

The prosody of rhetorical and information-seeking questions in German

Bettina Braun, Nicole Dehé, Jana Neitsch, Daniela Wochner, Katharina Zahner

University of Konstanz, Department of Linguistics, Germany

Running head: Prosody of rhetorical questions in German

Corresponding author:

Bettina Braun

University of Konstanz

Postfach 186

78457 Konstanz

Germany

Bettina.Braun@uni-konstanz.de

Abstract

This paper reports on the prosody of rhetorical questions (RQs) and information-seeking questions (ISQs) in German for two question types, polar questions and constituent questions (henceforth *wh*-questions). The results are as follows: Phonologically, polar RQs were mainly realized with H-% (high plateau), while polar ISQs mostly ended in H-^H% (high-rise). *Wh*-RQs almost exclusively terminated in a low edge tone, whereas *wh*-ISQs allowed for more tonal variation (L-%, L-H%, H-^H%). Irrespective of question type, RQs were mainly produced with L*+H accents. Phonetically, RQs were more often realized with breathy voiced quality than ISQs, in particular in the beginning of the interrogative. Furthermore, they were produced with longer constituent durations than ISQs, in particular at the end of the interrogative. While the difference between RQs and ISQs is reflected in the intonational terminus of the utterance, this does not happen in the way suggested in the semantic literature, and in addition, accent type and phonetic parameters also play a role. Crucially, a simple distinction between rising and falling intonation is insufficient to capture the realization of the different illocution types (RQ, ISQ), against frequent claims in the semantic and pragmatic literature. We suggest alternative ways to interpret the findings.

Keywords: rhetorical question, illocution, prosody, question types, German

Introduction

This paper deals with the prosody of two different illocution types in German, rhetorical questions (RQs) and information-seeking questions (ISQs). By prosody we mean both intonational events and phonetic realization. To this end we investigate two question types: polar questions and constituent questions (henceforth *wh*-questions), see (1) and (2), respectively. Syntactically, German polar questions start with a finite verb (1), *wh*-questions with a *wh*-word, followed by the finite verb in second position (2).

(1) Mag denn jemand Limonen?
Likes PRT anyone lemons?
'Does anyone like lemons?'

(2) Wer mag denn Limonen?
Who likes PRT lemons?
'Who likes lemons?'

To date, the literature discusses RQs mostly in terms of their semantic and pragmatic properties. Canonical ISQs are used to seek information from the addressee. The answer to an ISQ can only be given by the addressee and not by the speaker (Caponigro & Sprouse, 2007). In contrast, RQs do not require or expect answers from the addressee as stated by several authors (Banuazizi & Cresswell, 1999; Biezma & Rawlins, 2017; Han, 2002; Hudson, 1975; Ilie, 1995; Quirk, et al., 1985; Wilson & Sperber, 1988). Instead, they are used when the answer is already known or at least inferable to all interlocutors (Caponigro & Sprouse, 2007; Ilie, 2010; Sadock, 1971, 1974). Moreover, the purpose of RQs is to seek the addressee's commitment to the proposition that is presupposed by the question (Biezma & Rawlins, 2017). This latter definition excludes RQs that are uttered in monologues to change a topic or to engage the audience, as they occur, for example, in parliamentary speeches or in court (Ilie, 1995). It also excludes RQs that are used as retorts (e.g., *Is the Pope Catholic?*, cf. Sadock (1974), Schaffer (2005)). None of these types will be investigated in this paper.

Regarding syntactic form and lexical cues, a rhetorical illocution may be signaled by strong negative polarity items (e.g., *Who on earth needs holidays?*, cf. Gutiérrez Rexach (1998) and

Biezma and Rawlins (2017)), and, in German, by discourse particles: *schon* and *auch*,¹ cf. Meibauer (1986, p. 112ff), Biezma and Rawlins (2017). These markers are sufficient to trigger an RQ interpretation, but they are not necessary. Instead, RQs and ISQs can be string-identical on the surface, in which case they can be disambiguated by the context (3), as well as by their prosodic realization, which is the focus of this paper.

(3) **ISQ context:**

At a party, you offer cake made with limes. You would like to know which of the guests like this fruit and would like some of it. You say to your guests:

Q: *Does anyone like limes?*

RQ context:

Your aunt offers limes to her guests. However, it is known that this fruit is too sour to be eaten on its own. You say to your cousin:

Q: *Does anyone like limes?*

Regarding the prosodic realization of RQs and ISQs, there are generalizations based on introspection (Bartels, 1999; Han, 2002) as well as two corpus studies (Banuazizi & Cresswell, 1999; Hedberg, et al., 2010), all on English. Furthermore, there is one experimental pilot study on German (Wochner, et al., 2015). These studies will briefly be presented here to motivate the hypotheses for our study. Bartels (1999, p. 255) suggests that polar questions ending in H-H% (high-rising boundary tone) may be interpreted as ISQs or RQs, but polar questions ending in L-L% (low boundary tone) are very likely intended as RQs. Since *wh*-questions are typically falling in English (e.g., Quirk, et al., 1985), L-L% is argued not to be a signal for RQs in *wh*-questions. Instead, Bartels (1999, p. 257) claims that "non-tonal cues are necessary for their categorization as statements or questions", in particular contextual information. This leads to the asymmetric pattern summarized in Table 1.

¹ The lexical meaning of *schon* is 'already', that of *auch* is 'also, too'.

Edge tone	polar question	<i>wh</i> -question
H-H%	ISQ, RQ	ISQ ("addressee is seen as already committed to the presupposition"), cf. Bartels (1999, p. 256)
L-L%	RQ	ISQ, RQ

Table 1. Interpretation of edge tones as ISQ or RQ according to Bartels (1999).

In her syntactic and pragmatic analysis of RQs, Han (2002, p. 215) argues that polar rhetorical questions have falling intonation, while polar information-seeking questions end in a rise. This is partly in line with what Bartels (1999) claims for polar questions. Han (2002) relates the falling contour to assertive force, which is also often found in declaratives. Note, however, that in more recent works, authors disagree about the assumption that RQs necessarily have assertive force (cf. Biezma & Rawlins, 2017; Caponigro & Sprouse, 2007). For *wh*-questions, Han (2002, p. 217) remains very vague about prosodic realization, suggesting only "that the intonational contour serves as a cue that a *wh*-question is a rhetorical question that expresses an assertion", but without specifying the exact intonational means.

Corpus studies do not fully confirm the intonational patterns of RQs argued for in the semantic literature. Banuazizi and Cresswell (1999) investigated the intonation of polar questions in telephone conversations in English in the SWITCHBOARD corpus. They analyzed 102 polar RQs and 2106 polar ISQs. Of the 102 polar RQs, only 45 (44.1%) ended in a final fall (L-L%), and 57 (55.9%) ended in a final rise (H-H%). In comparison, in polar ISQs, 89.7% ended in a rise. These data speak against the analysis of Han (2002) and largely confirm Bartels' (1999) assumptions (middle column of Table 1).

Hedberg et al. (2010) studied *wh*-questions, both in terms of their function and in terms of their prosody. They distinguished between information-seeking and non-information-seeking questions (the latter comprised RQs, back-channel questions, and questions to self) as well as other pragmatic classifications such as floor-management, topic maintenance, interruption of the discourse, and givenness of the propositional content of the question. There were only 26 rhetorical *wh*-questions in their corpus. Of these, 21 ended in falling contours (ten high-falls, eight rise-falls, three low-falls) and five were rising. Note that rising *wh*-RQs are not present in Bartels' (1999) analysis. It is conceivable that the rising non-information-seeking questions in Hedberg et al. (2010) did not come from genuine RQs but from some other kind of non-information-seeking questions. Moreover, given the nature of the data, other linguistic properties of these *wh*-questions may have played a role, too. In any case, a controlled

experimental investigation of the intonational realization of string-identical RQs and ISQs is missing in the literature.

A pilot production study on German was a first attempt to fill this gap (Wochner, et al., 2015). In that study, participants were tested in pairs, each with his/her own screen, facing each other. One of the members was assigned the role of the speaker, the other the role of the addressee. The speaker read a context on screen (either triggering an information-seeking or a rhetorical interpretation) and then produced an interrogative. Simultaneously, the addressee saw a shortened version of that context on his/her screen and replied to the speaker's interrogative, using one of two suggested answer sentences presented on screen. One of these two possible answers was compatible with an RQ interpretation (thus the addressee voiced agreement with the speaker), the other one with an ISQ interpretation (thus the addressee provided information). The phonological analysis of pitch accent types and edge tones showed the following differences between German ISQs and RQs: Polar RQs showed a higher proportion of high plateaus (H-%) than polar ISQs (which mostly ended in H-^H%). *Wh*-questions of both illocution types ended mostly in a fall (L-%), but there was a higher proportion of L*+H nuclear accents in *wh*-RQs than in *wh*-ISQs. Given the explorative nature of the pilot study, the contexts and interrogatives were very diverse. In particular, the target interrogatives were not controlled for syntactic structure, and the contexts were more varied and longer than in the present experiment and were not controlled for information structure. Based on the English corpus data and the pilot study on German, we formulate the following two phonological hypotheses for the current, more controlled, study:

- H1: In polar questions, the edge tone distinguishes between RQs and ISQs: there are more high rises (H-^H%) in polar ISQs than in polar RQs, and more high plateaus (H-%) in polar RQs than in polar ISQs.²
- H2: *Wh*-questions end in a low edge tone (L-%) across illocution types, but RQs and ISQs differ in pitch accent types: We expect more late rises (L*+H accents) in RQs than in ISQs.

Questions have been argued to differ from 'non-questions' in terms of phonetic properties, too. Sicoli, et al. (2015), for instance, analyzed a large number of questions uttered in natural conversations in typologically different languages (ǀǁKha, Haillom, Danish, Dutch, English,

² We analysed the data in the GToBI framework (Grice, et al., 2005). Descriptive terminology such as high-rising, low-rising, etc. is used for ease of reading.

Italian, Japanese, Korean, Lao, Tzeltal, Yélf-Dnye). The questions were coded as information-seeking, confirmation-seeking or agreement-seeking. Confirmation-seeking and agreement-seeking questions (e.g., *Isn't that a horrible color?* or *The weather's just gorgeous isn't it?*) were grouped under the term evaluative questions. The analysis of the height of initial pitch in the first accented syllable – a measure they argue to be suitable for typologically different languages – showed that the evaluative questions started with an initial pitch in the top 10% of a speaker's range more often than information-seeking questions. In an experimental study, Heuven and Haan (2002) compared Dutch declarative questions to string-identical declarative statements and found higher prenuclear pitch peaks in declarative questions than in statements (cf. Petrone & Niebuhr, 2014 for the relevance of height, shape and alignment differences in prenuclear accents in German). Other researchers have argued that declarative questions differ from declarative statements in terms of duration: declarative questions are produced with a faster speaking rate than statements in German, Manado Malay, Orkney English and Dutch (e.g., Heuven & Zanten, 2005; Niebuhr, 2013a; Niebuhr, et al., 2010). Perception studies on Neapolitan Italian, on the other hand, showed no effect of duration (Cangemi & D'Imperio, 2013). It is therefore unclear how robust this cue is. If faster speaking rate is a cue to genuine questionhood, we expect ISQs to be produced with shorter durations than RQs. Another phonetic difference between ISQs and RQs is voice quality. As can be seen in (3) above, the contexts triggering RQ readings state that a certain fact is self-evident (e.g., "However, you know that this fruit is too sour to be eaten on its own."). These contexts may hence trigger an exasperated attitude, which in turn may result in a breathy voice quality (Schourup, 1985 for English). Finally, in Wochner et al.'s (2015) pilot experiment summarized above, RQs were produced with longer durations than ISQs and had a breathier voice quality. Given the lack of control of information structure in that study, and given the fact that information structure affects duration (Baumann, et al., 2006; Cooper, et al., 1985; De Jong, 2004; Féry & Kügler, 2008), the present study follows up on Wochner et al. (2015) and tests the effect of illocution type on phonetic parameters more directly. Given the above-mentioned findings, hypothesis H3 is stated as follows.

H3: Phonetically, RQs start with higher initial pitch, and have longer constituent durations and breathier voice quality than ISQs.

Production Experiment

In order to test hypotheses H1 through H3 given above, a production experiment was carried out. Participants read context descriptions presented on screen and produced the subsequent target interrogatives.

Methods

Materials

We constructed 22 *wh*-interrogatives that fitted both a rhetorical and an information-seeking reading (e.g., *Who likes celery?*). To this end, we used predications that – out of context – may be true for some people and false for others (e.g., 'liking celery'). From these *wh*-interrogatives, we derived polar questions by replacing the *wh*-word by the indefinite subject *anyone* and adapted the syntactic structure to verb-first (V1) (e.g., *Mag den jemand Sellerie?* 'Does anyone like celery?'). The polar questions thus contained an open element, similar to the *wh*-pronouns in *wh*-questions. In sum, we had 22 pairs of matched *wh*- and polar questions, henceforth referred to as interrogative pairs. Within the pairs, only the syntactic structure (*wh*-pronoun + verb vs. verb + subject) varied between question types, but the proposition expressed by the sentence radical was the same.

For each interrogative pair, we constructed two contexts, one context triggering an information-seeking and one context triggering a rhetorical interpretation. An example of the resulting quadruple is given in Table 2. The complete list of questions is provided in Table A1 in the Appendix. To control for information structure and specifically to avoid effects of information structure on nuclear accent position and type as well as on duration, both the verb and the nominal object of the target interrogative were always lexically given in the context the speaker was presented with (e.g., *liking celery* in Table 2).

ISQ	RQ
wh-question	
You cooked a dish with celery. You would like to know which of your guests likes this vegetable and would like some of it. You say to your guests:	In the canteen they have casserole with celery on the menu. However, you know that nobody likes this disgusting vegetable. You say to your friends:
<i>Wer mag denn Sellerie?</i> 'Who likes celery?'	
polar question	
You cooked a dish with celery. You would like to know whether your guests like this vegetable and will eat it or not. You say to your guests:	In the canteen they have casserole with celery on the menu. However, you know that nobody likes this disgusting vegetable. You say to your friends:
<i>Mag denn jemand Sellerie?</i> 'Does anyone like celery?'	

Table 2. Contextual settings with both polar and *wh*-questions in both illocution types evoking a rhetorical or an information-seeking question interpretation; contexts and target interrogatives translated from German.

The rhetorical contexts for a given interrogative pair (polar, *wh*) were identical. They all contained a sentence stating that it is generally known (or that the speaker knows) that nobody agrees with the proposition stated in the sentence radical. In some of the contexts, an explicit reason was stated for the speaker's knowledge that nobody agrees with the proposition (e.g., disgusting vegetable, banal roses, old-fashioned dance), which may have induced a certain emotional stance. The information-seeking contexts differed from the rhetorical contexts in that they stated that the speaker was looking for some piece of information. The information-seeking contexts were largely identical for the two question types and differed only in whether uncertainty was expressed about the polarity (in polar questions; e.g., *whether or not your guests like it*) or about the subject (in *wh*-questions; e.g., *who likes it*). Each target interrogative ended in a mostly sonorous sentence-final object noun, consisting of two to four syllables with primary lexical stress on the penultimate (six items) or antepenultimate syllable (five items). All target interrogatives contained the modal particle *denn*, which frequently occurs in both question types in German (Thurmair, 1991). The use of *denn* facilitated the creation of natural target sentences in both conditions without biasing the interpretation of the utterance towards a rhetorical or information-seeking reading (Thurmair, 1989). Given this make-up of contexts and target interrogatives, the actual illocution of the target was determined only by the information given in the contexts.

We used 28 filler sentences to distract the participants from the task. They were six questions with structural (PP-attachment) ambiguities, each of which occurred in two contexts, and 22 exclamatives with V1 word order, i.e., the same word order as in polar questions.

Procedure

Two basic experimental lists were constructed. Each list contained the polar question for half of the question-pairs and the *wh*-question for the other half. Illocution type was manipulated within-subjects. That is, each participant produced both the rhetorical and the information-seeking version of each target interrogative, but only one question type of each pair. The 28 filler items were added to each list. The experimental lists were randomized anew for each participant with the constraint that the two illocution types of one target interrogative (ISQ, RQ) were separated by at least four other trials. Each experiment started with four familiarization trials, followed by a short break, which participants were allowed to use for questions if anything was unclear. The experiment was controlled using the experimental software *Presentation* (Neurobehavioral-Systems, 2000). Each trial started with the visual display of the context, which the participant had to read silently, followed – upon button press – by the target interrogative on the next screen. The target sentence had to be produced aloud. Participants were asked to produce the questions in such a way that they were suitable in the given context. The experiment was self-paced. The recording started simultaneously with the appearance of the interrogative on screen. After the production of the target, participants pressed a button to proceed to the next trial. The recording of the previous target was stopped at this point. Participants were allowed to repeat the question in case of mispronunciation or other mistakes (participants only rarely used this option, < 0.5% of the cases). No feedback was provided during the actual experiment. The experiment lasted about 25 to 30 minutes. Productions were recorded using a headset-microphone (Shure SM10A) and digitized directly onto a PC (44.1 kHz, 16Bit, stereo).

Participants

Twelve monolingual native speakers of German (average age = 21.7, SD = 2.3; 10 female, 2 male) participated for a small payment. They were students at the University of Konstanz and were unaware of the purpose of the study. The participants were randomly assigned to one of the two experimental lists. None of them reported any speaking or hearing disorders.

Data treatment and analysis

In total we collected 528 target interrogatives (44 contexts x 12 participants), of which 26 realizations (4.9%) had to be excluded due to mispronunciation ($N = 14$), laughter ($N = 2$), technical errors ($N = 2$) or audible pauses between the syntactic constituents ($N = 8$). In case of

multiple recordings, the second recording was analyzed. The final data set consisted of 249 polar questions (125 ISQs and 124 RQs) and 253 *wh*-questions (126 ISQs and 127 RQs).

To be able to analyze duration and to provide acoustic validation for perceptual voice quality coding (see below), the target files were automatically segmented using MAUS (Kisler, et al., 2012; Schiel, et al., 2011, cf. <ftp://ftp.bas.uni-muenchen.de/pub/BAS/SOFTW/MAUS>), based on an orthographic representation of each sentence (Kipp, et al., 1996). Word boundaries were manually corrected using standard segmentation criteria (Turk, et al., 2006) in the software package Praat (Boersma & Weenink, 2018). The boundaries of stressed syllables were determined using the maximum onset principle. In case of intervocalic ambisyllabic consonants, the acoustic midpoint served as syllable boundary. For the extraction of acoustic parameters for the analysis of voice quality, we manually annotated the middle of the vowels in the stressed syllables of the first constituent (verb in polar questions, *wh*-word in *wh*-questions) and the sentence-final object noun in both question types. At these points, HNR (harmonics-to-noise ratio, expressed in dB) was extracted in Praat via Voice Report on the basis of a forward cross-correlation analysis with a sensitivity of 60dB and a time resolution of 12ms in the frequency range between 100 and 500 Hz. HNR is the ratio between the energy in the periodic part and the non-periodic part of the signal ($10 \times \log_{10}(\text{harmonic part}/\text{non-harmonic part})$, cf. Boersma, 1993). In modally voiced vowels, HNR values are greater, whereas more breathy voiced vowels (associated with asthenic voice and dysphonia) are characterized by lower HNR values (e.g., de Krom, 1993; Teixeira, et al., 2013). HNR was previously found to be the best single predictor of breathiness among a variety of frequency-domain parameters (de Krom, 1995)³.

For the analysis of intonation and voice quality, which was performed on a perceptual basis, the last three authors divided the items among them such that each of them annotated one third of the items. Voice quality was classified as modal, breathy or glottalized in the stressed syllables of the initial word (verb in polar questions, *wh*-word in *wh*-questions), of the second constituent (subject in polar questions, verb in *wh*-questions), and of the final object noun. A perceptual classification is more robust than acoustic measures, given variability in the

³ Other measures used in subsequent studies are incompleteness of closure (IC: the first formant's bandwidth (B1) normalized by the frequency of the first formant (F1), see Pützer and Wokurek (2015)) and H1-H2 (Niebuhr, et al., 2010). IC relies on a correct estimation of f0 and F1, which was often error prone with our data, resulting in a large number of missing data (approximately 35% of the data points). Regarding H1-H2, Simpson (2009, 2012) doubts its use as correlate of breathiness, because of inherent nasality in open vowels and sex-specific differences in harmonic spacing.

experimental materials regarding the quality of the stressed vowel and the word-prosodic structure of the words. For intonational analysis, pitch accents and edge tones were annotated according to the GToBI guidelines (Grice & Baumann, 2002; Grice, et al., 2005). The two rising bitonal accents $L+H^*$ and L^*+H , which have been shown to cause disagreement between transcribers (Braun, 2005; Grice, et al., 1996) were mainly differentiated based on the alignment of the L-tone. When the L-tone was aligned with the onset of the stressed syllable (acoustically before or at the start of the onset consonant) and the perceptual impression of the stressed syllable was high-pitched, the accent was labeled $L+H^*$. When the L-tone was aligned with the rhyme of the stressed syllable (leading to the impression of a low-rising accent), the accent was labeled L^*+H .

To check the reliability of perceptual annotations, approximately 20% of each person's annotations were annotated by one of the other annotators. This was done in a rotation system so that inter-annotator agreement and reliability was calculated between three different pairs of annotators. The data set for interrater checks included $N = 100$ questions, 49 RQs (22 polar, 27 *wh*) and 51 ISQs (26 polar, 25 *wh*). Interrater reliability was assessed by calculating Cohen's kappa (Cohen, 1960) with the *irr* package in R (Gamer, et al., 2012). For voice quality, annotators agreed in 89.7% of the cases (269 out of 300 labels). Interrater agreement was "substantial" (Landis & Koch, 1977, p. 165) ($\text{kappa} = 0.71$).⁴ For intonational labels, annotators agreed in 87.1% of the cases (526 out of 604 labels), $\text{kappa} = 0.84$ which is interpreted as "almost perfect" agreement (Landis & Koch, 1977, p. 165).⁵ Since agreement between annotators was very high, suggesting reliable annotations, the initial annotation of the first annotator was used for further analysis.

The measure "initial pitch" was operationalized as in Sicoli, et al. (2015). First, the f_0 -value at the point of maximum intensity in the first accented syllable was extracted (as will be shown in more detail below, this was typically the *wh*-word in *wh*-questions, and the subject *anyone* in polar questions). Next, we extracted the speaker's median pitch and the 90% quantile and

⁴ The cases of disagreement concerned the confusion of the label glottal vs. modal ($N = 17$), breathy vs. modal ($N = 10$) and breathy vs. glottal ($N = 4$). Most of the disagreements occurred on the initial word, i.e., the verb in polar questions or the *wh*-word in *wh*-questions ($N = 17$). The disagreements were distributed across question type and illocution type as follows: polar ISQ ($N = 5$), polar RQ ($N = 4$), *wh*-ISQ ($N = 9$), *wh*-RQ ($N = 13$).

⁵ There were 12.9% ($N = 78$) disagreements (most frequent between H^* / no accent ($N = 14$) and L^* / no accent ($N = 6$) in the prenuclear field, and between $L+H^*$ / L^*+H ($N = 8$) and L^* / L^*+H ($N = 6$) on the sentence-final object noun.

classified the data as high initial pitch if the extracted initial f0-value lay in the top 10% of the speaker's pitch range of this utterance.

For the statistical analyses of the categorical dependent variables accent type, edge tone (combination of phrase accent and boundary tone), initial pitch and voice quality, we ran logistic mixed effects regression models with *illocution type* (RQ vs. ISQ) and *question type* (polar vs. *wh*) as fixed factors, and *participants* and *items* as crossed random factors (adjustments of intercepts) (cf. Baayen, 2008; Baayen, et al., 2008).⁶ Then, we added random slopes for the fixed factors to the random-effects-structure and kept them if this improved the fit of the model (Bates, et al., 2015; Matuschek, et al., 2017). For model comparison, we used a comparison of the model's LogLikelihood, which is implemented in the R-function `anova()`. In case the dependent variable had more than two levels, one of the levels was coded as 1 and all other levels were coded as 0, and the effects of the fixed factors were calculated for this modified dependent variable (Agresti, 2002). For the statistical analyses of the phonetic variables, the dependent variables were analyzed using linear mixed effects regression models with the same specification and model fitting as described for the categorical data. To estimate the degrees-of-freedom (and arrive at *p*-values) we used the Satterthwaite approximation that is implemented in the R-library `lmerTest` (Kuznetsova, et al., 2016). To account for the fact that multiple variables were tested, we adjusted the *p*-values using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). For the sake of completeness, both the raw and the adjusted values are reported in this paper.

Results

In this section, we first analyze the phonological events final edge tones (combination of phrase accent and boundary tone), nuclear pitch accents, frequent nuclear tunes, and prenuclear accents, and then turn to the phonetic variables initial pitch, constituent durations and voice quality. In autosegmental-metrical (AM) theory, nuclear pitch accents, phrase accents and boundary tones are freely combinable (Pierrehumbert, 1980), but certain combinations of pitch accents, phrase accents and boundary tones occur more frequently than others (Dainora, 2006). In the German GToBI annotation system, which is couched in the AM framework, phrase accents and boundary tones are largely seen as one unit, which is why we analyse them together as final edge tones. Since work on the prosody-semantics interface often refers to pitch accents

⁶ Note that gender was not included as a fixed factor because of the imbalanced gender distribution and the need for much larger sample sizes to investigate gender effects (cf. Niebuhr, 2015).

and boundary tones separately (Hedberg & Sosa, 2007; Kohler, 2004; Pierrehumbert & Hirschberg, 1990; Truckenbrodt, 2012), we analyse pitch accents and edge tones separately.

Final edge tones (combination of phrase accent and boundary tone)

The distribution of final edge tones is plotted in Figure 1. Note that for this bar chart, four items with rare edge tones were removed (3x !H-%, 1x H-L%); the percentages are calculated with respect to the complete data set. Likewise, the statistical analysis is done for the complete data set. For polar questions, ISQs were more frequently produced with a high-rising H-^H% (88%) than RQs (23%, $\beta = 3.3$, $SE = 0.4$, $z = 7.7$, $p = p_{\text{adjusted}} = 0.0001$). Conversely, the mid-level plateau H-% was more frequent in RQs (67%) than in ISQs (< 1%, $\beta = 7.4$, $SE = 1.3$, $z = 5.5$, $p = p_{\text{adjusted}} = 0.0001$). Figure 2 shows representative realizations of the respective edge tones of polar ISQs (upper panel) and polar RQs (lower panel). There was no difference in the occurrence of the low edge tone L-% ($p = p_{\text{adjusted}} > 0.8$). The effect of illocution type on the low-rising L-H% could not be calculated because of zero occurrences in polar RQs.

Wh-questions typically ended in a final fall, L-%, in both illocution types (RQs: 94%, ISQs: 44%). The L-% occurred more than twice as often in RQs than in ISQs, a difference that was statistically significant ($\beta = 3.7$, $SE = 1.1$, $z = 3.5$, $p = 0.0005$, $p_{\text{adjusted}} = 0.001$). Additionally, in ISQs we found significantly more final rises (H-^H% and L-H%) than in RQs ($\beta = 4.6$, $SE = 0.7$, $z = 6.5$, $p = p_{\text{adjusted}} < 0.0001$, due to the small number of data points, these two rising edge tones were collapsed into one category 'rising'). There are interactions between illocution type and question type for H-% ($\beta = 6.8$, $SE = 1.6$, $z = 4.3$, $p = p_{\text{adjusted}} < 0.0001$) and L-% ($\beta = 4.0$, $SE = 0.9$, $z = 4.5$, $p = p_{\text{adjusted}} < 0.0001$), but not for H-^H% ($p = p_{\text{adjusted}} > 0.5$) or the combined category 'rising edge tones' ($p = p_{\text{adjusted}} > 0.9$). That is, question types differ in the use of edge tones for signalling illocution type for H-% and L-%, but not for H-^H%.⁷

⁷ One reviewer argues that the choice of nuclear accent and boundary tone may also depend on the number of syllables between nuclear accent and the utterance end (Grabe, 1998; Hanssen, 2017; Rathcke, 2006; Yu & Zahner, 2018). In our materials, all object nouns had a final unstressed syllable and stress was either penultimate or antepenultimate. Descriptive analyses showed that stress position did not affect the choice of accent types and boundary tones.

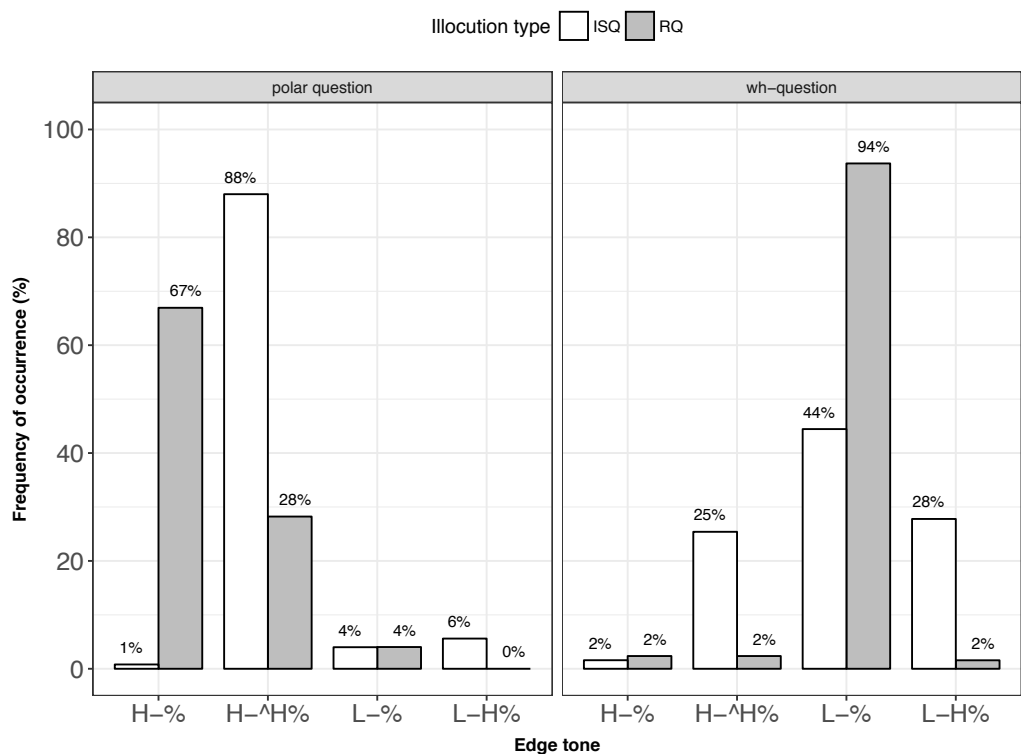
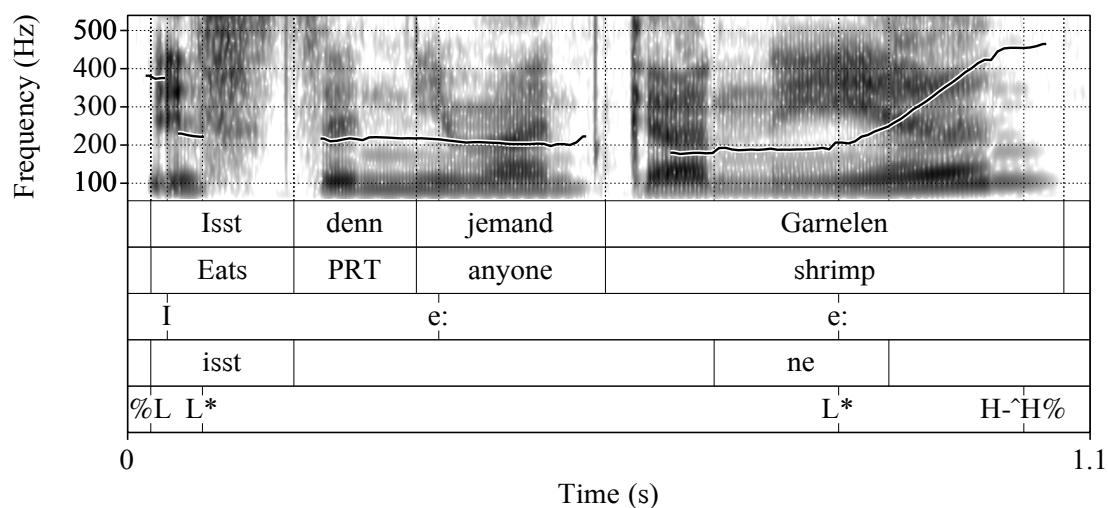


Figure 1. Distribution of frequent edge tones across illocution types. Only edge tones occurring more than 10 times in one of the two illocution types (i.e., more than 5%) are included.

Please insert Figure 1 about here



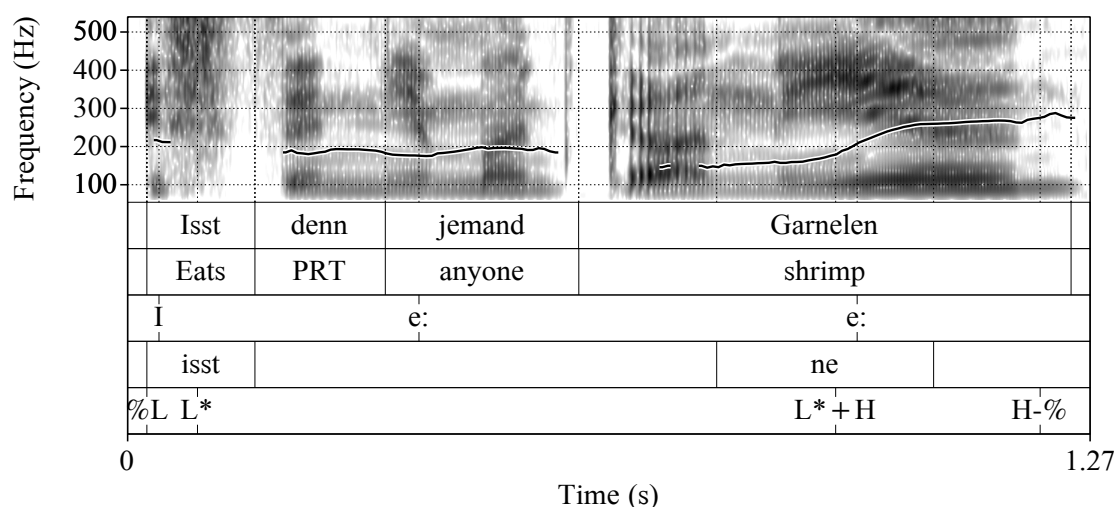


Figure 2. Example realization of an ISQ high-rising contour (upper panel) and an RQ plateau contour (lower panel). The third tier indicates the middle of the stressed vowel and its quality, the fourth layer the boundaries of the accented syllables and the fifth layer the GToBI annotation.

Please insert Figure 2 about here

Nuclear accents

The nuclear accent was generally associated with the object noun. We start with the most frequent accent types: In polar questions, the low-accent L^* was the most frequent nuclear accent in ISQs; it occurred significantly more often in ISQs than in RQs ($\beta = 2.8$, $SE = 0.4$, $z = 7.9$, $p = p_{\text{adjusted}} < 0.0001$, see Figure 3). The late-rise L^*+H was the most frequent accent type in RQs; it was significantly more frequent in RQs than in ISQs ($\beta = 5.1$, $SE = 0.8$, $z = 6.1$, $p = p_{\text{adjusted}} < 0.0001$). The early rise $L+H^*$, as well as deaccentuation of the object noun, were rare. $L+H^*$ did not differ significantly between illocution types ($p = 0.1$, $p_{\text{adjusted}} = 0.14$), but there were more unaccented object nouns in polar RQs than in polar ISQs ($\beta = 1.3$, $SE = 0.6$, $z = 2.2$, $p = 0.03$, $p_{\text{adjusted}} = 0.05$). When the object noun was deaccented, the nuclear accent was typically realized on the verb.

In *wh*-questions, the late-rise L^*+H was more frequent in RQs than in ISQs ($\beta = 3.7$, $SE = 0.5$, $z = 6.9$, $p = p_{\text{adjusted}} < 0.0001$), similar to what was found for polar questions. For ISQs, the most frequent accent type was the early rise $L+H^*$, which occurred more frequently in *wh*-ISQs than in *wh*-RQs ($\beta = 1.2$, $SE = 0.5$, $z = 2.6$, $p = p_{\text{adjusted}} < 0.0001$). The low accent L^* was more frequent in *wh*-ISQs than in *wh*-RQs ($\beta = 1.7$, $SE = 0.4$, $z = 4.0$, $p < 0.0001$, $p_{\text{adjusted}} = 0.0001$), H^* did not differ across illocution types ($p = p_{\text{adjusted}} > 0.2$) and $H+!H^*$ was significantly more

frequent in ISQs than in RQs ($\beta = 1.7$, $SE = 0.4$, $z = 4.0$, $p = 0.0007$, $p_{\text{adjusted}} = 0.002$). There was hardly any deaccentuation of the object noun in *wh*-questions (1% of the *wh*-questions), so this accent type is not statistically analysed.

The comparison of the distribution across illocution and question types shows no interaction between illocution type and question type for either L^*+H ($p = p_{\text{adjusted}} > 0.3$), or $L+H^*$ ($p = p_{\text{adjusted}} > 0.5$). The interaction approached significance for L^* ($p = 0.07$, $p_{\text{adjusted}} > 0.1$). This suggests similar differences in accent type distribution between ISQs and RQs across the two question types. Interactions for the other accent types could not be calculated due to zero occurrences in some of the conditions.

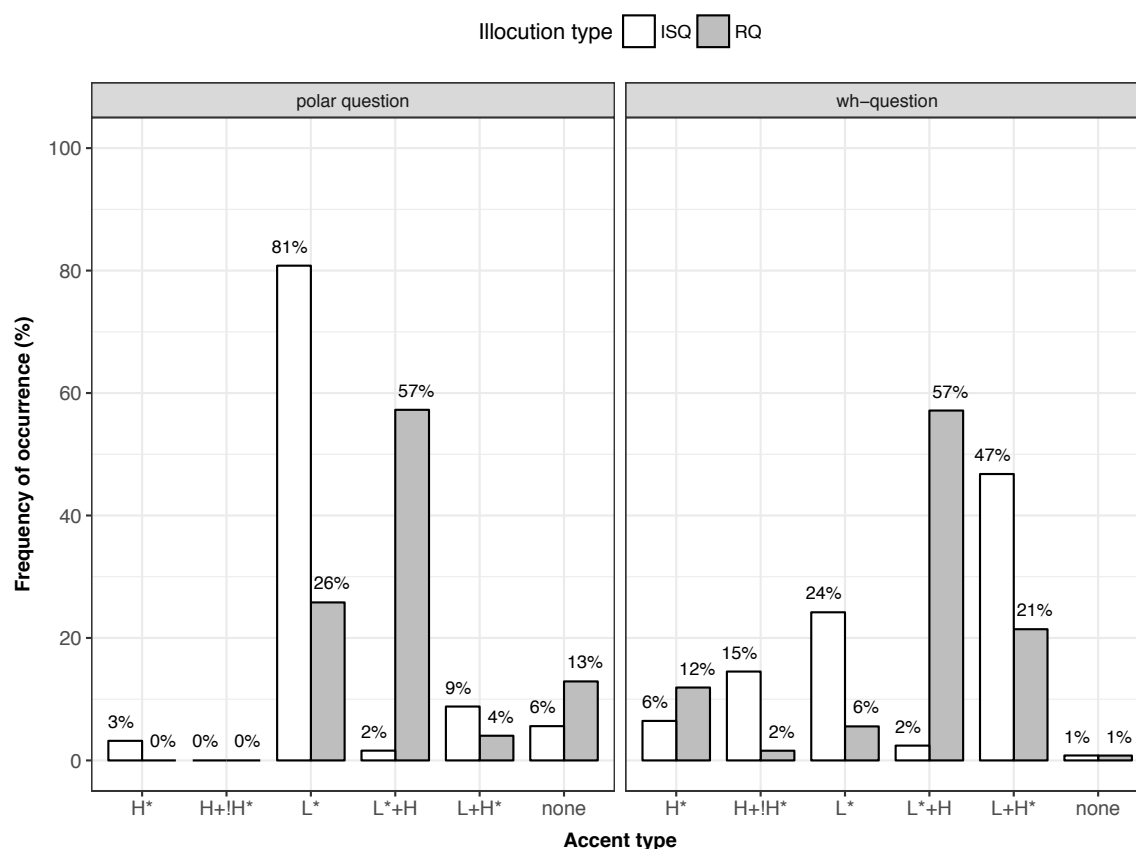
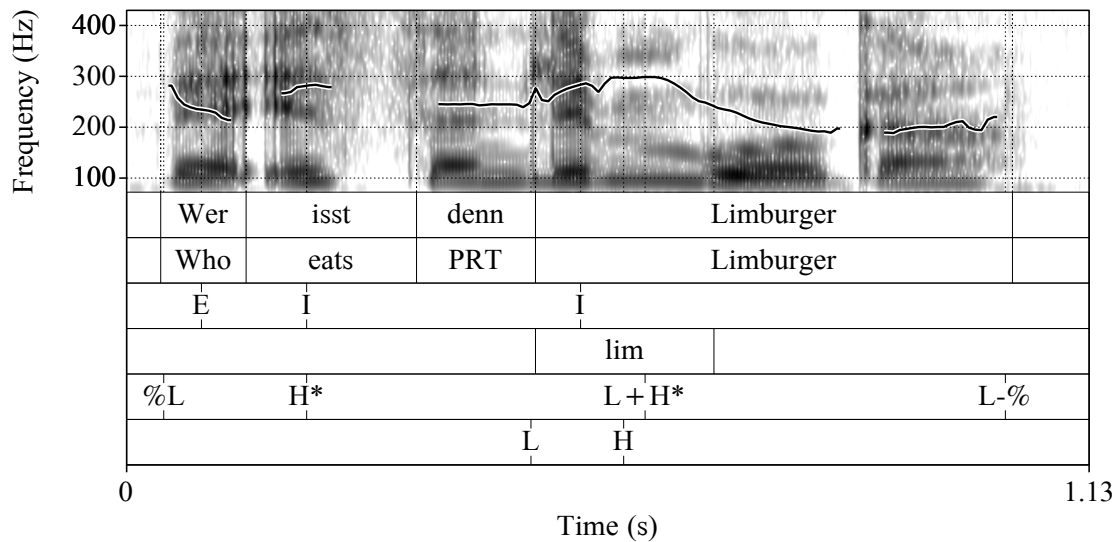


Figure 3. Distribution of frequent accent types across illocution types. Only edge tones occurring more than 10 times in one of the two illocution types were included (more than 5%). Note that the nuclear accent is generally associated with the object noun.

Please insert Figure 3 about here

In *wh*-questions, the most striking difference in pitch accent type was the contrast between nuclear L^*+H and $L+H^*$ (see Figure 3). Since this contrast has been argued to be difficult to label, we take a moment to argue for the distinction. In terms of interrater disagreements, the

two accent types were confused in only 8 of 71 cases (11%), which is not dramatic. Phonetically, the two accents clearly differed in L-alignment and in H-alignment. Alignment was calculated as the temporal distance between the respective tonal target and the start of the stressed syllable, divided by the duration of the stressed syllable (a value of .6 indicates that the tonal target is located 60% into the duration of the stressed syllable). L was aligned 23% later in L*+H than in L+H* accents (0.45 vs. 0.21 , $\beta = 0.23$, $SE = 0.03$, $df = 130.3$, $t = 7.4$, $p = p_{\text{adjusted}} < 0.0001$). The H target in L*+H was aligned 7% later than in L+H* (0.94 vs. 0.86 , $\beta = 0.07$, $SE = 0.02$, $df = 125.4$, $t = 4.4$, $p < 0.0001$, $p_{\text{adjusted}} = 0.0001$). There was no difference in the f0-excursion of the rise in st ($p = p_{\text{adjusted}} > 0.8$) and no difference in the slope of the rise in st/sec ($p = p_{\text{adjusted}} > 0.8$). Figure 4 provides an example contour for each accent type: L+H* for *wh*-ISQs (upper panel) and L*+H for *wh*-RQs (lower panel). It is noteworthy that the H is aligned within the stressed syllable in 77% of the L*+H accents, the accent that predominantly occurred in RQs. For the 23% of the cases in which H was positioned in the posttonic syllable, its distance to the stressed syllable was small: The average proportional H-alignment in these cases was 1.05 (range 1.002 to 1.16).



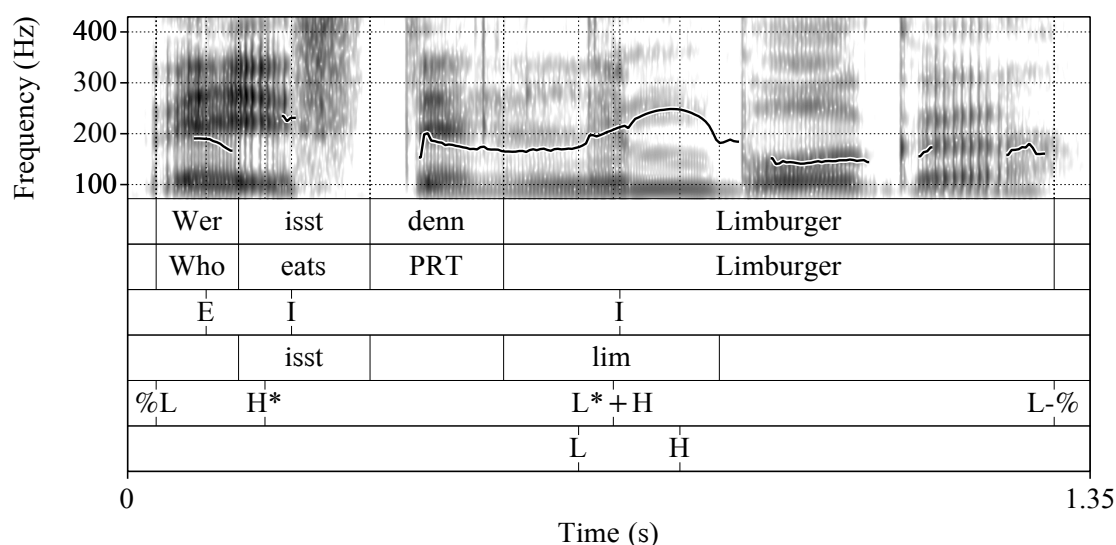


Figure 4. Examples of an L+H* nuclear accent in ISQs (upper panel) and an L*+H nuclear accent in RQs (lower panel); the nuclear accent is associated with *Limburger* (a cheese), which has antepenultimate word stress. The sixth their shows the position of the tonal targets L and H.

Please insert Figure 4 about here

Frequent nuclear tunes

Table 3 shows a list of frequent nuclear tunes across question types and illocution types, including only those nuclear tunes that occurred more than 10 times in one of the illocution types. For polar questions, the nuclear contour L*+H H-% was very frequent for RQs (68 times, 54.5%) but hardly occurred in ISQs (< 2%), rendering it a very specific contour for RQs. This contour is almost inexistent in *wh*-question (< 3 %). The contour L* H-^H% is most frequent in polar ISQs (98 times, 79.0%), but it also occurred in polar RQs (31 times, 24.8%) and in *wh*-ISQs (23 times, 19.7%). In *wh*-questions, L*+H L-% was most frequent for RQs (68 times, 54.0%). This contour was very specific for *wh*-RQs in the sense that it did not occur at all or very rarely in other question or illocution types (< 3%). In *wh*-ISQs, there was no contour that occurred more than 50% of the cases: There is a competition between L* H-^H% (25 times, 19.7%), L+H* L-% (27 times, 21.3%) and L+H* L-H% (29 times, 22.8%). This latter contour was inexistent in any of the other conditions. Generally, in the case of *wh*-ISQs, focusing on the nuclear accent alone (L+H*) gives a clearer pattern than an analysis based on nuclear tunes.

Nuclear tune	polar questions			<i>wh</i> -questions		
	RQ	ISQ	sum	RQ	ISQ	sum
	(N = 125)	(N = 124)		(N = 126)	(N = 127)	
L*+H H-%	68 (54.4%)	2 (1.6%)	70	3 (2.4%)	2 (1.6%)	5
L* H-^H%	31 (24.8%)	98 (79.0%)	129	3 (2.4%)	25 (19.7%)	28
L*+H L-%	3 (2.4%)	0	3	68 (54.0%)	0	68
L+H* L-%	2 (1.6%)	2 (1.6%)	4	28 (22.2%)	27 (21.2)	55
L+H* L-H%	0	0	0	0	29 (22.8%)	29
H+!H* L-%	0	0	0	2 (1.6%)	18 (14.2%)	20
H* L-%	0	0	0	14 (11.1%)	1 (0.8%)	15

Table 3. Overview of nuclear contours occurring more than 10 times in one of the illocution types

Prenuclear pitch accents

In polar questions, the verb was frequently unaccented with no difference between illocution types (RQs: 50 times, 46%, ISQs: 56 times, 48%, $p = p_{\text{adjusted}} > 0.5$). The next frequent accentuation patterns associated with the verb were H* (RQs: 25 times, 23%, ISQs: 30 times, 25%) and L*+H (RQs: 34 times, 27% ISQs: 21 times, 17% in ISQs). The subject was unaccented in 100% of the ISQs and in 97% of the RQs. There were no differences between illocution types for any of these accentuation patterns (all $p = p_{\text{adjusted}} > 0.4$).

In *wh*-questions, the *wh*-word was unaccented most of the times, but more so in RQs than in ISQs (RQS: 112 times, 89%, ISQs: 92 times, 73%). This effect of illocution type was significant ($\beta = 1.5$, $SE = 0.4$, $z = 3.7$, $p = 0.0002$, $p_{\text{adjusted}} = 0.0007$). Next in frequency was L* in RQs (RQs: 7 times, 6%, ISQs: 2 times, 2% in ISQs), and H* in ISQs (RQs: 4 times, 3%, ISQs: 18 times, 24%). The effect of illocution type on L* was not significant ($p = p_{\text{adjusted}} > 0.1$), but the effect on H* was ($\beta = 1.6$, $SE = 0.6$, $z = 2.9$, $p = 0.003$, $p_{\text{adjusted}} = 0.008$) The verb was mostly

unaccented, too, but slightly more so in ISQs than in RQs, a difference that approached significance (RQs: 73 times, 59%, ISQs: 85 times, 68%, $\beta = 0.6$, $SE = 0.3$, $z = 2.1$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$). If the verb was accented, it was mostly H* (RQs: 30 times, 24%, ISQs: 21 times, 17%) or L+H* (RQs: 10 times, 8%, ISQs: 13 times, 10%). There were no differences between illocution types (both $p = p_{\text{adjusted}} > 0.1$). In sum, in *wh*-ISQs, if there is a prenuclear accent, it is more often associated with the *wh*-word, whereas in RQs, it is more often associated with the verb.

Initial pitch

In 284 of the 501 recordings, there was no accent on any of the first two constituents, so these data points had to be excluded. In further 25 cases, f0 could not be extracted from the point with maximum intensity. This left 259 cases for the analysis of initial pitch (52%). There were 17 cases with high initial pitch in RQs and 15 in ISQs, a difference that was not significant ($p > 0.4$). Instead there were significantly more utterances with high initial pitch (i.e., in the top 10% of the speaker's range) in *wh*-questions than in polar questions (27 vs. 5 cases, $\beta = 1.9$, $SE = 0.5$, $z = 3.6$, $p = 0.0003$, $p_{\text{adjusted}} = 0.0008$). There was no interaction between illocution type and question type ($p = p_{\text{adjusted}} > 0.25$).

Duration

The overall duration of RQs was longer than that of ISQs, both in polar questions and in *wh*-questions (cf. Table 4). The average duration of polar RQs exceeded that of polar ISQs by 15.8%. The average duration of *wh*-RQs exceeded that of *wh*-ISQs by 17.6%.

	ISQ	RQ
polar question	1.20s (0.15s)	1.39s (0.20s)
<i>wh</i> -question	1.08s (0.15s)	1.27s (0.22s)

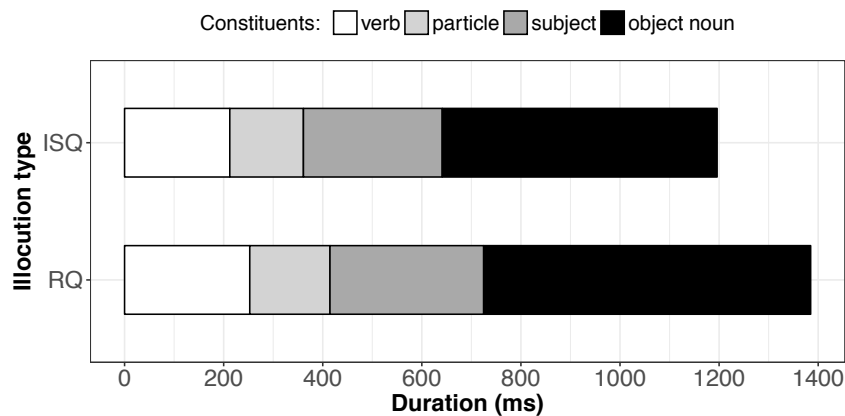
Table 4. Average duration and standard deviation of the interrogatives in the four conditions.

Results of a linear-mixed effects regression model showed significant effects of illocution type ($\beta = 0.19$, $SE = 0.03$, $df = 18.3$, $t = 6.4$, $p = p_{\text{adjusted}} < 0.0001$) and question type ($\beta = 0.12$, $SE = 0.01$, $df = 10.5$, $t = 10.0$, $p = p_{\text{adjusted}} < 0.0001$), and no interaction between the two factors ($p = p_{\text{adjusted}} > 0.5$).

To investigate whether the major constituents were lengthened proportionally, we first compared their absolute durations and then their relative durations, starting with the object

noun. The absolute durations of all constituents are plotted in Figure 5. The absolute duration of the object noun was 115ms longer in RQs than in ISQs (675ms vs. 559ms). The effect of illocution type was significant in both question types (polar questions: $\beta = 103.9$, $SE = 17.5$, $df = 17.5$, $t = 5.9$, $p = p_{\text{adjusted}} < 0.0001$; *wh*-questions: $\beta = 128.9$, $SE = 22.8$, $df = 14.3$, $t = 5.6$, $p < 0.0001$, $p_{\text{adjusted}} = 0.0002$). Furthermore, there was an interaction between illocution type and question type that approached significance (126ms vs. 96ms, $\beta = 25.8$, $SE = 12.4$, $t = 2.1$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$); it suggests stronger durational differences in *wh*-questions than in polar questions. It is, of course, conceivable that increased duration is caused not directly by illocution type, but indirectly by the choice of accent types and edge tones (e.g., words with rising pitch accents tend to be longer than other kinds of pitch accents, cf. Cambier-Langeveld & Turk, 1999). The influence of the phonological realization on phonetics is difficult to test, however, because of the uneven distribution of accent types and edge tones across illocution types. We return to this issue below, where we analyze phonetic differences in a subset of contours that occurred in ISQs and RQs alike.

The relative duration of the object noun in percent (divided by the duration of the entire interrogative, times 100) was also affected by illocution type ($\beta = 1.6$, $SE = 0.5$, $df = 12.3$, $t = 3.0$, $p = 0.02$, $p_{\text{adjusted}} = 0.03$) and question type ($\beta = 6.4$, $SE = 0.7$, $t = 12.9$, $z = 8.7$, $p = p_{\text{adjusted}} < 0.0001$), but there was no interaction between the two factors ($p = p_{\text{adjusted}} > 0.4$). This means that RQs are lengthened compared to ISQs in both question types to the same extent.⁸



⁸ One reviewer suggested we test whether the lengthening was purely phonetic or whether it involved the suppression of a phonological reduction process (i.e., schwa-elision in final unstressed syllables). We therefore coded whether schwa was elided or not in the final syllable of two items (*Limonen* 'limes' and *Schablonen* 'stencils'; e.g., [li.'mo:.nən] vs. li.'mo:.ŋ]). The results showed that schwa was realized in 62% of the ISQs and in 65% of the RQs. The lack of an effect of illocution type suggests that the lengthening was purely phonetic.

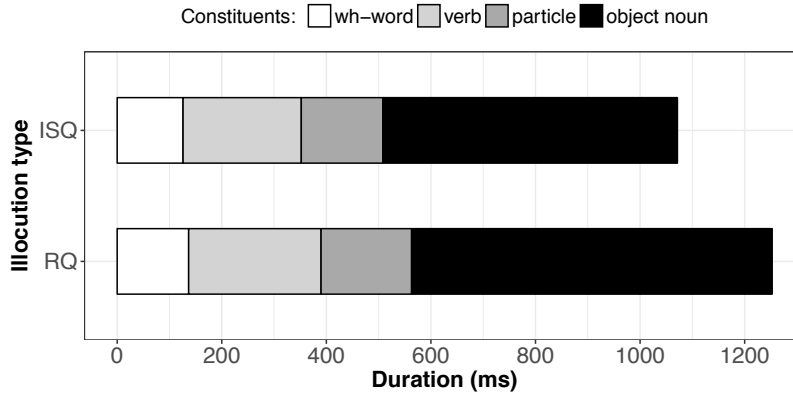


Figure 5. Average constituent durations in polar questions (upper panel) and *wh*-questions (bottom panel) in the two illocution types.

Please insert Figure 5 about here

In polar questions, the utterance-initial verb was 41ms longer in RQs than in ISQs (253ms vs. 212ms, $\beta = 37.0$, $SE = 9.6$, $df = 11.0$, $t = 3.9$, $p = 0.003$, $p_{\text{adjusted}} = 0.006$); however, the relative duration of the verb did not differ ($p = p_{\text{adjusted}} > 0.4$). The subject (*jemand* 'anyone') was also 27ms longer in RQs than in ISQs ($\beta = 28.8$, $SE = 4.1$, $df = 218.2$, $t = 7.0$, $p = p_{\text{adjusted}} < 0.0001$), but the relative duration of the subject in RQs was even slightly shorter than that of ISQs (22.3% in RQs vs. 23.4% in ISQs, $\beta = -1.0$, $SE = 0.3$, $df = 218$, $t = -3.8$, $p = 0.0002$, $p_{\text{adjusted}} = 0.0007$). There were too few prenuclear accents to statistically include accent type as a control predictor.

In *wh*-questions, the *wh*-pronouns did not differ in duration (140ms vs. 130ms, $p = p_{\text{adjusted}} > 0.15$), but the verb did: in absolute duration it was on average 29.3ms longer in RQs than in ISQs, $\beta = 29.4$, $SE = 7.8$, $df = 10.9$, $t = 3.8$, $p = 0.003$, $p_{\text{adjusted}} = 0.007$). In relative terms, the verb was less than 1% shorter in RQs than in ISQs, a difference that approached significance ($\beta = -0.8$, $SE = 0.4$, $df = 37.7$, $t = -2.1$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$).

Voice quality (perceptual classification)

Of the 1506 vowels, 78.8% were perceived as being produced with modal voice, 11.2% with glottalized voice and 10.0% with breathy voice. First, to validate the auditory voice quality classification, we tested whether breathy and modal voice differ in HNR. To this end we fitted

a linear regression model with HNR as dependent variable and label (breathy, glottalized, modal) as predictor. The results showed that vowels labelled as breathy voice had a significantly lower HNR (9.1 dB) compared to vowels labelled as modal voice (14.7 dB, $\beta = -5.6$, $SE = 0.4$, $t = -12.5$, $p < 0.0001$). The difference between breathy voice and glottalized was not significant ($p > 0.4$), but the results still largely support the validity of the manual voice quality labels. In the following, we report the proportion of breathy voice vowels across conditions and constituents (see Figure 6).

The distribution of all three categories (breathy, modal and glottalized vowels) is shown in Table A2 in the Appendix. Overall, there was a strong main effect of vowel position, with less breathy voice towards the end of the interrogative ($\chi^2(1) = 101.0$, $p = p_{\text{adjusted}} < 0.0001$, $p = p_{\text{adjusted}} < 0.0001$), and a strong main effect of illocution type ($\chi^2(1) = 25.3$, $p = p_{\text{adjusted}} < 0.0001$). Furthermore, there was an interaction between question type and illocution type ($\chi^2(1) = 5.4$, $p = 0.02$, $p_{\text{adjusted}} < 0.04$). The effect of question type was not significant ($\chi^2(1) = 3.5$, $p = 0.06$, $p_{\text{adjusted}} = 0.1$)⁹. A three-way interaction between vowel position, question type and illocution type could not be calculated due to the absence of breathy voice vowels in *wh*-ISQs. However, given the unequal distribution of breathy voice across vowel position, illocution type and question, we analyze each vowel separately. For the first vowel (vowel of the verb in polar questions and of the *wh*-word in *wh*-questions), there was a strong effect of illocution type ($\beta = 2.2$, $SE = 0.6$, $z = 4.0$, $p < 0.0001$, $p_{\text{adjusted}} = 0.0002$), no effect of question type ($p = p_{\text{adjusted}} > 0.4$), and no interaction between these two factors ($p > 0.13$). For the second vowel (subject pronoun in polar questions and verb in *wh*-questions), there was a significant effect of illocution type as well ($\beta = 1.8$, $SE = 0.5$, $t = 3.4$, $p = 0.0006$, $p_{\text{adjusted}} = 0.002$), an effect of question type ($\beta = 1.8$, $SE = 0.5$, $t = 3.2$, $p = 0.001$, $p_{\text{adjusted}} = 0.003$), but no interaction between the two factors ($p = 0.1$, $p_{\text{adjusted}} > 0.16$). For the third vowel (the object noun), the effect of illocution type approached significance ($\beta = 1.6$, $SE = 0.8$, $t = 2.0$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$). There was no effect of ($p = p_{\text{adjusted}} > 0.12$). Note that the interaction could not be calculated because there were no breathy vowels in *wh*-ISQs.

⁹ We report here first the type-III Sum-of-Squares which allow us to see whether there are main effects in addition to interactions.

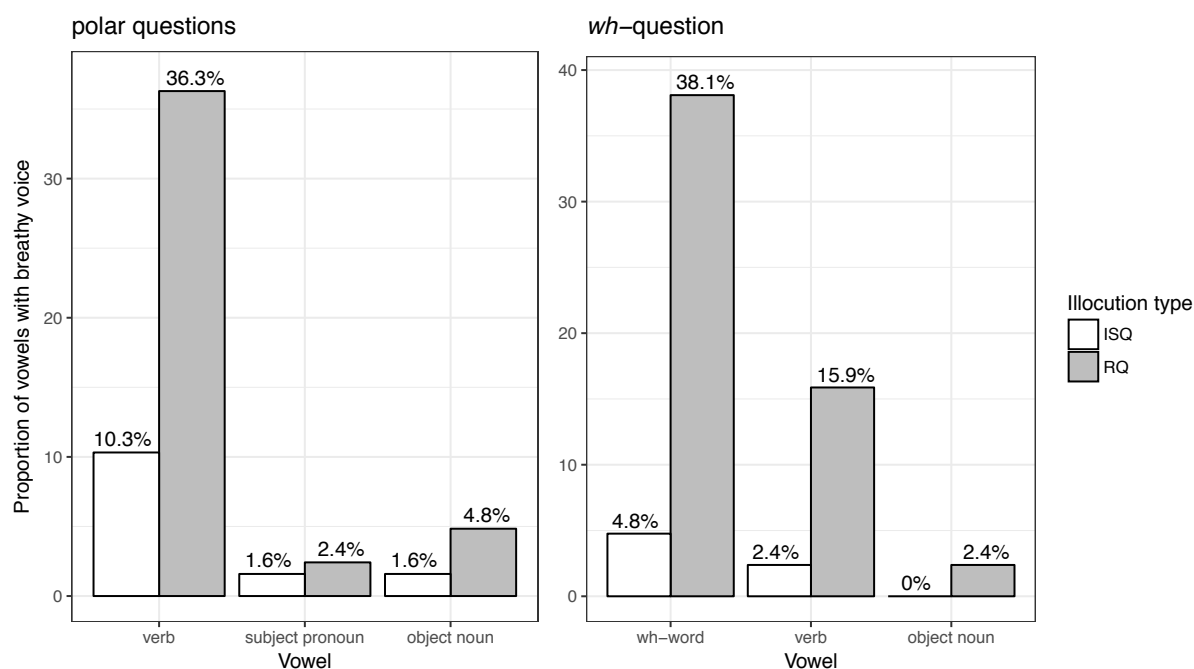


Figure 6. Percentages of vowels classified as breathy.

Please insert Figure 6 about here

Excursus: Phonetic differences of ambiguous nuclear contours

It is noticeable that there are a number of phonological contours which frequently occur in both illocution types and thus seem to be non-specific with respect to illocution type. This is particularly striking for two contours: the rising $L^* H-\hat{H}\%$ contour in polar questions (98 times in ISQs and 31 times in RQs), and the rising-falling $L+H^* L-\%$ contour in *wh*-questions (27 times in ISQs and 28 times in RQs). We take a moment to analyze these two subsets of contours more closely, because contours occurring in both illocution types are interesting for at least two reasons. First (i), they are suited to test whether phonologically identical contours may convey meaningful differences via their phonetic realization, which would suggest a direct link between pragmatics and phonetics. Second (ii), they allow us to test whether differences in phonetic realization are triggered directly by illocution type, or whether these differences occur because of a certain choice of pitch accent or boundary tone. If the phonetic realization is driven by intonational categories (e.g., longer constituent durations with L^*+H than with $L+H^*$ accents, longer durations with rising-falling than with rising contours), we expect no differences in the phonetic realization in the current subset. If, however, the phonetic realization is directly driven by illocution type, we expect the same differences in phonetic realization as in the main analysis.

Regarding (i), previous work has shown that utterances that do not distinguish illocution type by means of intonation contours differ in the phonetic realization of their contours. For instance, Michalsky (2017) compared the phonetic realization of final rises in L* H-^H% in polar questions vs. string-identical first parts of conditional statements (e.g., *Will Mone nachher zu Suse gehen*, lit: Wants Mone later to Suse go, 'Does Mone want to visit Suse later?' / 'If Mona wants to visit Suse later,'). He showed that questions had a higher f0-excursion than conditional statements (1.7 st on average). Rohloff and Michalsky (2018) tested rising-falling contours in *wh*-questions vs. nearly string-identical statements, showing that the pitch peak was on average 2st higher in questions. In the present study, we tested whether polar ISQs were realized with a greater excursion than polar RQs (parallel to the greater excursion of questions compared to statements in Michalsky's work), and whether *wh*-ISQs were realized with a higher peak scaling than *wh*-RQs. In polar questions, the f0-excursion of the final rise in L* H-^H% was larger in ISQs than in RQs, a difference that approached significance after the correction of p-values (11.1 st vs. 9.6 st, $\beta = 1.5$, $SE = 0.7$, $t = 116.9$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$). In *wh*-questions, the peak of the L+H* was on average 22.3 Hz higher in ISQs than in RQs (274.1 Hz vs. 251.8 Hz), a difference of 1.5st that was significant ($\beta = 27.8$, $SE = 5.2$, $df = 22.4$, $t = 5.3$, $p < 0.0001$, $p_{\text{adjusted}} = 0.0001$). Further analyses showed that this difference in peak scaling is due to a register difference: The f0-minimum before the rise was also higher in ISQs than in RQs (204.6 Hz vs. 186.5 Hz, $\beta = 21.1$, $SE = 7.3$, $t = 37.0$, $t = 2.9$, $p = 0.006$, $p_{\text{adjusted}} = 0.01$), but there was no effect of illocution type on the f0-excursion of the rise (5.21 st for ISQs vs. 5.22 st for RQs, $p = p_{\text{adjusted}} = 0.9$). This suggests that within *wh*-questions, *wh*-ISQs are realized in an overall higher register than *wh*-RQs.

Regarding (ii), in the subset of contours that occurred in both illocution types, we find a similar degree of lengthening of the duration of object nouns in RQs compared to ISQs. In polar ISQs, the object noun was 549ms long, compared to 632ms for RQs. This equals a lengthening of RQs of 15% compared to ISQs, and was significant ($\beta = 83.3$, $SE = 26.8$, $df = 17.2$, $t = 3.1$, $p < 0.006$, $p_{\text{adjusted}} = 0.01$). The relative duration of the object was also increased in RQs relative to ISQs ($\beta = 1.5$, $SE = 0.6$, $df = 99.8$, $t = 2.3$, $p = 0.02$, $p_{\text{adjusted}} = 0.05$). In *wh*-questions, the absolute proportional lengthening of the object in RQs relative to ISQs was 21% (683ms in RQs vs. 562ms in ISQs), a difference that was highly significant ($\beta = 85.8$, $SE = 21.9$, $df = 34.3$, $t = 3.9$, $p < 0.0004$, $p_{\text{adjusted}} = 0.001$). The relative duration of the object was also increased in RQs relative to ISQs ($\beta = 2.1$, $SE = 0.9$, $df = 34.4$, $t = 2.2$, $p = 0.04$, $p_{\text{adjusted}} = 0.07$). This suggests that illocution type directly affects the temporal structure of the constituents, rather than being influenced by the choice of the phonological contour (e.g., RQs are not longer than ISQs

because of more frequent L*+H accents than L+H* accents or because of a different distribution of boundary tones).

Discussion

This paper experimentally investigated prosodic differences between RQs and ISQs in German. The main prosodic differences between RQs and ISQs are summarized below:

Phonologically,

- polar RQs were frequently realized with H-% and polar ISQs with H-^H%. *Wh*-RQs almost exclusively ended in a low edge tone whereas *wh*-ISQs allowed for more tonal variation (L-%, L-H%, H-^H%).
- irrespective of question type, RQs were mainly realized with an L*+H nuclear accent, while polar ISQs were mostly realized with L* and *wh*-ISQs with L+H*.

Phonetically,

- irrespective of question type, RQs were realized with breathier voice quality than ISQs, in particular in the beginning of the utterance.
- irrespective of question type, RQs were generally longer than ISQs. This holds for all constituents except for the *wh*-pronouns; the object noun was particularly lengthened.

Below we evaluate the hypotheses set forth in the Introduction based on the results. We include parameters that, contrary to our predictions, did not differ between ISQs and ROs.

H1 hypothesized that polar ISQs are more often produced with a high rising edge tone (H-^H%) than polar RQs and that polar RQs are more often produced with mid level plateau H-% than polar ISQs. This was indeed the case. The distribution of edge tones in polar questions corroborates the findings of the pilot study by Wochner, et al. (2015). Importantly, given the role of the mid-high plateau H-% for RQs, our results stand in clear contrast to a simple binary distinction between final rise (H% or H-H%) vs. fall (L% or L-L%), which is typically modeled in the semantic literature (Bartels, 1999; Han, 2002). However, the high-rising H-^H% edge tone also occurred in RQs to a non-negligible extent (28% of the polar RQs). One possible explanation for the observed frequency of H-^H% in RQs in our data is its default status in German polar questions (e.g., Féry, 1993; Grice, et al., 2005; Kohler, 2004; Oppenrieder, 1988). It is conceivable that participants produced the default contour in some of the cases rather than varying the edge tone to mark illocution type. However, this explanation would not fully account for our findings, because the H-^H%, when produced in RQs, differed phonetically

from the H-[^]H% in ISQs. The rising edge tone had a smaller f₀-range in RQs than in ISQs. This suggests that participants were aware of the contrast in illocution type; they reduced the f₀-range of the high-rising edge tone H-[^]H% to signal that it is not information-seeking. We will return to this issue below when we discuss the phonetic differences in more detail. We now turn to the mid-level plateau (H-%), which was the most frequent edge tone in polar RQs (67% of the cases). Other than in polar RQs, this contour has also been referred to as *progreident* intonation to signal incomplete utterances (von Essen, 1964) and *turn-keeping* (Caspers, 1998 for Dutch). More recently, Niebuhr (2013b) related the plateau contour to reluctance on the part of the speaker to give in to a demand. It seems difficult to find a common semantic or pragmatic interpretation for these different uses of H-% (RQs, incomplete utterances, *turn-keeping*, reluctance to give in to a demand). However, it may be possible to use the same edge tone in these different contexts because the target sentences are very different otherwise, e.g., in their syntactic structure (RQ: interrogative form, elsewhere: declarative form) and position (e.g., *progreident* intonation: utterance-medially; RQs: utterance-finally).

In *wh*-questions, RQs almost always ended in a low edge tone (L-%), suggesting that *wh*-RQs with edge tones other than L-% are illicit. L-% also frequently occurred in *wh*-ISQs, but there was more variation, and in particular we found a higher proportion of rising edge tones in *wh*-ISQs (L-H%, H-[^]H%). The presence of rising edge tones in *wh*-ISQs is surprising, since rising *wh*-questions have been associated with echo-questions (Féry, 1993; Grice, et al., 2005; Isačenko & Schädlich, 1966; Pheby, 1969; von Essen, 1964). One possible explanation is that some of the *wh*-questions were interpreted as invitations or offers, which may have triggered a rising edge tone because of politeness or to signal openness towards the addressee (Kohler, 2004 for German; Sunwo, 2016 for the interpretation of edge tones in English polar questions). Future analyses of *wh*-questions with other question words are necessary to determine the source of the rising *wh*-ISQs in our corpus.

In sum, we observed a three-way distinction in edge tones, corresponding to different illocution types in syntactically interrogative utterances: L-% typically occurs in *wh*-RQs, H-% in polar RQs, and H-[^]H% in polar ISQs (cf. Wochner, et al., 2015). The only member of the quadruple (question type x illocution type) that did not have a very specific edge tone was the *wh*-ISQ. Since the propositions expressed in the experimental materials were the same across question types, it is difficult to semantically explain why *wh*-ISQs show more variability in edge tones than polar ISQs. In any case, our results show that the edge tone is not as reliably a marker of illocution types in *wh*-questions as it is in polar questions.

H2 hypothesized that *wh*-RQs are more often produced with L*+H nuclear accents than *wh*-ISQs, a hypothesis that was supported by our data. Since the L-% edge tone occurred in both *wh*-ISQs and *wh*-RQs, it is mainly the nuclear accent that distinguishes between illocution types in this question type. Furthermore, polar RQs also showed a higher proportion of L*+H accents than polar ISQs, so this accent type appears to be a prime candidate for signaling RQs, independent of question type. In ISQs, L* was most frequent in polar questions (combined with H-^H%) and L+H* was most frequent in *wh*-questions (frequently combined with L-%). Nuclear accents are often claimed to signal information structure (e.g., Baumann, et al., 2006; Brown, 1983; Ladd, 1980) and the information status of referents (e.g., Baumann & Grice, 2006). However, in our experiment we controlled for information structure such that it was the same across illocution types, and we can therefore safely exclude information structure as a possible source for differences in the choice of nuclear accent. In previous literature, the L*+H accent (followed by L-%) has been claimed to occur in self-evident statements (Kohler, 1991a, 1991b, 2004), to signal sarcasm (Lommel & Michalsky, 2017) and in prenuclear accents to signal contrastive topics (Braun, 2006). In a broader sense, L*+H may therefore have the function to signal a contrast, either in terms of lexical alternatives or in terms of propositional alternatives that are left implicit (contrastive topics and irony). In RQs, the contrast may be established to the proposition in the question (e.g., x likes lemons), for which the speaker in the current experiment knows that nobody agrees with it. Another generalization for the use of L*+H is that this accent reflects the speaker's attitude that the proposition (or the answer to an RQ) is self-evident or obvious, which, depending on the context, may signal rhetoricity or sarcasm. It is noteworthy that in many of the L*+H accents in our *wh*-RQs, both the low and the high tonal target are aligned within the stressed syllable, see Figure 4. This accent may hence constitute a new, different accent category altogether, given that in L*+H accents reported and tested elsewhere (Lommel & Michalsky, 2017) and (Braun, 2006), the peak is aligned clearly within the posttonic syllable. Future research is necessary, however, to investigate this interpretation and to test other contexts in which this special L*+H accents may occur.

In this context it makes sense to discuss the relation between rhetorical questions and emotions, also brought up by one of the reviewers for this paper. In particular, we address the issue of whether the prosodic realizations of RQs reported in this paper are specific of rhetorical questions or whether they signal emotional attitudes towards the proposition in the question radical. Generally speaking, many RQs are face-threatening and implicate a negative attitude towards the proposition in the question. Recent analyses of the Stanford Politeness Corpus (<http://www.cs.cornell.edu/~cristian//Politeness.html>) have shown that rhetorical questions are

often rated at the lower end of a politeness scale (Miriam Butt, personal communication). Therefore it is not surprising that some of the phonetic properties of RQs found in this study are similar to properties found for certain emotional attitudes (e.g., Abelin & Allwood, 2000; Banse & Scherer, 1996). However, it is in our view impossible to relate the prosodic realization to one particular emotional attitude, since the contexts used in the present experiments included a range of different triggers, among them disgust, derision, astonishment and surprise (classifications according to Plutchik, 2001). At the same time all RQs in our study are inherently attitudinal (see Neitsch, 2018 for a posthoc coding of the attitudinal strength of the RQ contexts). Therefore, we argue that the prosodic realizations are caused by illocution type, realizing that it is impossible to completely separate rhetorical illocution from attitudinal stance. More fine-grained analyses and control of emotional attitudes, ideally orthogonal to the current manipulation of illocution type, are necessary to address this aspect.

As part of the phonological analysis, we also analyzed prenuclear accents. However, there were few prenuclear accents overall and no strong differences between illocution types. In polar questions, there were no differences between illocution types at all. In *wh*-questions, we found the following difference: if a prenuclear accent was present, it was more often associated with the verb in RQs, whereas it was more often associated with the *wh*-word in ISQs. Accents on the verb are also found in polarity contrasts (Turco, et al., 2014; Turco, et al., 2013) and they may highlight a reversion of polarity from the RQ to what is actually meant. However, this assumption is highly speculative at this point because there are too few prenuclear accents to make strong generalizations. The sparsity of accents in the prenuclear region is perhaps due to our utterances being very short, which may have disfavored prenuclear accents altogether. In future studies, it will be worthwhile to include longer and possibly syntactically more complex utterances in order to trigger more prenuclear accents, which in turn will allow us to investigate phonological differences in the prenuclear area between illocution types. In addition, future perception studies will have to reveal which parts of the intonational contour (prenuclear area, nuclear accent, edge tone) contribute most to the prosodic marking and processing of illocution types. For the perception of German declarative questions vs. declarative statements, the prenuclear region played a role for listeners (Petrone & Niebuhr, 2014).

Taking a look at the complete nuclear contours of polar questions (i.e., combining pitch accent and edge tone), we observe that in polar ISQs, one tune is particularly frequent (L* H-^H%: 79% of the polar ISQs), while in polar RQs, there is a competition between two nuclear contours (L*+H H-%: 54% of the polar RQs; and L* H-^H%: 25% of the cases). In other words, while polar ISQs dispose of a very specific contour, there is variation for polar RQs. This might

suggest that RQs do not need a very specific realization but that it is instead sufficient to have a contour that differs from that of polar ISQs. However, the assumption that RQs merely need to be realized differently from ISQs is weakened by the results for *wh*-questions. Here, the rhetorical questions dispose of the most specific contour in *wh*-questions (L*+H L-% 54% of the cases in *wh*-RQs), while the most frequent contour for *wh*-ISQs (L+H* L-H%) occurred in only 23% of the *wh*-ISQs. It is, however, still possible that "being different" is enough in more general terms. After all, within question types, one illocution type comes with a typical contour, while the other one shows variation. For polar questions, more variation is found in RQs, in *wh*-questions the variation is found in ISQs.

In terms of phonetic realization, H3 predicted higher initial pitch, longer constituent durations and higher proportions of breathy voice quality in RQs than in ISQs, irrespective of question type. Our results do not support the hypothesis for initial pitch, as operationalized in Sicoli, et al. (2015). Contrary to predictions based on their cross-linguistic corpus study, RQs were *not* marked by higher initial pitch. There was only an effect of question type, with *wh*-questions starting with higher initial pitch than polar questions. This null effect of illocution type on initial pitch together with the main effect of question type suggests that parts of the effects of Sicoli, et al. (2015) may be attributable to differences in question types across conditions. In any case, the height of initial pitch, as operationalized in Sicoli, et al. (2015), does not qualify as a discriminator between ISQs and RQs in our data.

Constituent durations were longer in RQs than in ISQs, supporting H3. Differences in duration going into the same direction as in our results have also been found in previous comparisons of durations in questions and non-questions. For example, shorter constituent durations were found for declarative questions than for declarative statements (e.g., Heuven & Zanten, 2005; Niebuhr, 2013a; Niebuhr, et al., 2010). The increase in duration of RQs relative to ISQs in our data was on average 16%, which is comparable to the results of Niebuhr, et al. (2010), who reported an increase in speaking rate for questions of 10% relative to statements in monologues (and of 20% for dialogues). Furthermore, in our data, the final constituent (the object noun) was particularly lengthened, possibly to increase the salience of the object noun in RQs. The durational increase in RQs was phonetic in nature and did not have an effect on segmental phonological processes (i.e., presence or absence of schwa deletion). The phonetic differences in duration seem to be mostly triggered directly by illocution type and are not a concomitant of the choice of phonological contour, especially in polar questions. That is, when we focus on the contours that have the same nuclear tune across illocution types, we still see

the same lengthening of the object noun (relative to the prenuclear constituents) as in the entire data set.

For the analysis of voice quality, we used a perceptual classification, which is more robust to variations in vowel quality and word-prosodic structure than acoustic measures. In line with hypothesis H3, RQs were more often realized with breathy voice than ISQs. This was a main effect, which was present in both question types and in all positions. Furthermore, there were effects of vowel position and an interaction between illocution type and question type. Regarding vowel position, breathiness was most frequent at the beginning of the utterance in both question types, with a decrease of breathiness towards the end of the utterance. This is in line with studies on Irish showing that breathy voice often occurs at the beginning of the utterance (Yanushevskaya, et al., 2016). The precise reason for the positional restrictions on voice quality differences is an open issue. Some studies show that in utterance-final position, accented syllables are produced with a small open-quotient, i.e., with non-breathy voice quality (Chasaide, et al., 2013; Gobl, 1988; Yanushevskaya, et al., 2016; Yanushevskaya, et al., 2010), suggesting that accentuation reduces breathiness. Since the object nouns were mostly accented in our study, this may have blocked the use of voice quality as a cue to illocution type in this position. This only leaves the option to realize voice quality contrasts in the prenuclear area. It is also conceivable that certain word categories or syntactic positions block the spreading of breathy voice. For instance, in both polar questions and *wh*-questions, the constituent realized with more breathy voice furthest to the right in surface structure was the verb. Future research will have to test more varied syntactic structures to isolate the conditions under which breathy voice quality appears in RQs. Regarding the interaction between voice quality and question type, voice quality contrasts were used more in *wh*-questions than in polar questions. This is discussed next.

The phonetic analyses showed intriguing interactions between illocution type and question type: for both object noun duration and voice quality, the effect of illocution type was stronger in *wh*-questions than in polar questions. We interpret the stronger effects of illocution type on phonetic properties in *wh*-questions compared to polar question as a compensation for fewer intonational contrasts. In *wh*-questions, the intonational differences did not so much rely on the edge tone, since both RQs and ISQs were most frequently realized with L-% (RQs: 94%, ISQs: 44%), while polar questions show a clearer contrast in edge tones (for H-%: RQs: 67%, ISQs 1%, H-^H%: ISQ: 88%, RQs: 28%). Furthermore, the differences in nuclear accent type in *wh*-questions were not as strong as in polar questions (a more subtle contrast between L*+H vs. L+H* in *wh*-questions, i.e. a contrast of alignment only, compared to a more salient contrast

between L* and L*+H in polar questions, i.e. a contrast between monotonal and bitonal accents). This is similar to the trading relations between acoustic cues to the voicing contrast observed in fricatives (Repp, 1982). Another kind of trading relation is found between the use of phonological and phonetic realization in general. When speakers produced the same contour in the two illocution types (polar questions: L* H-^H%, *wh*-questions: L+H* L-%), there were still subtle phonetic differences: In polar questions, RQs with an H-^H% edge tone had a smaller f0-excursion than polar ISQs with an H-^H% edge tone. Similar effects of illocution type on the excursion of the final rise were reported by Michalsky (2017) for the comparison between the first part in conditional sentences and polar questions, and by Wochner and Dehé (2018) for the comparison of verb-first RQs, exclamatives and information-seeking questions (RQ < exclamatives < ISQ). For the *wh*-questions in our study, the rising-falling movement had a higher peak in ISQs than in RQs, which is in line with Rohloff and Michalsky (2018) findings on higher peaks in *wh*-questions than in declarative statements. In our data, the increased peak height is the result of a higher register for ISQs than for RQs. Hence, in cases in which phonology did not discriminate between illocution types, phonetic cues took over. It is not yet clear, however, if these phonetic differences in f0-excursion and f0-register are strong enough to affect listeners' interpretation, although it is conceivable that these differences play a role, given that phonetic differences in f0-excursion and scaling are used to distinguish illocution types in other languages (Rathcke, 2006 for Russian; Vanrell, et al., 2012 for Catalan).

Regarding the perception of cues to illocution type, it is an open issue how phonetic and phonological cues are weighted. In a first perception experiment, Neitsch, et al. (2018), tested the role of pitch accent type and voice quality for the classification of German *wh*-questions as rhetorical or information-seeking. The stimuli ended in L*+H L-% or H+!H* L-%, contours that were specific to ISQs and RQs, respectively. The object noun was realized with a breathy or modal voice quality. The L*+H L-% contour typically resulted in RQ judgments (with breathy voice: 93%, with modal voice over 61%), while H+!H* L-% resulted in mostly ISQ responses (modal voice: 92%, breathy voice: 72%). Breathy voice quality alone was hence not sufficient to cue an interrogative as RQ, and, in comparison, seems a weaker cue than the L*+H nuclear accent. Taken together, intonation interacts with voice quality, but in comparison, intonation seems to be the stronger cue to illocution type.

To conclude, *wh*-RQs are typically realized with an L*+H L-% nuclear contour, polar RQs with L*+H H-%. The common phonological denominator of RQs across question types is the L*+H nuclear pitch accent, which also plays a crucial role for the perception of RQs. For semantic descriptions, it is necessary to turn the binary contrast previously assumed with respect

to edge tones (rising vs. falling) into a three-way contrast ($H\text{-}^{\wedge}H\%$, $H\text{-}\%$, $L\text{-}\%$) and to include the nuclear pitch accent type. To complete the picture, phonetic parameters (increase in duration and breathy voice quality) must also be included. Since the current data are based on a sample with imbalanced gender distribution, it is an open question how representative the reported contours are for each gender, given that earlier, large scale studies found gender differences in the intonational realization of polar and *wh*-questions in German (Niebuhr, 2015). Moreover, the results mostly generalize to speakers from Southern Germany. Future studies will therefore have to validate our findings with (a) a larger group of speakers with a more balanced gender distribution, (b) with speakers from different varieties, and (c) with more varied materials. Finally, it will be necessary to test whether the prosodic forms found in our experiment data also surface in spontaneous productions.

Acknowledgements

We thank Sophie Egger and María Biezma for discussion of the materials, Phoebe Braunwarth, Mona Dörr, Nicole Saks, Helena Schlipf and Annika Schilk for help with testing and segmental annotation. We are grateful to Wolfgang Wokurek and Melanie Weirich for advice on voice quality measures, to María Biezma and Maribel Romero for discussion of the contexts, and to Stefan Baumann, Jan Michalsky, Oliver Niebuhr and María Biezma for discussion of the results. All remaining errors or shortcomings are our own. The research presented here was funded by the DFG as part of research unit "Questions at the Interface" (FOR 2111, project P6), grant numbers BR 3428/4-1 and DE 876/3-1.

Appendix

<i>wh</i> -questions	polar questions
Wer studiert denn Algebra? 'Who studies PRT algebra?'	Studiert denn jemand Algebra? 'Does PRT anyone study algebra?'
Wer trägt denn Angora? 'Who wears PRT magenta?'	Trägt denn jemand Angora? 'Does PRT anyone wear angora wool?'
Wer isst denn Bolognese? 'Who eats PRT pasta Bolognese?'	Isst denn jemand Bolognese? 'Does PRT anyone eat pasta Bolognese?'
Wer isst denn Garnelen? 'Who eats PRT shrimps?'	Isst denn jemand Garnelen? 'Does PRT anyone eat shrimps?'
Wer isst denn Innereien? 'Who eats PRT innards?'	Isst denn jemand Innereien? 'Does PRT anyone eat innards?'
Wer will denn Kamille? 'Who wants PRT camomile?'	Will denn jemand Kamille? 'Does PRT anyone want camomile?'
Wer tanzt denn Lambada? 'Who dances PRT Lambada?'	Tanzt denn jemand Lambada? 'Does PRT anyone dance Lambada?'
Wer mag denn Leber? 'Who likes PRT liver?'	Mag denn jemand Leber? 'Does PRT anyone like liver?'
Wer mag denn Lebertran? 'Who likes PRT cod-liver oil?'	Mag denn jemand Lebertran? 'Does PRT anyone like cod-liver oil?'
Wer schenkt denn Lilien? 'Who gives PRT lilies (as a present)?'	Schenkt denn jemand Lilien? 'Does PRT anyone give lilies (as a present)?'
Wer isst denn Limburger? 'Who eats PRT Limburger?'	Isst denn jemand Limburger? 'Does PRT anyone eat Limburger?'
Wer isst denn Limonen? 'Who eats PRT limes?'	Isst denn jemand Limonen? 'Does PRT anyone eat limes?'
Wer kennt denn Mangold? 'Who knows PRT parsnip?'	Kennt denn jemand Mangold? 'Does PRT anyone know parsnip?'
Wer mag denn Mayonnaise? 'Who likes PRT mayonnaise?'	Mag denn jemand Mayonnaise? 'Does PRT anyone like mayonnaise?'
Wer will denn ins Museum? 'Who wants PRT to go to the museum?'	Will denn jemand ins Museum? 'Does PRT anyone want to go to the museum?'
Wer liest denn Novellen? 'Who reads PRT novellas?'	Liest denn jemand Novellen? 'Does PRT anyone read novellas?'
Wer liest denn Romane? 'Who reads PRT novels?'	Liest denn jemand Romane? 'Does PRT anyone read novels?'
Wer möchte denn Rosen? 'Who would like PRT roses?'	Möchte denn jemand Rosen? 'Does PRT anyone want roses?'
Wer isst denn Rosenkohl? 'Who eats PRT Brussels sprouts?'	Isst denn jemand Rosenkohl? 'Does PRT anyone eat Brussels sprouts?'
Wer braucht denn Schablonen? 'Who needs PRT stencils?'	Braucht denn jemand Schablonen? 'Does PRT anyone need stencils?'
Wer mag denn Sellerie? 'Who likes PRT celery?'	Mag denn jemand Sellerie? 'Does PRT anyone like celery?'
Wer züchtet denn Würmer? 'Who breeds PRT worms?'	Züchtet denn jemand Würmer? 'Does PRT anyone breed worms?'

Table A1. Target interrogatives

	polar questions						<i>wh</i>-questions					
	ISQ			RQ			ISQ			RQ		
	b	g	m	b	g	m	b	g	m	b	g	m
Vowel 1	10.3	11.9	77.8	36.3	8.9	54.8	4.8	15.9	79.4	38.1	11.9	50.0
Vowel 2	1.6	4.8	93.6	2.4	1.6	96.0	2.4	7.9	89.7	15.9	15.1	69.0
Vowel 3	1.6	25.4	73.0	4.8	14.5	80.6	0	6.3	93.7	2.3	10.3	87.3

Table A2. Percentages of breathy (b), glottalized (g) and modal (m) voice quality in three vowels (*wh*-word, verb and object in *wh*-questions, and verb, subject and object in polar questions).

References

- Abelin, A., & Allwood, J. (2000). *Cross-linguistic interpretation of emotional prosody*. Paper presented at the Proceedings of the ISCA ITRW on Speech and Emotion, Belfast, Northern Ireland.
- Agresti, A. (2002). *Categorical data analysis*. New York: John Wiley.
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personal and Social Psychology*, 70, 614-636.
- Banuazizi, A., & Cresswell, C. (1999). *Is that a real question? Final rises, final falls, and discourse function in yes-no question intonation*. Paper presented at the 35th Annual Meeting of the Chicago Linguistics Society Chicago.
- Bartels, C. (1999). *The Intonation of English Statements and Questions. A Compositional Interpretation*. New York & London: Garland Publishing.
- Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. R. (2015). Parsimonious mixed models. *arXiv preprint, arXiv:1506.04967*.
- Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, 38, 1636-1657.
- Baumann, S., Grice, M., & Steindamm, S. (2006). *Prosodic marking of focus domains - categorical or gradient?* Paper presented at the 3rd International Conference on Speech Prosody, Dresden, Germany.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 289-300.
- Biezma, M., & Rawlins, K. (2017). Rhetorical questions: Severing asking from questioning. In D. Burgdorf, J. Collard, S. Maspong & B. Stefánsdóttir (Eds.), *Proceedings of SALT 27* (pp. 302-322).
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, 17, 97-110.
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer (Version 6.0.40). <http://www.praat.org/>, retrieved 11 May 2018.
- Braun, B. (2005). *Production and Perception of Thematic Contrast in German*. Frankfurt am Main: Peter Lang Publishing.
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4), 451-493.
- Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 67-77). Berlin: Springer.
- Cambier-Langeveld, T., & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs English. *Journal of Phonetics*, 27(3), 255-280.
- Cangemi, F., & D'Imperio, M. (2013). Tempo and the perception of sentence modality. *Laboratory Phonology*, 4, 191-219.
- Caponigro, I., & Sprouse, J. (2007). Rhetorical questions as questions. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung 11* (pp. 121-133). Barcelona: Universitat Pompeu Fabra.
- Caspers, J. (1998). Who's next? The melodic marking of questions vs. continuation in Dutch. *Language and Speech*, 41(3-4), 375-398.
- Chasaide, A. N., Yanushevskaya, I., Kane, J., & Gobl, C. (2013). *The voice prominence hypothesis: the interplay of F0 and voice source features in accentuation*. Paper presented at the INTERSPEECH.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Cooper, W. E., Eady, S. J., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer pairs. *Journal of Acoustic Society of America*, 77(6), 2142-2156.

- Dainora, A. (2006). Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. Whalen & C. Best (Eds.), *Laboratory Phonology 8* (pp. 107 – 132). Berlin, New York: Mouton de Gruyter.
- De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, 32(493-516).
- de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language, and Hearing Research*, 36(2), 254-266.
- de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research*, 38, 794-811.
- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.
- Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36, 680-703.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. <https://CRAN.R-project.org/package=irr>.
- Gobl, C. (1988). Voice source dynamics in connected speech. *STL-QPSR*, 1(1988), 123-159.
- Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, 26, 129-143.
- Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 191, 267-298.
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In J. Sun-Ah (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing* (pp. 55-83). Oxford: Oxford University Press.
- Grice, M., Reyelt, M., Benz Müller, R., Mayer, J., & Batliner, A. (1996). *Consistency in transcription and labelling of German intonation with GToBI*. Paper presented at the 4th International Conference on Spoken Language Processing, Philadelphia, USA.
- Gutiérrez Rexach, J. (1998). Rhetorical questions, relevance and scales. *Revista Alicantina de Estudios Ingleses*, 11, 139-155.
- Han, C.-H. (2002). Interpreting interrogatives as rhetorical questions. *Lingua*, 112, 201-229.
- Hanssen, J. (2017). *Regional variation in the realization of intonation contours in the Netherlands*. Radboud University, LOT.
- Hedberg, N., & Sosa, J. M. (2007). The prosody of topic and focus in spontaneous English dialogue. In C. Lee, M. Gordon & D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 101-120). Dordrecht: Springer.
- Hedberg, N., Sosa, J. M., Gürgülü, E., & Mamani, M. (2010). *Prosody and pragmatics of wh-interrogatives*. Paper presented at the 2010 Meeting of the Canadian Linguistics Association.
- Heuven, V. J. v., & Haan, J. (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 61-86). Berlin: Mouton de Gruyter.
- Heuven, V. J. v., & Zanten, E. v. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*(47), 87-99.
- Hudson, R. A. (1975). The meaning of questions. *Language*, 51(1), 1-31.
- Ilie, C. (1995). *The validity of rhetorical questions as arguments in the courtroom*. Paper presented at the Special Fields and Cases. Proceedings of the Third International Conference on Argumentation, Amsterdam.
- Ilie, C. (2010). Rhetorical questions. In L. Cummings (Ed.), *The Pragmatics Encyclopedia* (pp. 405-408). London, New York: Routledge.
- Isačenko, A., & Schädlich, H. (1966). *Untersuchungen über die deutsche Satzintonation*. Berlin: Deutsche Akademie der Wissenschaften zu Berlin.
- Kipp, A., Wesenick, M.-B., & Schiel, F. (1996). *Automatic Detection and Segmentation of Pronunciation Variants in German Speech corpora*. Paper presented at the ICSLP 1996, Philadelphia, USA.
- Kisler, T., Schiel, F., & Sloetjes, H. (2012). *Signal processing via web services: the use case of WebMAUS*. Paper presented at the Proceedings of Digital Humanities, Hamburg, Germany.
- Kohler, K. J. (1991a). A model of German intonation. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 25, 295-360.

- Kohler, K. J. (1991b). 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)*, 25, 115-185.
- Kohler, K. J. (2004). Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. In G. Fant, H. Fujisaki, J. Cao & Y. Xu (Eds.), *From traditional phonology to modern speech processing - Festschrift für Professor Wu Zongji's 95th Birthday* (pp. 205-215). Beijing: Foreign Language Teaching and Research Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. <https://CRAN.R-project.org/package=lmerTest>.
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Lommel, N., & Michalsky, J. (2017). Der Gipfel des Spotts. Die Ausrichtung von Tonhöhengipfeln als intonatorisches Indiz für Sarkasmus [Peak alignment as intonational cue to sarcasm]. In N. Levkovich & A. Urdze (Eds.), *Diversitas Linguarum* 42. Bremen: Universitätsverlag Dr. N. Brockmeyer.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. R., & Bates, D. M. (2017). Balancing type 1 error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315.
- Meibauer, J. (1986). *Rhetorische Fragen*. Tübingen: Niemeyer.
- Michalsky, J. (2017). *Frageintonation im Deutschen. Zur intonatorischen Markierung von interrogativität und Fragehaltigkeit*. Tübingen: Niemeyer.
- Neitsch, J. (2018). *The prosody of rhetorical questions in consideration of context*. Paper presented at the Workshop "Meaning in non-canonical questions", Konstanz, Germany.
- Neitsch, J., Braun, B., & Dehé, N. (2018). *The role of prosody for the interpretation of rhetorical questions in German*. Paper presented at the 9th International Conference on Speech Prosody, Poznan, Poland.
- Niebuhr, O. (2013a). The acoustic complexity of intonation. In E.-L. Asu & P. Lippus (Eds.), *Nordic Prosody XI* (pp. 15-29). Frankfurt: Peter Lang.
- Niebuhr, O. (2013b). *Resistance is futile – The intonation between continuation rise and calling contour in German*. Paper presented at the 14th Annual Conference of the International Speech Communication Association, Lyon, France.
- Niebuhr, O. (2015). *Gender differences in the prosody of German questions*. Paper presented at the Proceedings of the International Congress of the Phonetic Sciences, Glasgow, UK.
- Niebuhr, O., Bergherr, J., Huth, S., Lill, C., & Neuschulz, J. (2010). Intonationsfragen hinterfragt! - Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Frageätzen mit deklarative Syntax [On the complexity of prosodic differences between declaratives and declarative questions]. *Zeitschrift für Dialektologie und Linguistik*, 77, 304-346.
- Oppenrieder, W. (1988). Intonatorische Kennzeichnung von Satzmodi. In H. Altmann (Ed.), *Intonationsforschungen* (pp. 169-206). Tübingen: Niemeyer.
- Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, 57(1), 108-146.
- Pheby, J. (1969). *Intonation und Grammatik im Deutschen*. Berlin: Akademie-Verlag.
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Unpublished PhD Thesis, MIT, Bloomington.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge: MIT Press.
- Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344-350.
- Pützer, M., & Wokurek, W. (2015). Acoustic- and EGG-parametrisations of Phonatory Quality Provide Voice Profiles of Normal Speakers. *Laryngorhinootologie*, 94(5), 303-310.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (Eds.). (1985). *A Comprehensive Grammar of the English Language* (Vol. Longman): New York.

- Rathcke, T. (2006). A perceptual study on Russian questions and statements. In J. M. Harrington, C. Mooshammer & F. Kleber (Eds.), *Arbeitsberichte des Instituts für Phonetik und Kommunikation (AIPUK)* (Vol. 37, pp. 51-62). Kiel, Germany.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81-110.
- Rohloff, M., & Michalsky, J. (2018). *Pitch scaling as a question cue in German wh-questions*. Paper presented at the Conference on Phonetics & Phonology in German-speaking countries, Berlin, Germany.
- Sadock, J. M. (1971). *Queclaratives Papers from the seventh regional meeting, April 16-18, 1971*: Chicago Linguistics Society.
- Sadock, J. M. (1974). *Toward a Linguistic Theory of Speech Acts*. New York, San Francisco, London: Academic Press.
- Schaffer, D. (2005). Can rhetorical questions function as retorts?: Is the Pope Catholic? *Journal of Pragmatics*, 37(4), 433-460.
- Schiel, F., Draxler, C., & Harrington, J. (2011). *Phonemic segmentation and labelling using the MAUS technique*. Paper presented at the New Tools and Methods for Very-Large-Scale Phonetics Research, University of Pennsylvania.
- Schourup, L. C. (1985). *Common discourse particles in English conversation*. London: Routledge.
- Sicoli, M. A., Stivers, T., Enfield, N. J., & Levinson, S. C. (2015). Marked initial pitch in questions signals marked communicative function. *Language and Speech*, 58(2), 204-223.
- Simpson, A. P. (2009). *Breathiness differences in male and female speech. Is H1-H2 an appropriate measure?* Paper presented at the Proceedings of FONETIK.
- Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, 40(3), 477-490.
- Sunwo, J. (2016). *Conventions in prosody for affective meanings: non-canonical terminal contours in English polar interrogatives*. Paper presented at the Proceedings of the 7th International Conference on Speech Prosody, Boston, USA.
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis—jitter, shimmer and hnr parameters. *Procedia Technology*, 9, 1112-1122.
- Thurmair, M. (1989). *Modalpartikeln und ihre Kombinationen*. Tübingen: Niemeyer.
- Thurmair, M. (1991). *Zum Gebrauch der Modalpartikel 'denn' in Fragesätzen. Eine korpusbasierte Untersuchung*. Tübingen: Niemeyer.
- Truckenbrodt, H. (2012). Semantics of intonation. In C. Maienborn, K. von Stechow & P. Portner (Eds.), *Semantics. An International Handbook of Natural Language Meaning* (pp. 2039-2069). Berlin: de Gruyter.
- Turco, G., Braun, B., & Dimroth, C. (2014). When contrasting polarity, the Dutch use particles, Germans intonation. *Journal of Pragmatics*, 62(94-106).
- Turco, G., Dimroth, C., & Braun, B. (2013). Intonational means to mark verum focus in German and French. *Language and Speech*, 56, 460-490.
- Turk, A., Satsuki, N., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schließer (Eds.), *Methods in Empirical Prosody Research* (pp. 1-28.). Berlin, New York: De Gruyter.
- Vanrell, M. M., Mascaró, I., Torres-Tamarit, F., & Prieto, P. (2012). Intonation as an encoder of speaker's certainty: information and confirmation yes-no questions in Catalan. *Language and Speech*, 56(2), 163-190.
- von Essen, O. (1964). *Grundzüge der Hochdeutschen Satzintonation*. Ratingen: Henn Verlag.
- Wilson, D., & Sperber, D. (1988). Mood and the analysis of non-declarative sentences. In J. Dancy, J. M. E. Moravcsik & C. C. W. Taylor (Eds.), *Human Agency: Language, Duty and Value* (pp. 77-101). Stanford, CA: Stanford University Press.
- Wochner, D., & Dehé, N. (2018). *Prosody meets pragmatics: a production study on German verb-first sentences*. Paper presented at the Speech Prosody, Poznan, Poland.
- Wochner, D., Schlegel, J., Braun, B., & Dehé, N. (2015). *The prosody of rhetorical questions in German*. Paper presented at the Proceedings of Interspeech, Dresden, Germany.

- Yanushevskaya, I., Chasaide, A. N., & Gobl, C. (2016). *The interaction of long-term voice quality with the realisation of focus*. Paper presented at the Speech Prosody.
- Yanushevskaya, I., Gobl, C., Kane, J., & Ní Chasaide, A. (2010). *An exploration of voice source correlates of focus*. Paper presented at the Eleventh Annual Conference of the International Speech Communication Association.
- Yu, J., & Zahner, K. (2018). *Truncation and compression in Southern German and Australian English*. Paper presented at the Proceedings of Interspeech, Hyderabad, India.