*, and *judgment* of the conversation participants from (Malamud and Stephenson 2014) and correlate them with the semantic/pragmatic categories *speaker certainty* and *target of confirmation* defined in (Clausen and Scheffler 2020).

After discussing the previous studies by Malamud and Stephenson (2014) on English tag questions and Clausen and Scheffler (2020) on German tag questions in more detail, we present the design and results of our experiment. We show that this kind of experiment is a reliable source of data for semantic/pragmatic analyses. For example, our

results for German declarative clauses, which we tested as a control, match the predictions from the literature. Further, while the German question tags often overlap in their felicity ratings, we identify particular constraints for certain tags, thus demonstrating that (1) "tag question" is not a uniform category with uniform semantics and pragmatics, and (2) German question tag variants differ along other dimensions than the English RP and SP constructions. We conclude by discussing what properties of German tag questions have not yet been taken into account in the existing discourse model.

2 Background

In this section we outline the idea behind the discourse model by Malamud and Stephenson (2014) and its criteria for distinguishing between English RP and SP tag questions. Further, we provide an overview of the variability of German tag questions based on previous work.

2.1 English tag questions

The discourse model (Malamud and Stephenson 2014) is composed of the following components:

1. DC_X: for each participant X, the set of X's public discourse commitments.
2. *Table*: stack of issues to be resolved (the top issue first), where issues are represented as sets of propositions.
3. *Common Ground (CG)*: the set of propositions that all speakers are publicly committed to.
4. *Projected CGs*: a set of potential CGs giving possible resolution(s) of the top issue on the Table in the expected (canonical) next stage of the conversation; this 'next stage' is typically reached within the next few moves responding to the current move.
5. *Projected* DC_X: for the speaker and the hearer(s); allow for moves that give tentative commitments (by adding propositions to the speaker's projected, rather than present, commitments), or to offer the speaker's best guess of commitments of other participants (by adding to others' projected commitment sets).

In this model, commitments are crucial for the analysis of the RP and SP tags. Along with further criteria, such as whose judgment is at issue (speaker's or hearer's) and who has the knowledge about the expressed prejacent proposition, they provide for a clear distinction between the use cases of these tags.

Examples (1) and (2) from (Malamud and Stephenson 2014: 279) show how the felicity conditions for RP and SP tags attached to the same utterance change depending on the situational context.

(1) 'Blushing/Innuendo': A and B are gossiping. A doesn't know anything about B's neighbor. B says, blushing, 'You've GOT to see this picture of my new neighbor!' Without looking, A replies:

A: He's attractive, isn't he?
OK *A: He's attractive, is he?*

(2) 'Seeking agreement': A and B are discussing various traits of their mutual acquaintances. B says, 'I think Bill, more than anything else, is just a really nice guy'. A replies:

OK *A: (But) he's attractive too, isn't he?*
A: He's attractive too, is he?

A typical use case of English RP tags is when the interlocutors are establishing points of agreement (2), so that independent commitments of both speaker and hearer are at issue. Both speaker and hearer are informed, but the speaker is expressing an opinion and seeks agreement (the judgments of both speaker and hearer are at issue). The commitments of the speaker can be definite or tentative, where tentativeness is associated with some lack of confidence.

English SP tags on the other hand involve independent commitments of the hearer, and may or may not involve dependent commitments of the speaker (1). The hearer is informed, whereas the uninformed speaker is making a guess about potential hearer commitment. Hearer judgment is at issue.

2.2 German tag questions

Contrary to English, German does not have variable tags, but offers a large inventory of invariant tags instead (cf. Clausen and Scheffler 2020). The tags show a considerable overlap in use, but are also subject to constraints and display preferences for certain (situational/conversational) contexts. For example, a clear contrast between the tags *oder* and *ne* has been repeatedly mentioned in the literature (see e.g. König 2017; Clausen and Scheffler 2020).

Whereas *oder* is mostly used to express speaker uncertainty and requests confirmation, *ne* prefers contexts where the speaker is certain and therefore the hearer's reaction is not necessarily required. However, there are also contexts that are easily available for both (cf. Example (3) from (Clausen and Scheffler 2020: 2)).

(3) Emma and her family returned from a vacation. It is Monday morning, her mother comes to the kitchen and finds Emma watching movies instead of getting ready for school. Mother says:

Du musst heute nicht in die Schule, {*ne?* | *oder?*}
'You don't have to go to school today, right?'

Further, in the previous corpus study we have shown that tag questions expressing confirmation requests for acceptance of a command most frequently occur with the tag *ja* (cf. Clausen and Scheffler 2020: 24).

In the corpus study, we investigated the usage patterns of the five most common German tags based on the syntactic and pragmatic properties of the utterance a tag attaches to (the anchor). We identified that speaker certainty and target of confirmation are crucial in distinguishing between the types of German tags. Speaker certainty reflects the speaker's epistemic status regarding the expressed anchor proposition, i.e. either sure or unsure. The target of confirmation encodes the speaker's expectation for the hearer's reply to the tag question. It includes four types: *p is true* (hearer is asked to confirm whether the anchor proposition is true); *H believes p* (hearer is asked to confirm that they (also) believe the proposition, i.e. show agreement with it); *H perceives p* (hearer is asked to express their awareness of the proposition); *H accepts p* (hearer is asked to confirm that they accept the command expressed in the anchor).

In this study, we test whether the model by (Malamud and Stephenson 2014) is applicable to German tags. First, we conduct an experiment to test in which contexts different tags are (dis)preferred or may not occur at all. Based on the elicited judgments, we then determine the kinds of commitments different tags involve. To group different types of tag questions and contexts, we additionally adopt the categories from (cf. Clausen and Scheffler 2020) which play an important role in the differentiation of German tag variants.

3 Experiment

We conducted an experiment in which the felicity of German tag questions in certain contexts was assessed by native German speakers. We used the contexts from (Malamud and Stephenson 2014), which involve varying degrees of speaker and hearer commitment and knowledge. Additionally, we tested the contexts from (Wiltschko et al. 2018), where the use of declaratives and tag questions with Canadian *eh* is compared on the basis of speaker and hearer knowledge. We translated these examples into German and adapted them where necessary, replacing the utterances in question with the German tag questions. Additionally, we included two of our own examples.

Our experiment contained 20 items, each consisting of a description of a conversational situation (i.e., context) and an utterance produced by one of the conversation participants (all items are listed in Appendix A).

We tested every item in five variants, each time with a different ending: a period (declarative) or one of the common German question tags (*ja*, *ne*, *nicht*, or *oder*). Every participant saw each item only once, i.e. with only one of the five possible endings. To ensure this, we created five questionnaires containing the same utterances each with one of the five different endings, which we randomly distributed among participants (a Latin square design). Figure 1 shows an example of the participants' view of the experiment item WI-06 with the ending *ja* (see Section A.3 of Appendix A for translation). The same item with the other endings was shown to different participants.

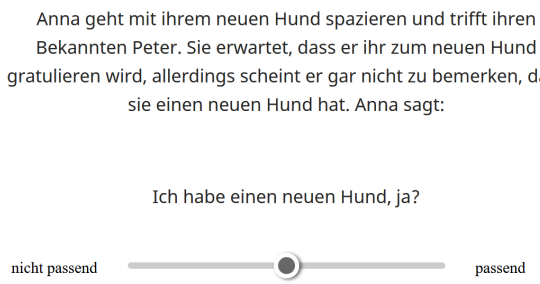

Figure 1: Example of the participants' view of the experiment item WI-06.

Along with the actual items, each questionnaire contained 20 fillers that were constructed analogously to the test items and consisted of a context and an utterance to be judged. For the fillers, we

used various sentence types, as well as tag questions with tags other than the four we are interested in (e.g. *nicht wahr*?), varying the utterances between well-fitting and more awkward.

The questionnaires were created using `_magpie`[1] and participants were recruited via the crowd-sourcing platform Prolific[2] and paid for participation. We selected the participants based on the following criteria: born and currently living in Germany, monolingually raised speakers of German, German nationality, no language related disorders. They were instructed to carefully read the contexts, paying attention to which of the interlocutors is the speaker of the utterance in question. The participants were asked to rate on a continuous scale how well each utterance fits into the given context. The scale ranged internally from zero to 100 which were displayed as *nicht passend* 'not fitting' and *passend* 'fitting' (see Figure 1). The items were presented in a randomized order one at a time. The participants could control how quickly they wanted to switch to the next item.[3] Once they made a decision, they could not go back and change it.

We obtained data from 149 participants.[4] We used MACE (Hovy et al. 2013) to detect unreliable annotators and excluded 21 submissions with MACE score below 0.5.[5] This left us with 128 submissions with the following number of judgments per questionnaire: 21, 23, 23, 29 and 32.

We represent the results per item aggregated over all participants as box plots. Within the results for each item, we perform pair-wise comparisons of the ending variants with the independent t-test[6] and measure the effect size of differences between the variants with Cohen's d[7] (Cohen 1988).

[1]Minimal Architecture for the Generation of Portable Interactive Experiments: https://github.com/magpie-ea

[2]https://www.prolific.co/

[3]Participants who took longer than the time allotted by the platform were automatically replaced by new participants. Additionally, we rejected submissions that were completed exceptionally fast.

[4]One participant experienced a technical problem with submission, so that their data is not available for us.

[5]MACE clusters annotators based on the overlap between their annotations, providing a quality score for individual annotators even in the absence of gold labels; https://github.com/dirkhovy/MACE.

[6]We used the t-test implementation in statsmodels package v0.11.1: https://www.statsmodels.org/stable/generated/statsmodels.stats.weightstats.ttest_ind.html

[7]We used the implementation in pingouin package v0.3.3: https://pingouin-stats.org/generated/pingouin.compute_effsize_from_t.html

4 Results

In this section, we summarize the results of our experiment. We find that, as expected from the discussion of English tag questions in the literature, the contexts and utterances form groups based on the involved speaker and hearer commitments, knowledge, and judgments. However, these distinctions alone are insufficient to differentiate the German tag variants. Adding speaker certainty and target of confirmation allows for clearer separation between the tag variants, although there are still several items that do not fit into any of the groups and require further study.

We describe the identified groups according to the involved commitments including one representative example per group. A full list of contexts, utterances, box plots, and statistics of participants' ratings is provided in Appendix A.

4.1 Speaker and hearer commitments

The utterances in this group involve independent commitments of both speaker and hearer. Both interlocutors are informed and establishing points of agreement. Therefore, both their judgments are at issue. The speaker is sure about their own opinion and wants to know whether the hearer has the same belief regarding the proposition (target of confirmation *H believes p*). We use item MS-03 (Example (4) and Figure 2) as representative of this group. Items MS-04, WI-05 and WI-11 are listed in Appendix A.

(4) MS-03: Eva and Laura are talking about their mutual friend. Laura says: "I think Mark is, first and foremost, a very nice guy.". Eva answers:

Aber er ist auch attraktiv{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'But he's attractive too{. | right?}'

In English, RP tags, *eh* and declaratives are fine in these contexts (cf. Malamud and Stephenson 2014; Wiltschko et al. 2018). Similarly, in our data, declaratives as well as *ne*, *nicht* and *oder* were rated high. However, *ja* received significantly lower ratings than the other tags (across all items tested in this group) and declaratives (in all items but MS-03). Note that the overall felicity rating for *ja* is still high, even though significantly lower than for the other tags. This example represents a classic use

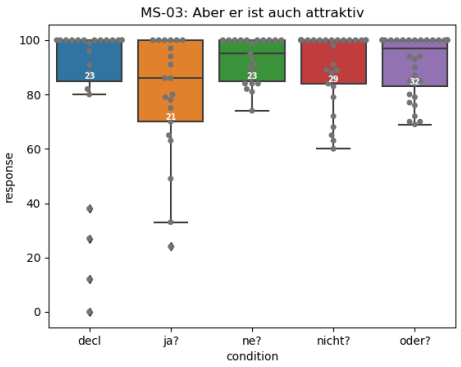

Figure 2: Ratings for item MS-03.

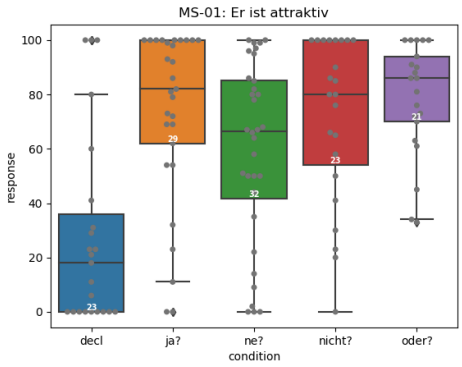

Figure 3: Ratings for item MS-01.

of tag questions – establishing agreement between interlocutors. Consequently, all tags are felicitous here.

4.2 Hearer commitments

This group comprises utterances in contexts where the speaker is uninformed and the hearer's commitment, judgment and knowledge are at issue. The speaker is unsure and wants to obtain confirmation regarding the truth of the expressed proposition (target of confirmation *p is true*). Item MS-01 serves as a representative example of this group (Example (5) and Figure 3). Items MS-02, WI-04 and WI-07 are listed in Appendix A.

(5) MS-01: Anne and Lena are chatting. Anne doesn't know anything about Lena's new neighbor. Lena says: "You HAVE to see this picture of my new neighbor!" and blushes. Anne answers without looking at the picture:

Er ist attraktiv{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'He is attractive{. | right?}'

In (5), the speaker Anne is uninformed about the physical appearance of Lena's neighbor, but is making a guess based on Lena's blushing. She wants Lena to confirm whether the neighbor actually is attractive (target of confirmation). The judgment involved would necessarily be Lena's (the hearer's), because the speaker has not seen the neighbor. Therefore, any (projected) commitment wrt. the neighbor's attractiveness must also be the hearer's.

In English, SP tags and *eh* are preferred over declaratives in these contexts. We observe the same tendency in our data: tag questions were rated higher than the declaratives (statistically significant in three items: MS-01, MS-02, WI-04). Among the tags, *oder* is preferred over *ne* (statistically significant in two items: MS-01, MS-02).

The item WI-07 is less straightforward. In this context, the speaker Ben knows that his friend Maria has wanted to adopt a dog for a long time, and when he finally sees her with a dog, he says "You have a new dog{. | right?}". Here, the declarative was rated relatively high (contrary to declaratives in other items in this group), and there are no significant differences between the tags. The felicity of the declarative in this context can be explained by the fact that Ben knew about the adoption plan and merely states the obvious. However, since he doesn't know this for sure, he still expects a reply, which is indicated by the high ratings of the tags.

4.3 Speaker commitments

In this group, the utterances are speaker centered: the speaker is informed and sure regarding the expressed proposition. Speaker's commitment and judgment are at issue, whereas the hearer is merely asked to express their awareness of the anchor proposition (target of confirmation *H perceives p*).

In the English variants of these examples, declaratives are felicitous. In our German test items, declaratives have continuously been rated higher than the variants with the tags (significant in all items). However, the distributions of the tag ratings are not homogeneous among the six items in this group. Rather, the items form three subgroups: In the first subgroup, the tags were rated consistently lower than declaratives (see Example (6) and

Figure 4 for WI-10 and Appendix A for ON-01).

(6) WI-10: Tom and Laura have decided to attend a lecture together. Tom came on time and has already heard the beginning of the lecture. As Laura arrives late, Tom says to her:

Dem kann man gar nicht folgen{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'It's not really possible to follow him{. | right?}'

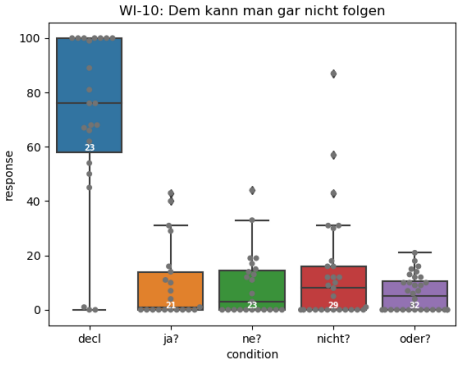

Figure 4: Ratings for item WI-10.

In the second subgroup, the tags received higher ratings compared to the first subgroup, but they display large variation along almost the whole scale, indicating dissent or uncertainty among participants (see Example (7) and Figure 5 for WI-08 and Appendix A for ON-02).

(7) WI-08: Elisa has organized a surprise party for Louis. A friend was meant to get Louis to the party at a certain time. As Louis opens the door, everyone shouts "Surprise!". Louis says:

Das ist vielleicht eine Überraschung{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'Well, that is a surprise{. | right?}'

In the third subgroup, *ne* is rated significantly higher than the other tags, even though with large variation. Additionally, in both items, *ja* is significantly better than *oder* (see Example (8) and Figure 6 for WI-03 and Appendix A for WI-06).

(8) WI-03: Eric is taking a stroll. Suddenly, a bird starts talking to him. At first he is confused. Then a woman approaches him and says:

Überraschung! Du bist bei "Versteckte Kamera" {. | *ja?* | *ne?* | *nicht?* | *oder?*}

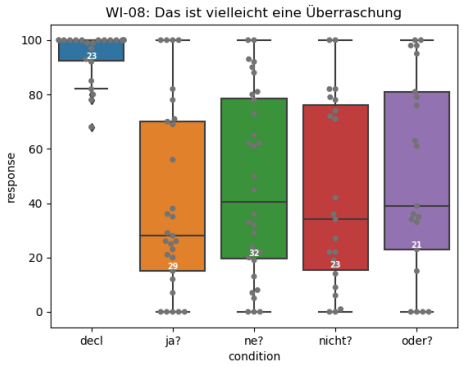

Figure 5: Ratings for item WI-08.

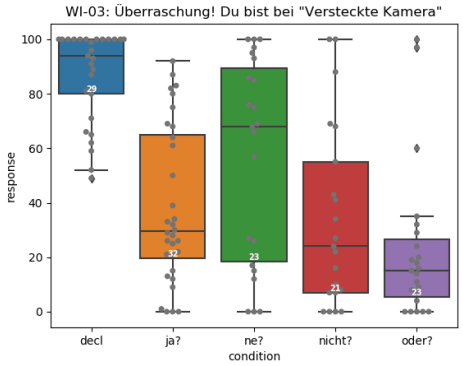

Figure 6: Ratings for item WI-03.

'Surprise! You're on "Candid Camera" {. | right?}'

We attribute these effects to the allocation of knowledge. In the first subgroup, the speaker conveys completely new information to the hearer, which is why the plain declarative is the best option. Since the hearer has no prior knowledge of the proposition at issue, which the speaker is just bringing up for the first time, all kinds of question tags are infelicitous.

In the second subgroup, there exists a certain amount of common ground about the issue raised by the speaker (the surprise for Louis planned by the hearers in WI-08). For this reason, the speaker can assume that the expressed proposition is not completely new to the hearer. This makes the use of tags more appropriate for some language users, who use them to ask the hearer for their opinion/confirmation.

In the third subgroup, there is also (visual) evidence (the weird bird incident in WI-03), for which

the utterance serves as a kind of explanation. Here, the speaker wishes not a confirmation or reply, but merely asks the hearer to take note of what has been said. The tags *ja* and, especially, *ne* are more strongly associated with this target of confirmation, *H perceives p*, than the other tags (cf. Clausen and Scheffler 2020: 26), and this is confirmed by the felicity ratings here, as well.

4.4 Directives

We tested the felicity of German tags in directives (see Examples (9) and (10) for items WI-01 and WI-02, respectively and Figure 7). These utterances are characterized by the speaker requesting confirmation from the hearer about the acceptance of the expressed command (target of confirmation *H accepts p*).[8] Consequently, the hearer's judgment is at issue. The question of commitments is not straightforward with this type of utterances, and they are not discussed in (Malamud and Stephenson 2014). However, in the subsequent work on the Chinese question tag *ba*, which can occur with imperatives, Ettinger and Malamud (2015) ascribe the commitments in directives to the hearer.

(9) WI-01: Lukas and Robin are students and roommates in a dorm. One evening, they are sitting on the couch and watching a movie. Lukas says:

Holst du mir mal ein Bier{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'Will you bring me a beer{. | right?}'

(10) WI-02: Tobias and Max are students and roommates in a dorm. One evening, they are sitting on the couch and watching a movie. Max gets up and heads to the kitchen. Tobias says:

Bringst du mir mal ein Bier mit{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'Will you bring me a beer{. | right?}'

Both items describe similar situations, where two fellow students are spending an evening in a dorm, probably the way they usually do. In both cases, one student is asking the other one to bring him a beer. The difference in WI-02 is that upon hearing the utterance Max is already on his way to the kitchen, and it is more likely that he will comply with it.

As Figure 7 shows, in both cases, declaratives and the tag *ja* were rated higher than the other tags. The use of the untagged variant seems to be the most preferred one when expressing a command in both situations. However, whereas in WI-02, where the hearer is already on his way to the kitchen, almost all participants agree, in WI-01, their consensus is somewhat lower. A possible reason is a lack of courtesy resulting with the use of a plain declarative in this context.

The felicity of *ja* in these items is not surprising. It is significantly better than *ne* (WI-01) and *nicht* and *oder* (both items). This confirms the findings in our corpus study that *ja* strongly correlates with directives (cf. Clausen and Scheffler 2020: 26). The situation with *ne* is more interesting – the ratings are widely spread across the middle range of the scale in WI-01 (both interlocutors are sitting on the couch). In WI-02 (the hearer is on the way to the kitchen), the use of *ne* becomes much more felicitous. It is significantly better than *nicht* in WI-01 and both *nicht* and *oder* in WI-02.

The ratings for the untagged and the *ne* utterances go in line with the English version of these examples, where the declarative is fine in both cases, but Canadian *eh* is felicitous only in WI-02 (cf. Wiltschko et al. 2018: 581,582). Since *ne* prefers situations where the speaker is sure (cf. Clausen and Scheffler 2020), it becomes more acceptable in WI-02, where the speaker has more certainty that his friend will bring him a beer, but wants it to be confirmed.

4.5 Remaining cases

There are four remaining items that did not fit into any of the described groups. For reasons of space, we discuss only two of the items here, but see Section A.5 in Appendix A for all items.

The items MS-06 and MS-07 (Examples (11) and (12) and Figure 8) both involve tentative speaker commitments and independent hearer commitments and are associated with speaker uncertainty. However, they differ with respect to the target of confirmation, as well as involved judgments and knowledge. In these examples, the speaker is making a guess about some issue they are unsure about, while the hearer could resolve this issue in their reply.

(11) MS-06: Tom and Ole are working at the store

[8] Although, strictly speaking, these utterances do not express commands, we adopt this phrasing from (Clausen and Scheffler 2020) to distinguish these utterances from the usual requests essentially encoded in all tag questions.

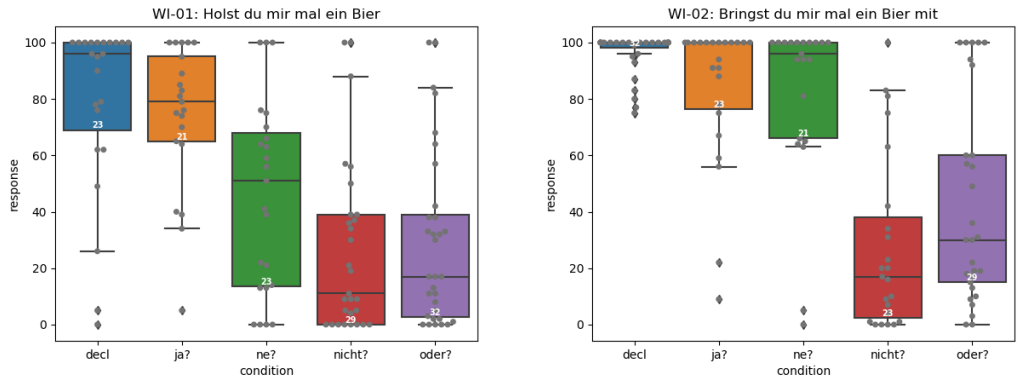

Figure 7: Ratings for items WI-01 and WI-02.

part time, they are sorting the paint buckets in two categories: "red" and "orange". Tom points at one of the buckets and asks: "How would you describe this color?". Ole says:

Es ist rot{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'It's red{. | right?}'

(12) MS-07: Maik is being questioned about the state capitals during an examination. The teacher says: "What is the capital of Hessen?'. Maik is not sure, but thinks it might be Wiesbaden. He answers:

Das ist Wiesbaden{. | *ja?* | *ne?* | *nicht?* | *oder?*}
'This is Wiesbaden{. | right?}'

In English, both utterances are felicitous with RP tags and declaratives (cf. Malamud and Stephenson 2014: 281,282), although declaratives fail to convey the uncertainty of the speaker. In German, we find that differences wrt. target of confirmation prevent them from being put into one group. In MS-06, the interlocutors strive to come to an agreement about an opinion, and both their judgments are at issue (target of confirmation *H believes p*). In MS-07, only the hearer's judgment and knowledge matter (target of confirmation *p is true*), whereas the speaker lacks knowledge.

In both items, declaratives and tags are mostly fine. However, in MS-06 *ja* is significantly worse than other options. In MS-07, *ne* was rated somewhat lower. The tag *oder* was rated consistently high and is significantly better than *ne* and *ja*. This is a typical use case for *oder*, where an uncertain speaker is seeking confirmation from an informed hearer.

4.6 Summary

Table 1 summarizes the distribution of commitments in the English and German variants of the examples discussed in this study.[9] In the most typical use of tags, i.e. establishing agreement where both speaker and hearer commitments are at issue, all German tags are felicitous. Only the tag *ja* is somewhat degraded here. With cases where only hearer commitments are at issue, again, all German tags can be used. However, the preferences for specific tag variants seem to depend on the nature of the context: the tags *ja*, *nicht* and *ne* become less felicitous in situations where the speaker is not just asking for confirmation, but is making a guess regarding the potential hearer commitments based on their behavior. Finally, in the contexts where only speaker commitments are at issue, the only available German tag is *ne*, and, in case of directives, also (and especially) *ja*.

This overview shows that based on the commitments alone we can confirm the difference between the tags *ne* and *oder* shown in our corpus study (Clausen and Scheffler 2020) as well as previous literature: *oder* is infelicitous unless hearer commitments are involved (S+H or H), and *ne* is virtually the only available tag when only speaker commitments are at issue. Also, the special preference for *ja* in directives is confirmed by the results of the experiment.

However, since *ne*, for example, can also be used where hearer commitments are involved, further differentiating criteria are needed to pin down the more subtle differences between the tags. In fact,

[9] Judgments for English reverse and same polarity tags (RP/SP) are taken from (Malamud and Stephenson 2014), and judgments for the Canadian *eh* are taken from (Wiltschko et al. 2018).

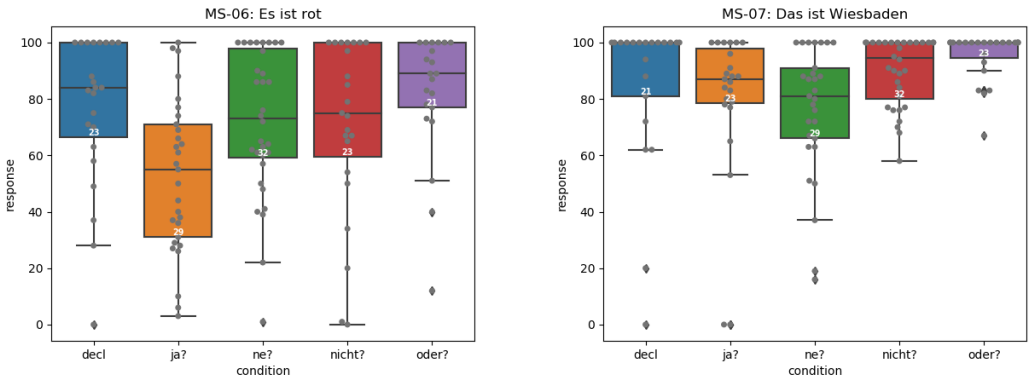

Figure 8: Ratings for items MS-06 and MS-07.

item	RP	SP	*eh*	*ja*	*ne*	*nicht*	*oder*
S+H							
MS-03	✓	#		✓	✓	✓	✓
MS-04	✓	#		✓?	✓	✓	✓
WI-05			✓	✓?	✓	✓	✓
WI-11			✓	✓?	✓	✓	✓
H							
MS-01	#	✓		✓	✓?	✓	✓
MS-02	#	✓		?	?	?	✓
WI-04			✓	✓	✓	✓	✓
WI-07			✓	✓	✓	✓	✓
S							
WI-03			#	#	✓?	#	#
WI-06			✓	#	✓?	#	#
WI-08			#	#?	#?	#?	#?
WI-10			#	#	#	#	#
ON-01				#	#	#	#
ON-02				?	?	?	?
S_{dir}							
WI-01			#	✓	?	#	#
WI-02			✓	✓	✓	#	#

Table 1: Commitments in English and German question tags for all of the studied items except for the remaining cases in Section 4.5: both speaker's and hearer's (S+H), hearer's (H), speaker's (S), and speaker's in directives (S_{dir}).

commitments seem to also not suffice to account for the use of the Canadian tag *eh*, which is fine both where RP tags as well as SP tags are used. Additionally, its felicity conditions vary within the group(s) of speaker commitments.

To resolve the unclear cases, several items need further testing: items WI-08 and ON-02 might be ambiguous and need to be reformulated to restrict the context interpretation.[10] Further, all unclear cases (marked with ✓? and #? in Table 1) should be retested with more participants.

5 Discussion

The method of an online experiment used in this study proved to be a reliable way of assessing the subtle differences in the use of tag variants (and other utterances). The high/low ratings for certain well-studied utterances were conform with the expected outcomes in most cases. For example, declaratives were consistently rated high where speaker commitments are involved. In typical situations, where the tags are used to establish agreement (e.g. item MS-03), the experiment results followed the patterns demonstrated in previous work. This indicates general diligence of participants' replies, that can additionally be filtered for an even higher reliability with an automatic method (MACE).

Adopting various discourse contexts from previous studies on English allowed us to test the use of German tagged utterances under similar conditions. Malamud and Stephenson (2014) differentiate English RP and SP tags based primarily on speaker/hearer commitments. RP tags are felicitous where both speaker and hearer commitments are at issue, and a speaker can make either definite or tentative commitment. SP tags are used in utterances where only hearer commitments are at issue. For German, it is not possible to make a clear distinction between different tag variants based on the involved commitments, which is why we also make use of speaker certainty and target of confirmation. We find that the distinction sure/unsure roughly corresponds to definite/tentative commitments. All our items where speakers make tentative commitments coincide with 'unsure' and cases involving definite

[10] We thank our reviewers for pointing to this issue.

commitments with 'sure'. Target of confirmation is not tied to any other category. Items with the same commitments, such as MS-06 and MS-07, where in English RP tags are felicitous, involve different targets of confirmation, and the German tag ratings differ between these items (see Section 4.5).

Due to a large degree of overlap among the German tags, no direct correspondences to the English RP tags and SP tags can be established. Both English tag variants are matched (to varying degrees) by all German tags we tested. However, it stands out that *ja* is the least felicitous tag in the contexts where RP tags are used. Whereas the judgments for the other tags show good consensus among the raters, *ja* displays considerably more variation.

We find that certain German tags behave consistently regarding commitments: *oder* is felicitous whenever hearer commitments are at issue (both speaker and hearer or only hearer). Overall, it appears that when no hearer commitments are involved, the use of German tags becomes much more restricted.

Finally, we observe that many items show great inter-speaker variation in ratings, indicated by the long boxes in plots. This may point to participants' different interpretations of the contexts, or to differences in opinions on the uses of tags (when we observe a 'gap' between the ratings that seem to split the opinions into positive vs. negative camps, e.g. *ne* in WI-03). Alternatively, even if participants have the same interpretations, these might be vague or they might be uncertain about them.

As future work, we plan to set up a follow up experiment with a larger number of participants to retest the problematic and less clear items using reformulated as well as new contexts. Further, we will look into adapting the described discourse model to the German tags. There are several things that need to be considered. First, there exist utterances not discussed in (Malamud and Stephenson 2014) – with the speaker centered contexts and target of confirmation *H perceives p*. Second, directives expressing requests for acceptance of commands might involve a different kind of speaker commitments. Finally, and perhaps most importantly, it needs to be tested whether the variability of German tags can be explained by the existing model.

Funding Information

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 317633480, SFB 1287, project A03.

References

Yulia Clausen and Tatjana Scheffler. 2020. A corpus-based analysis of meaning variations in German tag questions: Evidence from spoken and written conversational corpora. *Corpus Linguistics and Linguistic Theory*, pages 1–31. Published online ahead of print.

Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences*, 2 edition. Lawrence Earlbaum Associates, Hillsdale, NJ.

Allyson Ettinger and Sophia A. Malamud. 2015. Mandarin utterance-final particle ba in the conversational scoreboard. In *Proceedings of Sinn und Bedeutung*, volume 19, pages 232–251.

Donka Farkas and Kim Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics*, 27.

Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard Hovy. 2013. Learning whom to trust with MACE. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1120–1130, Atlanta, Georgia. Association for Computational Linguistics.

Katharina König. 2017. Question tags als Diskursmarker? – Ansätze zu einer systematischen Beschreibung von *ne* im gesprochenen Deutsch. In H. Blühdorn, A. Deppermann, H. Helmer, and T. Spranz-Fogasy, editors, *Diskursmarker im Deutschen. Reflexionen und Analysen*, pages 233–258. Verlag für Gesprächsforschung, Göttingen.

Manfred Krifka. 2015. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. In *Proceedings of SALT 25*, 0, pages 328–345.

Sophia A. Malamud and Tamina Stephenson. 2014. Three Ways to Avoid Commitments: Declarative Force Modifiers in the Conversational Scoreboard. *Journal of Semantics*, 32(2):275–311.

Brian Reese and Nicholas Asher. 2007. Prosody and the interpretation of tag questions. In *Proceedings of Sinn und Bedeutung*, volume 11, pages 448–462.

Martina Wiltschko, Derek Denis, and Alexandra D'Arcy. 2018. Deconstructing variation in pragmatic function: A transdisciplinary case study. *Language in Society*.

A Experimental items

The items labeled with MS have been adapted from (Malamud and Stephenson 2014), the items labeled with WI have been adapted from (Wiltschko et al. 2018), the items labeled with ON were created by the authors of the paper. We present the items in groups according to the description in the paper. For each item, we present the German context and utterance with their translation, and the box plot of results. Thresholds and effect sizes for significant differences between variants are summarized for each item group in a table at the end of the group.

A.1 Speaker and hearer commitments group

MS-03 Eva und Laura diskutieren über ihren gemeinsamen Bekannten. Laura sagt: "Ich finde Mark ist, vor allem, ein sehr netter Typ.". Eva antwortet:

Aber er ist auch attraktiv{. | ja? | ne? | nicht? | oder?}

'Eva and Laura are talking about their mutual friend. Laura says: "I think Mark is, first and foremost, a very nice guy.". Eva answers:

But he's attractive too{. | right?}'

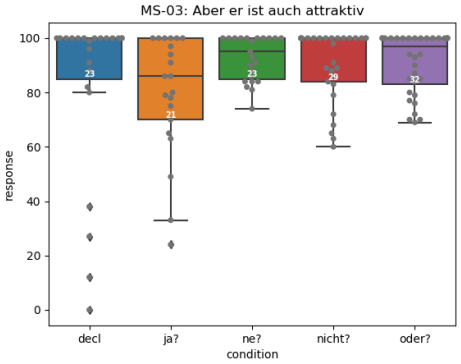

Figure 9: Ratings for item MS-03.

MS-04 Marie und Antonia sind Grundschullehrerinnen. Sie besprechen, was jeder von ihren gemeinsamen Schülern bei dem Tag der offenen Tür vorführen wird. Marie schlägt vor: "Nina könnte ein Gedicht vortragen, weil sie das gut kann.". Antonia sagt:

Aber sie tanzt auch schön{. | ja? | ne? | nicht? | oder?}

'Marie and Antonia are primary school teachers. They are talking about what each of their mutual students will present at the open day at school. Marie suggests: "Nina could read a poem, since she's good at it." Antonia says:

But she dances well too{. | right?}'

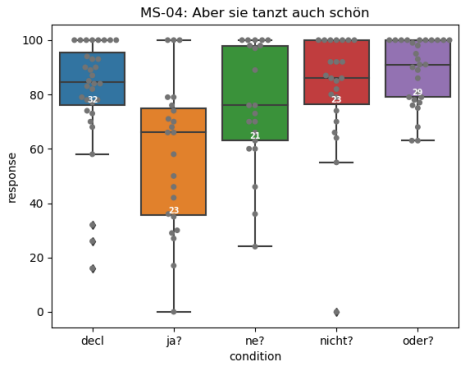

Figure 10: Ratings for item MS-04.

WI-05 Emma und Simon haben zusammen im Kino einen Film gesehen. Als sie das Kino verlassen, sagt Emma:

Der Film war gut{. | ja? | ne? | nicht? | oder?}

'Emma and Simon have seen a movie together. As they leave the theatre, Emma says:

The movie was good{. | right?}'

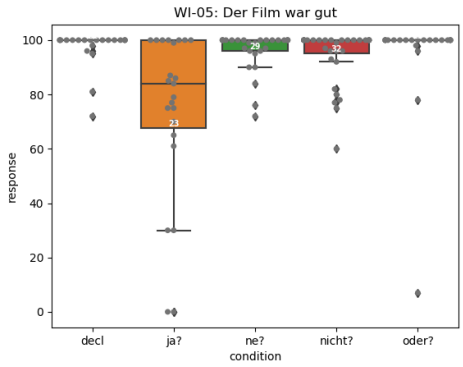

Figure 11: Ratings for item WI-05.

WI-11 Iris und Liam sitzen in einem Vortrag. Iris versteht nicht, worum es geht. Sie guckt Liam an – er sieht ebenso verwundert aus. Iris sagt:

Dem kann man gar nicht folgen{. | ja? | ne? | nicht? | oder?}

'Iris and Liam are attending a lecture. Iris does not understand what it is about. She looks at Liam – he looks equally bewildered. Iris says:

It's not really possible to follow him{. | right?}'

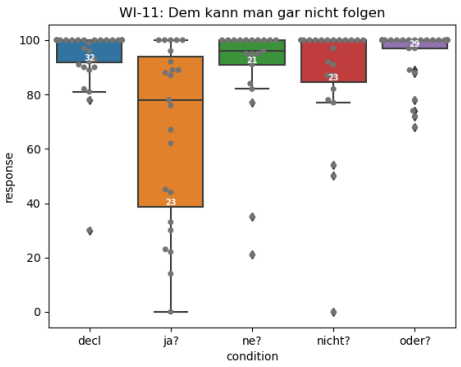

Figure 12: Ratings for item WI-11.

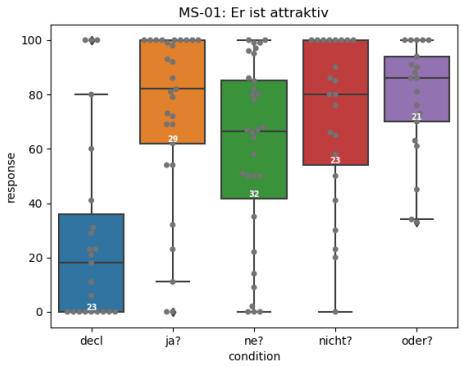

Figure 13: Ratings for item MS-01.

	item ID	*t*	*p*	df	*d*
decl vs. *ja?*	MS-04	3.55	=.001	53	1.0
	WI-05	3.35	=.002	42	1.0
	WI-11	4.25	<.001	53	1.2
ne? vs. *ja?*	MS-03	2.63	=.012	42	0.8
	MS-04	2.48	=.017	42	0.7
	WI-05	3.75	<.001	50	1.0
	WI-11	2.61	=.013	42	0.8
nicht? vs. *ja?*	MS-03	2.11	=.040	48	0.6
	MS-04	3.38	=.002	44	1.0
	WI-05	3.49	=.001	53	1.0
	WI-11	2.37	=.022	44	0.7
oder? vs. *ja?*	MS-03	2.43	=.019	51	0.7
	MS-04	5.53	<.001	50	1.5
	WI-05	2.71	=.010	44	0.8
	WI-11	4.45	<.001	50	1.2

Table 2: Significance tests: t-test and Cohen's *d* for the group **speaker and hearer commitments**.

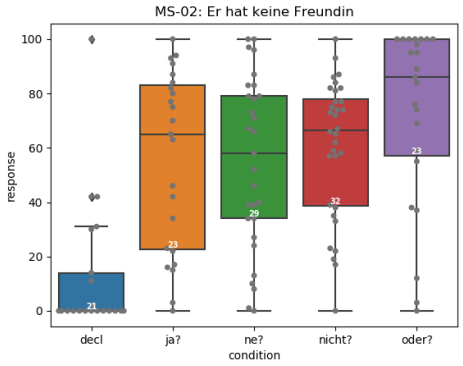

Figure 14: Ratings for item MS-02.

A.2 Hearer commitments group

MS-01 Anne und Lena unterhalten sich. Anne weiß noch nichts über Lenas neuen Nachbarn. Lena sagt: "Du MUSST dir mal dieses Bild von meinem neuen Nachbarn ansehen!" und wird dabei rot. Anne antwortet, ohne auf das Bild zu schauen:

Er ist attraktiv{. | ja? | ne? | nicht? | oder?}

'Anne and Lena are chatting. Anne doesn't know anything about Lena's new neighbor. Lena says: "You HAVE to see this picture of my knew neighbor!" and blushes. Anne answers without looking at the picture:

He is attractive{. | right?}'

MS-02 Sarah und Julia unterhalten sich. Sarah weiß noch nichts über Julias neuen Nachbarn. Julia sagt: "Du MUSST meinen neuen Nachbarn kennenlernen" und wird dabei rot. Sarah sagt:

Er hat keine Freundin{. | ja? | ne? | nicht? | oder?}

'Sarah and Julia are chatting. Sarah doesn't know anything about about Julia's new neighbor. Julia says: "You HAVE to meet my knew neighbor!" and blushes. Sarah says:

He doesn't have a girlfriend{. | right?}'

WI-04 David und Finn sitzen in einer Bar und tauschen sich über die letzten Neuigkeiten aus. Finn glaubt David bei der gestrigen Episode der "Versteckten Kamera" gesehen zu haben. Er möchte wissen, ob es wirklich David war und sagt:

Du warst gestern bei "Versteckte Kamera"{. | ja? | ne? | nicht? | oder?}

'David and Finn are sitting in a bar and discussing the latest news. Finn believes he saw David on last night's episode of "Candid Camera". He wants to know whether it really was David and says:

You were on "Candid Camera" yesterday{. | right?}'

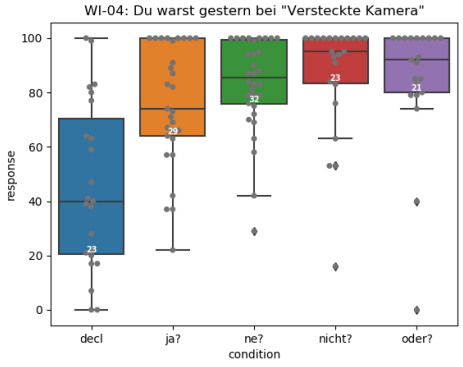

Figure 15: Ratings for item WI-04.

WI-07 Bei einem Spaziergang trifft Sven seine Bekannte Maria. Maria hat ihren Hund dabei. Sven weiß, dass sie schon seit langem einen Hund adoptieren wollte. Er nimmt an, dass sie das nun endlich getan hat. Sven sagt:

Du hast einen neuen Hund{. | ja? | ne? | nicht? | oder?}

'During a stroll Sven meets his acquaintance Maria. Maria has her dog with her. Sven knows that she's been wanting to adopt a dog for a while. He assumes that she has finally done so. Sven says:

You have a new dog{. | right?}'

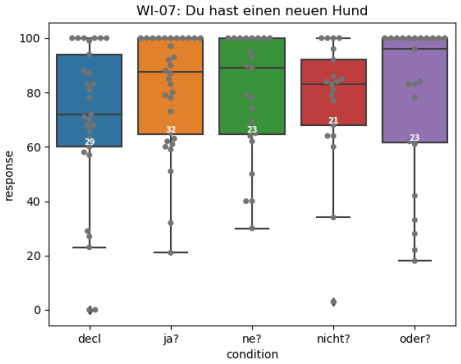

Figure 16: Ratings for item WI-07.

	item ID	*t*	*p*	df	*d*
ja? vs. decl	MS-01	4.86	<.001	50	1.4
	MS-02	4.82	<.001	42	1.5
	WI-04	4.09	<.001	50	1.1
ne? vs. decl	MS-01	3.33	=.002	53	0.9
	MS-02	5.03	<.001	48	1.4
	WI-04	5.74	<.001	53	1.6
nicht? vs. decl	MS-01	4.48	<.001	44	1.3
	MS-02	6.74	<.001	51	1.9
	WI-04	5.27	<.001	44	1.6
oder? vs. decl	MS-01	5.66	<.001	42	1.7
	MS-02	6.71	<.001	42	2.0
	WI-04	4.68	<.001	42	1.4
oder? vs. *ne?*	MS-01	2.32	=.025	51	0.7
	MS-02	1.99	=.052	50	0.6

Table 3: Significance tests: t-test and Cohen's *d* for the group **hearer commitments**.

A.3 Speaker commitments group

WI-10 Tom und Laura haben beschlossen, zusammen zu einem Vortrag zu gehen. Tom war rechtzeitig da und hat bereits den Anfang des Vortrags angehört. Als Laura verspätet reinkommt, sagt Tom zu ihr:

Dem kann man gar nicht folgen{. | ja? | ne? | nicht? | oder?}

'Tom and Laura have decided to attend a lecture together. Tom came on time and has already heard the beginning of the lecture. As Laura arrives late, Tom says to her:

It's not really possible to follow him{. | right?}'

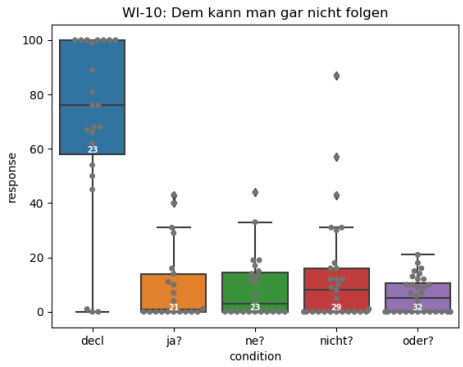

Figure 17: Ratings for item WI-10.

ON-01 Tanja war im Kino und hat einen neuen Film gesehen. Als sie sich später am Abend mit Jana trifft, die diesen Film noch nicht gesehen hat, sagt Tanja:

Dieser Film hat einen Oscar gewonnen und der war auch total gut{. | ja? | ne? | nicht? | oder?}

'Tanja was at the theatre and has seen a new movie. As she later meets up with Jana, who hasn't seen this movie yet, Tanja says:

This movie has won an Oscar and it was very good indeed{. | right?}'

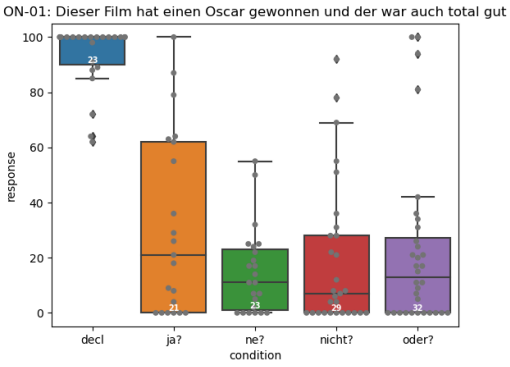

Figure 18: Ratings for item ON-01.

WI-08 Elisa hat für Louis eine Überraschungsparty organisiert. Ein Freund sollte Louis zu einer bestimmten Uhrzeit zu der Party bringen. Als Louis die Tür aufmacht, schreien alle auf einmal "Überraschung!". Louis sagt:

Das ist vielleicht eine Überraschung{. | ja? | ne? | nicht? | oder?}

'Elisa has organized a surprise party for Louis. A friend was meant to get Louis to the party at a certain time. As Louis opens the door, everyone shouts "Surprise!". Louis says:

Well, that is a surprise{. | right?}'

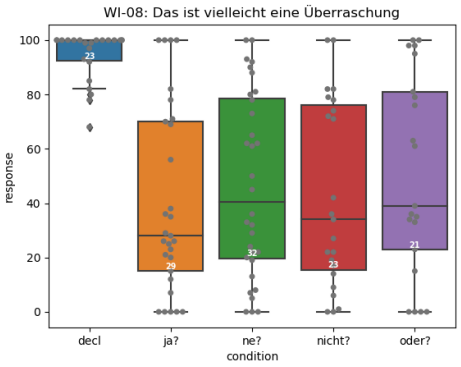

Figure 19: Ratings for item WI-08.

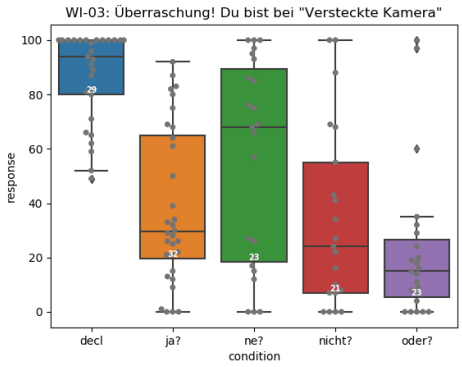

Figure 21: Ratings for item WI-03.

ON-02 Fabian hat sich früher um die Pflanzen im Garten seiner Eltern gekümmert. Jetzt lebt er in einer kleinen Wohnung und möchte dort auf dem Balkon etwas pflanzen. Dafür hat er einige pflegeleichte Pflanzensorten gekauft. Am Telefon mit seinen Eltern sagt Fabian:

Die brauchen natürlich nicht so viel Arbeit{. | ja? | ne? | nicht? | oder?}

'Fabian used to take care of the plants in his parents' garden. Now he lives in a small apartment and wants to plant something on the balcony. For this he bought some easy-care plant varieties. On the phone with his parents Fabian says:

They obviously don't need that much work{. | right?}'

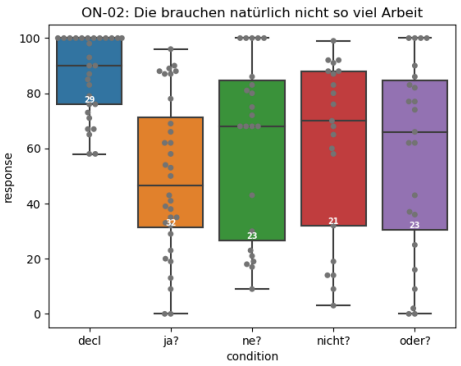

Figure 20: Ratings for item ON-02.

WI-03 Eric macht einen Spaziergang. Plötzlich spricht ihn ein Vogel an. Erst ist er völlig verwirrt. Dann kommt eine Frau zu ihm und sagt:

Überraschung! Du bist bei "Versteckte Kamera"{. | ja? | ne? | nicht? | oder?}

'Eric is taking a stroll. Suddenly, a bird starts talking to him. At first he is confused. Then a woman approaches him and says:

Surprise! You're on "Candid Camera"{. | right?}'

WI-06 Anna geht mit ihrem neuen Hund spazieren und trifft ihren Bekannten Peter. Sie erwartet, dass er ihr zum neuen Hund gratulieren wird, allerdings scheint er gar nicht zu bemerken, dass sie einen neuen Hund hat. Anna sagt:

Ich habe einen neuen Hund{. | ja? | ne? | nicht? | oder?}

'Anna goes for a walk with her new dog and runs into her acquaintance Peter. She expects him to congratulate her on getting a new dog, however, he seems not to notice, that she has a new dog. Anna says:

I have a new dog{. | right?}'

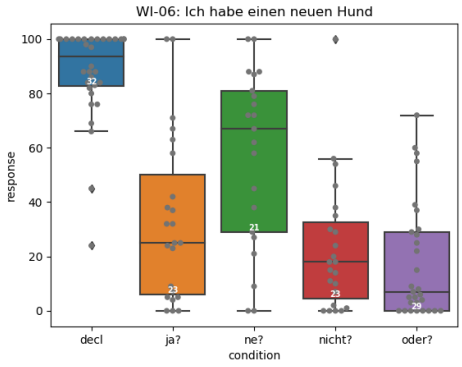

Figure 22: Ratings for item WI-06.

A.4 Directives group

WI-01 Lukas und Robin sind Studenten und Mitbewohner einer WG. An einem Abend sitzen sie auf der Couch und schauen einen Film. Lukas sagt:

Holst du mir mal ein Bier{. | ja? | ne? | nicht? | oder?}

'Lukas and Robin are students and roommates in a dorm. One evening, they are sitting on the couch and watching a movie. Lukas says:

Will you bring me a beer{. | right?}'

	item ID	*t*	*p*	df	*d*
decl vs. *ja?*	WI-03	7.95	<.001	59	2.0
	WI-06	8.36	<.001	53	2.3
	WI-08	7.35	<.001	50	2.1
	WI-10	7.73	<.001	42	2.3
	ON-01	8.45	<.001	42	2.6
	ON-02	6.32	<.001	59	1.6
decl vs. *ne?*	WI-03	3.99	<.001	50	1.1
	WI-06	4.46	<.001	51	1.3
	WI-08	6.65	<.001	53	1.8
	WI-10	8.35	<.001	44	2.5
	ON-01	19.3	<.001	44	5.7
	ON-02	3.70	<.001	50	1.0
decl vs. *nicht?*	WI-03	7.30	<.001	48	2.1
	WI-06	11.6	<.001	53	3.2
	WI-08	6.95	<.001	44	2.0
	WI-10	7.55	<.001	50	2.1
	ON-01	12.6	<.001	50	3.5
	ON-02	3.77	<.001	48	1.1
decl vs. *oder?*	WI-03	10.3	<.001	50	2.9
	WI-06	14.0	<.001	59	3.6
	WI-08	5.48	<.001	42	1.7
	WI-10	10.7	<.001	53	2.9
	ON-01	10.7	<.001	53	2.9
	ON-02	4.03	<.001	50	1.1
ne? vs. *ja?*	WI-03	1.96	=.055	53	0.5
	WI-06	2.51	=.016	42	0.8
	ON-01	-2.17	=.036	42	-0.7
ne? vs. *nicht?*	WI-03	2.01	=.051	42	0.6
	WI-06	4.00	<.001	42	1.2
ne? vs. *oder?*	WI-03	3.42	=.001	44	1.0
	WI-06	5.16	<.001	48	1.5
ja? vs. *oder?*	WI-03	2.04	=.046	53	0.6
	WI-06	2.14	=.037	50	0.6

Table 4: Significance tests: t-test and Cohen's *d* for the group **speaker commitments**.

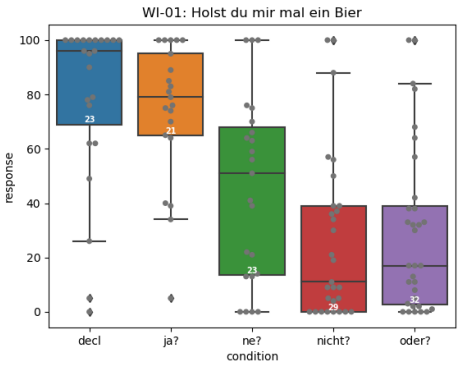

Figure 23: Ratings for item WI-01.

WI-02 Tobias und Max sind Studenten und Mitbewohner einer WG. An einem Abend sitzen sie auf der Couch und schauen einen Film. Max steht auf und geht in Richtung Küche. Tobias sagt:

Bringst du mir mal ein Bier mit{. | ja? | ne? | nicht? | oder?}

'Tobias and Max are students and roommates in a dorm. One evening, they are sitting on the couch and watching a movie. Max gets up and heads to the kitchen. Tobias says:

Will you bring me a beer{. | right?}'

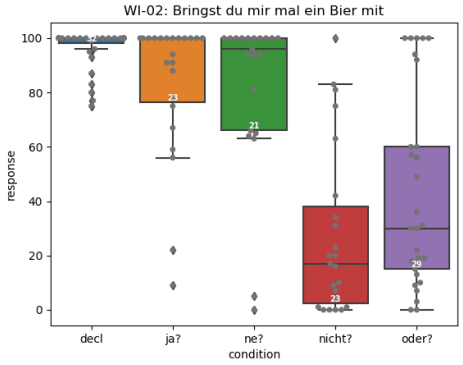

Figure 24: Ratings for item WI-02.

A.5 Remaining cases

MS-05 Sophia und Lisa unterhalten sich. Lisa hat den Nachbarn von Sophia noch nie getroffen und fragt: "Was hältst du von deinem neuen Nachbarn?". Sophia ist sich nicht sicher, was genau damit gemeint ist: seine Tauglichkeit als Nachbar oder als mögliches Date. Sophia sagt:

Er ist attraktiv{. | ja? | ne? | nicht? | oder?}

'Sophia and Lisa are talking. Lisa has not met Sophia's neighbor yet and asks: "What do you think of your new neighbor?". Sophia is not sure, what exactly is meant by that: his neighborliness or suitability for dating. Sophia says:

He's attractive{. | right?}'

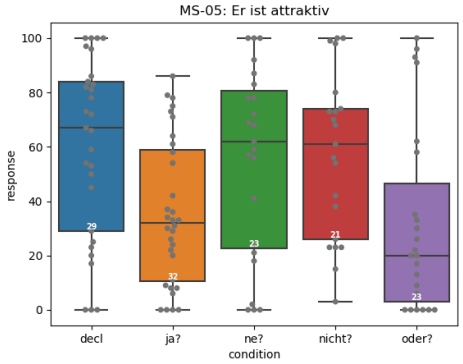

Figure 25: Ratings for item MS-05.

	t	*p*	df	*d*
decl vs. *ja?*	3.18	=.002	59	0.8
decl vs. *oder?*	3.00	=.004	50	0.8
ne? vs. *ja?*	2.35	=.022	53	0.6
ne? vs. *oder?*	2.29	=.027	44	0.7
nicht? vs. *ja?*	2.71	=.009	51	0.8
nicht? vs. *oder?*	2.58	=.013	42	0.8

Table 5: Significance tests: t-test and Cohen's *d* for item MS-05.

MS-06 Tom und Ole haben einen Nebenjob in einem Geschäft und sortieren gerade Farbeimer in zwei Kategorien: "rot" und "orange". Tom zeigt auf einen Eimer mit orange-roter

Farbe und fragt: “Wie würdest du diese Farbe bezeichnen?”. Ole sagt:

Es ist rot{. | ja? | ne? | nicht? | oder?}

‘Tom and Ole are working at the store part time, they are sorting the paint buckets in two categories: “red” and “orange”. Tom points at one of the buckets and asks: “How would you describe this color?”. Ole says:

It’s red{. | right?}’

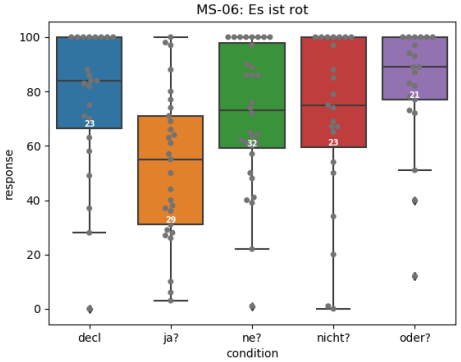

Figure 26: Ratings for item MS-06.

	t	*p*	df	*d*
decl vs. *ja?*	3.15	=.003	50	0.9
ne? vs. *ja?*	2.79	=.007	59	0.7
nicht? vs. *ja?*	2.21	=.031	50	0.6
oder? vs. *ja?*	3.99	<.001	48	1.1

Table 6: Significance tests: t-test and Cohen’s *d* for item MS-06.

MS-07 Maik wird bei einer Prüfung nach den Landeshauptstädten gefragt. Der Lehrer sagt: “Was ist die Hauptstadt von Hessen?”. Maik ist sich nicht sicher, denkt aber, dass es Wiesbaden sein könnte. Er antwortet:

Das ist Wiesbaden{. | ja? | ne? | nicht? | oder?}

‘Maik is being questioned about the state capitals during an examination. The teacher says: “What is the capital of Hessen?’. Maik is not sure, but thinks it might be Wiesbaden. He answers:

This is Wiesbaden{. | right?}’

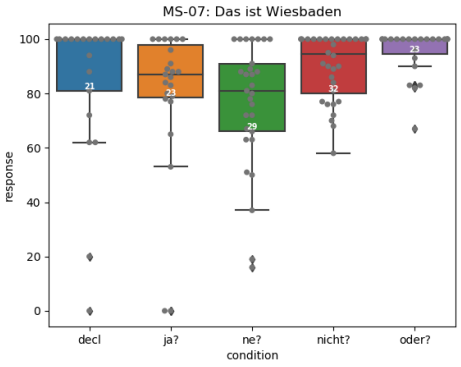

Figure 27: Ratings for item MS-07.

	t	*p*	df	*d*
nicht? vs. *ne?*	2.92	=.005	59	0.7
oder? vs. *ne?*	3.77	<.001	50	1.1
oder? vs. *ja?*	2.66	=.011	44	0.8

Table 7: Significance tests: t-test and Cohen’s *d* for item MS-07.

WI-09 Ines hat für Ben eine Überraschungsparty organisiert. Ein Freund sollte Ben zu einer bestimmten Uhrzeit zu der Party bringen. Als Ben die Tür aufmacht, schreien alle auf einmal “Überraschung!”. Ines sagt:

Das ist vielleicht eine Überraschung{. | ja? | ne? | nicht? | oder?}

‘Ines has organized a surprise party for Ben. A friend was meant to get Ben to the party at a certain time. As Ben opens the door, everyone shouts “Surprise!”. Ines says:

Well, that is a surprise{. | right?}’

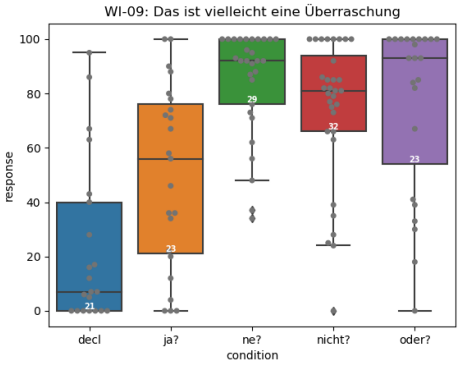

Figure 28: Ratings for item WI-09.

	t	*p*	df	*d*
ja? vs. decl	2.71	=.010	42	0.8
ne? vs. decl	8.68	<.001	48	2.5
nicht? vs. decl	6.38	<.001	51	1.8
oder? vs. decl	5.62	<.001	42	1.7
ne? vs. *ja?*	4.69	<.001	50	1.3
nicht? vs. *ja?*	2.97	=.005	53	0.8
oder? vs. *ja?*	2.75	=.009	44	0.8

Table 8: Significance tests: t-test and Cohen’s *d* for item WI-09.