



The prosodic marking of rhetorical questions in German

Daniela Wochner, Jana Schlegel, Nicole Dehé, Bettina Braun

Department of Linguistics, University of Konstanz, Germany

{Daniela.Wochner, Jana.2.Schlegel, Nicole.Dehe, Bettina.Braun}@uni-konstanz.de

Abstract

This study investigates the prosody of rhetorical questions in German. In an interaction study we examined how speakers use boundary tones, pitch, duration and voice quality to mark syntactically ambiguous questions as rhetorical or information-seeking. To this end, speakers produced identical interrogatives (polar and wh-questions) in rhetorical and information-seeking contexts. The results show that, phonologically, rhetorical questions end in low boundary tones more often, but the difference is minimal (5%), especially compared to the effect of interrogative type on boundary tones (polar questions are rising and wh-questions are falling in more than 90% of the cases, respectively). Phonetically, rhetorical questions are characterized by longer utterance durations and lower initial pitch. Furthermore, the first constituent (wh-word or verb) is produced with a softer voice compared to information-seeking questions. Overall the results show that there are prosodic differences between RQs and ISQs for both polar and wh-questions. There seems to be a tendency for rhetoricity to be signalled early-on in the utterance in both polar and wh-questions. Furthermore, the phonetic cues for disambiguation are stronger in wh-questions than in polar questions.

Index Terms: rhetorical questions, prosody, pragmatics, intonation, interrogativity, German

1. Introduction

Rhetorical questions (RQs) have been described as a hybrid type of utterances [1]. They have the syntactic surface structure of an interrogative but may be used as an assertion (e.g., [2, 3] [4]) of the opposite polarity of their interrogative structure [4]. E.g., *Did he lift a finger to help you?* suggests that he did not. This leads to a mismatch between form and function [5]. Unlike information-seeking questions (ISQs), which elicit answers, RQs tend to seek affirmation. Some rhetorical questions contain lexical triggers, such as the negative polarity item "lift a finger" ([4]), but this need not be the case. For instance, the plain interrogative utterance *Does she like chocolate?* can likewise be used as an RQ, given the appropriate context and/or prosody. Mostly, interlocutors do not have problems recognizing questions as rhetorical or information-seeking. The recognition of the correct illocutionary force may be facilitated by the discourse situation and/or the prosodic realization of the interrogative. This paper focuses on the prosodic realization of RQs in comparison to ISQs, an area where only little research is available so far (some of which provides conflicting results). For instance, in her paper on the semantics of RQs, [4] proposes falling intonation contours for RQs and ascribes this pattern to their assertive function. However, results of studies

for English show that RQs may also have rising intonation contours (e.g., [6], [7]), which suggests that boundary tones may not be such a crucial means for the distinction between RQs and ISQs, at least not for English. Possibly, boundary tones may be more relevant to signal discourse cohesion, turn-taking, etc. ([8], [9]). Recently, [10] argued that evaluative questions (assessments that seek agreement, e.g., "Isn't this just delicious?" and RQs that do not seek an answer, e.g., "Are you nuts?") differ from ISQs in that they have higher initial pitch compared to a speakers' median pitch (in five out of ten tested languages, not including German). However, it is not stated how many RQs were included in the corpus and it is therefore an open question whether RQs consistently have this property. A related line of research tested how interrogativity is marked in questions with declarative syntax (e.g., *Peter went to Paris?*), and provide findings that may be relevant for the distinction between RQs (no interrogative force) and ISQs (interrogative force) as well: a) Final boundary tones are high in questions and low in statements ([11], see also [4]). b) Pitch range is higher in questions than in statements already in the prenuclear region (e.g., [12]) c) Pitch scaling is a crucial feature for the strength of interrogative force. The higher the pitch scaling, the more likely it is that the utterance has an interrogative function [13] (e.g., for inverted questions and conditionals). d) Utterance duration is shorter in questions than in statements [14].

To analyze the prosody and intonation of RQs and ISQs in German, we conducted a production experiment, which was designed as a dialogue. The experiment tested how illocution type (ISQ vs. RQ, manipulated within-subjects) affects the prosodic realization of interrogatives in polar and wh-questions in different positions in a speaker's turn (turn-medial and turn-final).

Based on the literature on English RQs (e.g., [6], [7]) we did not predict the boundary tone to be a crucial marker to rhetoricity. Instead, we predicted a smaller pitch range for RQs compared to ISQs in the prenuclear region. Furthermore, we hypothesized longer durations for RQs than for ISQs. Additionally, we compared acoustic manifestations of voice quality in the first constituent of the target-questions (wh-word and verb).

2. Experiment

Participants were tested pairwise in speaker (S)-addressee (A) pairs. S uttered interrogatives that were presented in rhetorical and information-seeking contexts (RQs vs. ISQs). This allowed us to compare the prosodic features S used to signal the two illocution types in identical sentences. Since boundary tones have been mentioned as a distinctive feature for RQs (but also for discourse organization), we placed the target interrogatives in turn-final and turn-medial position. Consequently, S produced target sentences in the eight conditions shown in Table 1. A was also presented with

contexts, but they were less detailed than those for S. In order to simulate a dialogue situation, A's task was to react to S's utterance by choosing one of two replies, one compatible with a rhetorical, the other with an information-seeking illocution.

turn position	polar question		turn position	wh-question	
final	ISQ	RQ	final	ISQ	RQ
medial	ISQ	RQ	medial	ISQ	RQ

Table 1. *Summary of experimental conditions.*

2.1. Methods

2.1.1. Materials

We selected 39 wh-questions and 39 polar questions that were syntactically and lexically ambiguous between a rhetorical and an information-seeking reading. They all contained the particle "denn" which is frequently used in both RQs and ISQs in German [15]. All target interrogatives and contexts were constructed such that the nuclear accent was expected on the final object (e.g., *Domino* in *Wer spielt denn Domino?* 'Who plays Domino?'). These words were mostly sonorant, and had penultimate or antepenultimate stress, facilitating the identification of pitch accent type and final boundary tone.

For each of these 78 interrogatives, we generated two short contexts, one that favored a rhetorical reading, one an information-seeking reading. The target utterance was placed within each context such that it was either in turn-final position, or it was followed by exactly one more sentence placing the target in "turn-medial" position. The context situations were as short and informative as necessary. In rhetorical contexts, the question was related to aforementioned information that helped to commit S to the proposition of the target utterance (e.g., "Everybody knows that [the game] *Domino* is totally out of date", see Appendix). In information-seeking contexts, the answer to the question was presented as unknown and the question was directed at A in order to get new information (e.g., "Since she does not know who would like to play which game, she asks the children", see Appendix). The 78 descriptions (consisting of context and target interrogative in turn-final position) were validated in a web-experiment with 132 participants (average age: 28.9 years, SD = 9.4 years, 93 female, 39 male) in six experimental lists containing 13 descriptions each (22 participants for each list). In the web-experiment, participants were given one context at a time (for each interrogative in either rhetorical or information-seeking condition) and had to judge whether the descriptions contained an RQ or not (RQs were defined as having an obvious answer and asserting a proposition). The 32 contexts that were recognized best in both the rhetorical and information-seeking version were then selected for the production experiment (16 polar questions, 16 wh-questions)

Furthermore, 16 filler descriptions (in which target interrogatives contained structural and lexical ambiguities) and 5 practice items (with unambiguous questions and statements) were constructed. They were otherwise similar to the experimental items.

2.1.2. Participants

We tested 48 monolingual native speakers of German (average age: 21.3 years; SD = 3.0; 36 female, 12 male). They took part in pairs and received 6€ each for their participation. None of them had taken part in the web-experiment mentioned above.

2.1.3. Procedure

Illocution type was manipulated within-subjects and position between-subjects. We constructed two experimental lists, one containing all trials in turn-medial, one in turn-final position. Each list contained all 32 experimental descriptions. The rhetorical and information-seeking version of a given target interrogative was separated by at least 4 other target or filler trials. The order of presentation of each interrogative (RQ or ISQ first) was counterbalanced across target descriptions. Five practice trials were added to each list.

The two members of a pair of participants were seated comfortably in a soundproof booth, facing each other in order to create a communicative situation. Both were wearing dynamic headset-microphones (Shure SM10A). They were recorded on two different channels of a Tascam HD-P2 portable stereo audio Recorder (44.1kHz, 16Bit). Context descriptions were presented one-by one on two MacBook Pro computer screens in black Calibri 32font on white background, using MSPowerPoint. Participants could not see each other's screens. In each trial, S and A read their context situations silently and S started the interaction by saying out loud the target utterance (in the turn-final condition) or the target utterance plus one additional sentence (in the turn-medial condition). A reacted as soon as possible by choosing and uttering the reply she considered most appropriate. Then they turned to the next trial by pressing a button on their laptops. Each session started with the five practice trials so that participants became familiar with the procedure. They were allowed to ask questions after the practice phase. This practice period was followed by the experimental session.

The participants were instructed to produce the sentences as naturally as possible. No feedback was provided. Participants were randomly assigned as speaker and addressee. We had 7 male and 17 female participants in the role of S, and 4 male and 20 female participants in the role as A.

2.1.4. Data treatment and analysis

Eighty of the 768 utterances were excluded because S did not produce them as given on screen, which made a phonetic comparison impossible. The remaining audio files were analyzed using Praat [16]. The target interrogatives were labeled at the level of the morphological word. Tones were labeled following the tonal guidelines of GToBI [17]. The first author labelled wh-questions, the second polar questions. Initial pitch (first pitch value of utterance) was automatically extracted using Praat. Pitch range in st of the first constituent (wh-word or verb) was calculated based on manual annotation of low and high targets within the constituents. The duration of the entire utterance was automatically extracted. Since longer utterance durations may be caused by duration adjustments in different constituents (e.g. [14]), we also extracted the duration of the first constituent and of the sentence final object (both normalized by the duration of the whole utterance to account for differences in overall speech rate). Voice quality was operationalized as spectral tilt. We measured the differences in amplitudes between H1 and A3 at the centre of the first stressed vowel (wh-word in wh-questions, verb in polar questions) ([18]). Larger values indicate softer voice quality.

2.2. Results

Polar questions were mainly rising (ISQ: 93%; RQ: 90%; see Figure 1; Figure 3 for an example). Moreover, RQs often did

not rise as strongly as ISQs (higher proportion of H-% compared to H-^H% boundary tones, see Figure 1). Wh-questions were mainly falling (ISQs: 92%; RQs: 99%; example in Figure 4). In addition, we see a difference in the distribution of nuclear accents (Figure 2): Rhetorical wh-questions show a higher proportion of L*+H nuclear accents (44%) than ISQs (34% difference).

For the statistical analysis, boundary tones were classified as high (H-^H%, H-%) vs. low (L-%); this factor was entered as dependent variable into a logistic linear mixed effects regression model with *position* (turn-medial vs. turn-final), *interrogative type* (polar vs. wh) and *illocution type* (RQ vs. ISQ) as fixed factors and *participants* and *items* as crossed random factors ([19]). Initially, we included all main effects and interactions and eliminated interactions and factors that were not significant (Benjamini-Hochberg corrected significance level [20]). We only report factors that were significant in the most parsimonious model. The choice of boundary tone was affected by *interrogative type* (more low boundary tones for wh-questions, $p < 0.0001$) and *illocution type* (more low boundary tones for RQs; $p = 0.005$) only.

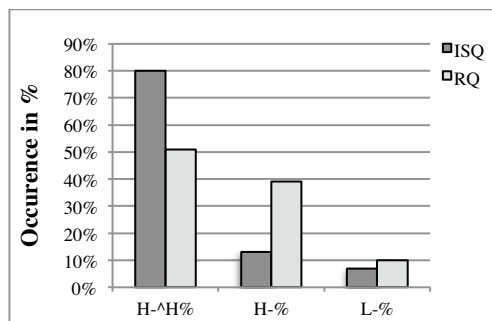


Figure 1. Boundary tones in polar questions.

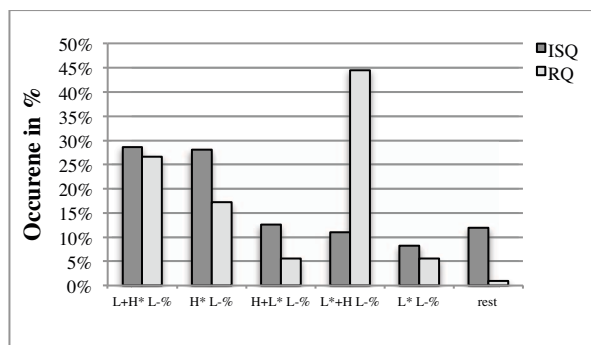


Figure 2. Nuclear tunes (nuclear pitch accent and boundary tone) in wh-questions.

To analyze the phonetic data, we ran a series of general linear mixed effects regression models with the same specification and procedure as above (*position*, *interrogative type*, and *illocution type* as fixed factors and *participants* and *items* as crossed random factors). For pitch analyses, we further included *gender* as a fixed factor.

Table 2 summarizes the variables for which *illocution type* had a significant effect (✓) or not (✗), split by polar and wh-questions. Naturally, we expected differences between wh- and polar questions in most of the measures, due to the differences in their syntactic structure. In the description below, we focus on effects of *illocution type*. Different

outcomes for polar and wh-questions are supported by a significant interaction between *illocution type* and *interrogative type* ($p < 0.0001$).

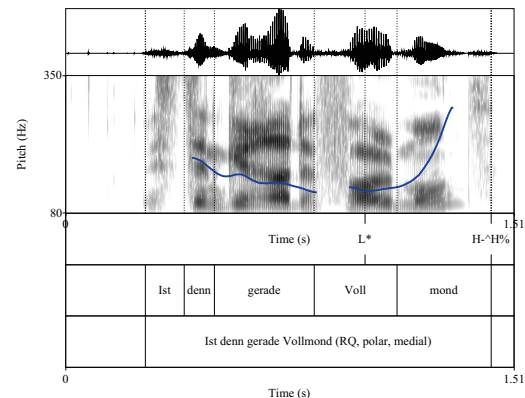


Figure 3: Example polar RQ in turn-medial position. F0 is smoothed and shown between 80 - 350Hz.

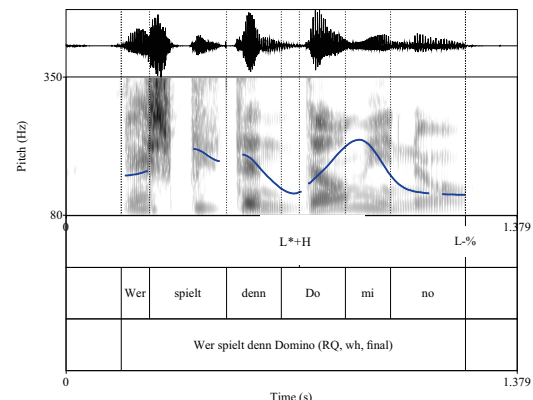


Figure 4: Example wh-RQ in turn-final position.

Measured variables	polar questions	wh-questions
Final boundary tone	✓	✓
Initial pitch	✓	✓
Overall utterance duration	✓	✓
Normalized duration of sentence final object	✗	✓
Normalized duration of first constituent (wh-word or verb)	✗	✓
Pitch range of first constituent	✗	✓
Spectral tilt (H1-A3) at the center of first vowel	✓	✓

Table 2. Significant differences (✓) between RQs and ISQs (Benjamini-Hochberg corrected significance level).

Initial pitch was significantly lower in RQs than in ISQs ($p = 0.003$) throughout. RQs started on average 9.2 Hz lower than ISQs (118.9 vs. 128.1 Hz).

Overall utterance duration was generally longer in RQs than in ISQs ($p = 0.002$). RQs were on average 39.9ms longer than ISQs (1455.4 vs. 1055.5ms).

The normalized duration of the sentence-final object was 3% longer for RQs than for ISQs, but only for wh-questions ($p < 0.00001$). The object takes up 41% of the utterance duration in RQs compared to 38% in ISQs.

The normalized duration of the first constituent was 2% shorter for RQs than for ISQs ($p < 0.02$), but only for wh-questions (14% vs. 16%).

The pitch range in the first constituent was 1.6st narrower for RQs than for ISQs, but only for wh-questions (3.8 vs. 5.4 st; $p < 0.0001$).

Spectral tilt was steeper for RQs than for ISQs ($p = 0.014$), indicating that RQs were generally produced with a breathier voice than ISQs.

3. Discussion

The results of the production study show that there are considerable differences in the prosodic realization of RQs and ISQs. What we also observe is that polar questions are marked less systematically for rhetoricity than wh-questions.

From a phonological perspective, RQs end with a falling contour 5% more often than ISQs, however, the effect is very small, in particular compared to the strong effect of *interrogative type* (polar questions are rising most of the time, wh-questions are falling). Therefore, the results do not confirm the categorical claim in [4] that RQs have falling intonation to signal their assertive function, but instead add to the variation reported in experimental findings of [6] and [7]. Interestingly, we see differences in the choice of the high boundary tone in polar questions; speakers produced more utterances ending in H-% in RQs than in ISQs and more H-^H% in ISQs. This can be seen as conventionalized pitch scaling to signal interrogativity [13], see also [21]. Interestingly, rhetorical wh-questions show a higher proportion of L*+H nuclear accents, an observation that has not been reported before. In descriptions on German intonation, this accent type has been argued to signal surprise ([22]) and sarcasm ([23]).

From an information-structural point of view, one would have expected more H+L* and H+!H* accents in RQs, accent types that signal that the respective referents are discourse-given ([24]; in our case mentioned in the context descriptions). This is not the case, however. Moreover, the sentence-final object was longer in RQs than in ISQs (in wh-questions), which is not expected if speakers simply use prosody to mark discourse-giveness ([25]). Taken together, these tonal and durational data suggest that speakers did more than signaling discourse-giveness. Instead we argue that these prosodic realizations are related to illocutionary force. Future research will have to minimize the influence of information structure when comparing the prosodic realization of RQs and ISQs, e.g., by making the respective referents discourse-given in both illocution types.

Phonetically, RQs started with a lower initial pitch, which led to a larger f_0 -excursion on the phrase-initial wh-word. On the other hand, no difference in f_0 -excursion was observed for the sentence-initial verb in polar questions. The same asymmetry is also found for the duration of the first constituent (only wh-words were shortened in RQs). This interaction between illocution type and interrogative type is most likely caused by the different intonational structures of polar and wh-questions: while the sentence-initial verb was hardly accented, the wh-word usually received a prenuclear accent. One strategy for RQs could hence be to lower the pitch value of the initial boundary tone (contradicting [10]) and to reduce the prominence of the first prenuclear accent. This generalization, however, is subject to future research.

Furthermore, RQs were longer than ISQs, similar to what [14] report for declarative questions in a number of languages. Hence, it seems that the temporal organization of utterances

may be used by speakers to signal questions. Future research will have to show whether this durational cue can be reliably used by listeners (which was not the case for the durational contrast between statements and intonation questions, [26]).

Generally speaking, rhetoricity was already marked early in the utterance (similar to declarative questions, [12]): the wh-phrase was made less prominent in RQs than in ISQs; the initial pitch was lower and the speakers' voice was softer. Softer voice may serve to attenuate the assertive force of RQs but more research is necessary in this respect. On the other hand, we see differences in phrase-final position, such as slight differences in the distribution of boundary tones and accent types. Future research will have to show whether listeners are sensitive to these phonological and phonetic differences.

4. Summary and Outlook

Our data show that, along with pragmatic and semantic factors, RQs and ISQs differ with respect to prosody (i.e. phonological and phonetic features). Even though RQs are syntactically identical to true interrogatives, prosodically they are different. Crucial properties of RQs were longer durations and a softer voice compared to ISQs, cues that already occur early-on in the interrogative utterance. In future studies, we will investigate the online interpretation of RQs and ISQs by listeners, testing the role context and prosody play in this respect. To this end, interrogatives produced by a trained speaker will be resynthesized in order to control for a number of parameters, and will be placed in information-seeking and rhetorical contexts. Participants will have to evaluate the manipulated stimuli as matching or mismatching. In addition we will compare our data with spoken corpus data.

5. Acknowledgments

This research was supported by a research grant from the University of Konstanz, awarded to the last two authors. We thank the audience of the workshop on questions at the DGfS conference 2015 in Leipzig for comments and discussion.

6. Appendix

Context situation ISQ (speaker):
Die Sprecherin, eine Mutter, teilt die Kinder am Geburtstag ihrer Tochter in verschiedene Gruppen ein. Da sie nicht weiß, wer gerne welches Spiel spielt, erkundigt sie sich bei den Kindern. 'On her daughters birthday, the speaker (the mother) separates the children into different groups. Since she does not know who would like to play which game, she asks the children.'
Target utterance: <i>Wer spielt denn Domino?</i> Der geht bitte an den Tisch dort hinten.
'Speaker (target utterance): Who plays Domino? Please go to the table over there.'
Context situation RQ (speaker):
Die Oma des Sprechers erkundigt sich, ob dieser Lust habe mit ihr und den anderen Enkelkindern eine Runde Domino zu spielen. Jeder weiß jedoch, dass Domino total veraltet ist. 'The speakers' grand mother is asking her grand child whether he wants to play Domino with her and the other grand children. However, everybody knows that Domino is totally out of date.'
Target utterance: <i>Wer spielt denn Domino?</i> Playstation ist doch viel cooler.
'Speaker (target utterance): Who plays Domino? Playstation is much cooler.'

Table 3: Context-example (RQ and ISQ in turn-medial position).

7. References

- [1] A. Grésillon, "Zum linguistischen Status rhetorischer Fragen," *Zeitschrift für Germanistische Linguistik*, vol. 8 no. 3: pp. 273-289, 1980.
- [2] J. M. Sadock, "Queclaratives," *Papers from the Seventh Regional Meeting Chicago Linguistic Society*. April 16-18 1971, vol. 7, pp. 223-231.
- [3] J. M. Sadock, *Toward a linguistic theory of speech acts*. New York: Academic Press, 1974.
- [4] C.-H. Han, *Interpreting interrogatives as rhetorical questions*. *Lingua*, vol. 112, no. 3, pp. 201-229, 2002.
- [5] H. Rohde, "Rhetorical Questions as Redundant Interrogatives," *Sand Diego Linguistic Papers*, vol. 2, pp. 134-168, 2006.
- [6] A. Banuazizi and C. Creswell, "Is that a real question? final rises, final falls, and discourse function in yes-no question intonation," *CLS*, vol. 35, pp. 1-14, 1999.
- [7] N. Hedberg, J. M. Sosa, E. Görgülü and M. Mameni, "Prosody and pragmatics of wh-interrogatives," in *2010 Annual Meeting of the Canadian Linguistics Society, Proceedings*, 2010.
- [8] E. Dombrowski and O. Niebuhr, "Shaping phrase-final rising intonation in German", *5th International Conference on Speech Prosody*. Chicago, Illinois, USA, 2010.
- [9] S. Kaiser and S. Baumann, "Satzmodus und die Diskurspartikel hm: Intonation und Interpretation," *Linguistische Berichte*, no. 236, pp. 473-496, 2013.
- [10] M.A. Sicoli, T. Stivers, N. J. Enfield and S. C. Levinson, "Marked Initial Pitch in Questions Signals Marked Communicative Function," *Language and Speech*, pp. 1-20, 2014.
- [11] W.F. Heeren, S. A. Bibyk, C. Gunlogson and M. K. Tanenhaus, "Asking or Telling - Real-time Processing of Prosodically Distinguished Questions and Statements," *Language and Speech*, pp. 1-28, 2015.
- [12] V.J. van Heuven and J. Haan, "Phonetic correlates of statement versus question intonation in Dutch," in *Intonation*. Springer: Netherlands. pp. 119-143, 2000.
- [13] J. Michalsky "Pitch scaling as a marker of interrogativity in canonical questions of German," presented at *The 37th annual meeting of the DGfS*. AG 6, Leipzig, Germany. 2015.
- [14] V.J. van Heuven and E. van Zanten, "Speech rate as a secondary prosodic characteristic of polarity questions in three languages," *Speech Communication*, vol. 47 No. 1-2, pp. 87-99, 2005.
- [15] M. Thurmair, "Zum Gebrauch der Modalpartikel 'denn' in Fragesätzen: Eine korpusbasierte Untersuchung," pp. 377-387, 1991.
- [16] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*. 2012.
- [17] M. Grice and S. Baumann, "Deutsche Intonation und GToBI," *Linguistische Berichte*, vol. 191, pp. 267-298, 2002.
- [18] P. A. Keating and C. Esposito, "Linguistic Voice Quality," *UCLA Working Papers in Phonetics U6*, vol. 105, pp. 85-91, 2007.
- [19] D. J. Barr, et al., "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language*, vol. 68, no. 3, pp. 255-278, 2013.
- [20] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, vol 57, no 1, Series B (Methodological), pp. 289-300, 1995.
- [21] C. Gussenhoven, "Intonation and interpretation: Phonetics and phonology," *1st International Conference on Speech Prosody*, Aix-en-Provence, 2002.
- [22] K. Kohler, "Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics," *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, vol. 25, pp. 115-185, 1991.
- [23] M. Grice, S. Baumann, and R. Benz Müller, *German Intonation in Autosegmental-Metrical Phonology*, in *Prosodic Typology*, OUP. 2005.
- [24] S. Baumann and M. Grice, "The Intonation of Accessibility," *Journal of Pragmatics*, vol. 38, pp. 1636-1657, 2006.
- [25] C. Fowler, "Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction," *Journal of Memory and Language*, vol. 26 no. 5, pp. 489-504, 1987.
- [26] F. Cangemi, and M. D'Imperio, "Tempo and the perception of sentence modality," *Laboratory Phonology*, vol. 4, pp. 191-219, 2013.