Proc. of the 17th Speech Science and Technology Conference (SST 2018), Sydney, Australia. 125-128.

Classification of interrogatives as information-seeking or rhetorical questions

Bettina Braun, Daniela Wochner, Katharina Zahner, Nicole Dehé

Department of Linguistics, University of Konstanz, Germany

{bettina.braun, daniela.wochner, katharina.zahner, nicole.dehe}@uni-konstanz.de

Abstract

Rhetorical questions (RQs) differ from information-seeking questions (ISQs) in that they do not require an answer and instead seek commitment of the addressee regarding the underlying proposition. We tested the prosodic differences between ISQs and RQs in German and showed that polar RQs were mainly realized with a high plateau (H-% in the GToBI system) and polar ISQs with a high-rise (H-^H%). *Wh*-RQs almost exclusively ended in a low edge tone whereas *wh*-ISQs showed more tonal variation (L-%, L-H%, H-^H%). Irrespective of question type, RQs were mainly produced with L*+H accents. Phonetically, RQs were – compared to ISQs – more often realized with breathy voice quality in the beginning of the utterance and with longer constituent durations. The object noun was particularly lengthened. Classification and regression trees showed that polar questions can be classified on the basis of the boundary tone alone, and *wh*-questions by an interaction between phonological events and duration. We discuss the findings with respect to the phonology-semantics interface.

Index Terms: rhetorical question, information-seeking question, prosody, classification, German

1. Introduction

This paper deals with the prosody of rhetorical and information-seeking questions in German in terms of intonational events and phonetic realization, and addresses the question of which of the parameters are needed to automatically classify utterances with an interrogative form as either rhetorical or information-seeking. RQs can have the form of a polar question (with subject-verb inversion in German, see (1)), or a *wh*-question with a fronted *wh*-element (see (2)). Other question types are also possible, but here we only investigate polar questions and *wh*-questions.

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

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

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 [1]. In contrast, RQs do not require or expect answers from the addressee as stated by several authors [2-8]. Instead, they are used when the answer is obvious or at least inferable to all discourse participants [1, 9-11]. Moreover, the purpose of RQs is to seek the addressee's commitment to the proposition that is presupposed by the question [7]. Other functions of RQs are to change a topic or to engage the audience in monologues or retorts, (e.g., *Is the Pope Catholic*?, cf. [10], [12]). These other functions are not 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. [13], [7]), and, in German, by the discourse particles *schon* and *auch*[1], cf. [14], [7]. These discourse particles are sufficient to trigger an RQ interpretation, but they are not necessary. That is, RQs and ISQs can be string-identical on the surface and can be disambiguated by the context (3), as well as by their prosodic realization.

(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?

Previous pilot data from German [15] showed that polar RQs have a higher proportion of high plateaus (H-% in the GToBI annotation system [22]) than polar ISQs, which were typically produced with a high rise (H-^H%). *Wh*-questions generally ended in a fall (L-%), with a higher proportion of L*+H nuclear accents in *wh*-RQs than in *wh*-ISQs. RQs were also produced with longer constituent durations than ISQs and had a breathier voice quality. Here we present results from a more controlled production experiment. Based on the pilot data and on claims made in the literature on English [6, 8, 16, 17], we tested the following prosodic parameters in the realization of string-identical ISQs and RQs in German: nuclear pitch accent type, boundary tone, voice quality in the major constituents (verb, subject, and object noun in polar questions; *wh*-word, verb, and object noun in *wh*-questions), constituent durations (here operationalized as speech rate), and voice quality. In this paper, we focus on the usefulness of these parameters for the automatic classification of illocution type (RQ vs. ISQ). Given previous claims (often inaccurate) about the meaning of boundary tones in previous relevant literature [8, 16], our results are highly relevant to semantic modeling, as well as to the extraction of the functions of interrogatives in human-computer interaction.

[1] These particles also have a lexical meaning (*schon:* 'already', *auch*: 'also, too'). The lexical meaning does not trigger RQs.

2. Production data

2.1. Methods

2.1.1. Materials

We constructed 11 *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 pronominal subject *anyone* and adapted the syntactic structure to verb-first (V1). 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. Within RQs, the set of propositions denoted by the *wh*-interrogative and the set of propositions denoted by the polar interrogative with the indefinite subject are roughly the same.

For each interrogative pair, we constructed two contexts, one triggering an information-seeking interpretation of the interrogative and one triggering a rhetorical one. An example of the resulting quadruple is given in Table 1. To control for information structure and specifically to avoid effects of information structure on nuclear accent position and type, each context introduced the predication expressed in the sentence radical (e.g., *liking celery* in Table 1), rendering the referents of the constituents in the verb phrase discourse-given (see [18] for more details).

Table 1. *Contextual settings for polar and* wh-*questions in both illocution types (ISQ, RQ); contexts and target interrogatives are translated from German.*

ISQ	**RQ**
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:
Mag denn jemand Sellerie? 'Does anyone like celery?'	
***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:
Wer mag denn Sellerie? 'Who likes celery?'	

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 a certain proposition (e.g., *you know that nobody likes celery*). 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 lexical stress on the penultimate or antepenultimate syllable. All target interrogatives contained the modal particle *denn,* which frequently occurs in both question types in German [19]. 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 [20]. Hence, the illocution of the target was determined only by the contextual information.

As fillers, we used six questions with structural (PP-attachment) ambiguities, each of which occurred in two contexts. In addition, we constructed 22 exclamatives with V1 word order, i.e., the same word order as in polar questions.

2.1.2. 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 interrogative pair. The 34 filler items were added to each list. The experimental lists were randomized anew for each participant with the constraint that two readings of a target interrogative were separated by at least four other trials. Each experiment started with four familiarization trials, followed by a short break in which participants were allowed to ask 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 the 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 that 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).

2.1.3. Participants

Twelve monolingual speakers of German (average age=21.7 years, SD=2.3; 10 female, 2 male) participated for a small payment. They were students at the University of Konstanz and unaware of the purpose of the study. The participants were randomly assigned to one of the two lists (6 in each list). None of them reported any speaking or hearing disorders.

2.1.4. Data Treatment

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 259 polar (RQ: 124, ISQ: 125) and 253 (RQ: 127, ISQ: 126) *wh*-questions.

The files were annotated at the word level using standard segmentation criteria [21] in the software package Praat [22]. Voice quality was classified as modal, breathy or glottalized in the initial word (verb in polar questions, *wh*-word in *wh*-questions), the second constituent (subject in polar questions, verb in *wh*-questions), and the final object noun. A perceptual classification was deemed more robust than acoustic measures, given variation in the 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 [23, 24]. The annotations were done by three trained annotators with substantial interrater reliability (kappa>0.71, [25]).

The continuous variables were analyzed with linear mixed effect regression models with illocution-type (ISQ vs. RQ) as fixed factor and participants and items as crossed random factors (adjustment of intercepts). Random slopes were added if this improved the fit of the models. P-values were calculated using the Satterthwaite approximation of degrees-of-freedom. Categorical variables were coded as 0 or 1 and analyzed using logistic mixed models. To avoid Type I errors, p-values were adjusted by means of the Benjamini-Hochberg correction [26].

2.2. Results

The production data resulted in the following differences between illocution types [18]. Phonologically, polar RQs typically ended in H-% (67%) and polar ISQs in H-^H% (88%). *Wh*-RQs almost exclusively ended in a low edge tone L-% (94%) whereas *wh*-ISQs exhibited more tonal variation (L-%: 44%, L-H%: 28%, H-^H%: 25%). Irrespective of question type, RQs were mainly realized with an L*+H nuclear accent (polar: 57%; *wh*: 57%), while polar ISQs were mostly realized with L* (81%) and *wh*-ISQs with L+H* (47%). Phonetically, irrespective of question type, RQs were realized more often with breathier voice quality than ISQs in the first word (37% vs. 7%, p<0.0005). Furthermore, RQs were on average 190ms longer than ISQs (p<0.0005), a lengthening of 15% relative to ISQs. The object noun was over-proportionally lengthened.

We trained separate classification and regression trees for the two question types (using the package rpart in R [27]). To this end, we excluded parameters that occurred less than 5 times in the illocution type where it was most frequent. The initial model included all phonological and phonetic properties. Instead of absolute object duration, which is dependent on the lexical material, we included speech rate (number of intended phones per second). Acoustic measures of voice quality (harmonics-to-noise ratio and H1*-A3*) were also included [28, 29]. To avoid overfitting, the resulting tree was pruned using the complexity parameter that resulted in the lowest cross-validation error [30]. To test the generalizability of the tree, we used a 10-fold cross-validation procedure (splitting the data in 10 random sets, training the tree on 8 sets and testing it on 2 sets). Accuracy was calculated on the 20% unseen data.

For polar questions, three data points with !H-% were excluded. Utterances ending in H-% were classified as RQ with only one exception (83 of 84 items were classified correctly). The other edge tones were mostly classified as ISQ (117 of 157). For the unseen data, classification accuracy was 87.5%.

For *wh*-questions, we removed eight data points (with the rare accent !H* and the rare boundary tones H-% and !H-%). The classification results showed that the initial split was caused by accent type, see Figure 1. Further factors were speech rate (duration of the final object) and the final boundary tone. An unseen test set (20% of the data) was classified correctly in 85.4% of the cases.

3. Discussion and Conclusion

The results of the production study showed that illocution type affected both intonational information (nuclear pitch accent types and boundary tones) and phonetic parameters (object duration and voice quality) in polar questions and *wh*-questions. For automatic classification, only the final boundary tone was used for polar questions. Utterances ending in H-% were classified as RQs in 99% of the cases; all other boundary tones as ISQs. RQs were classified more accurately than ISQs, which either suggests a very specific prosodic realization of RQs and more variability in ISQs, or a classification bias towards RQs, or both. Possibly, ISQs allow for the coding of other facets, such as politeness, emotional attitude, etc. Semantically, it is sufficient to use the boundary tone to model RQs, in line with previous semantic approaches, but against those approaches, it is a high plateau rather than a falling contour that specifies RQs. Similar intonational findings have been reported for English [31].

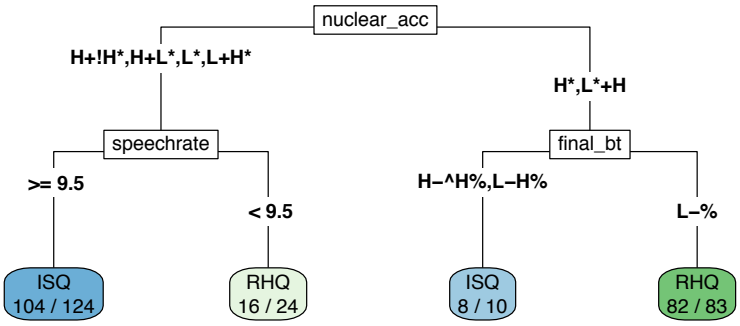

Figure 1. *Result of a pruned classification and regression tree for* wh*-questions. The values below the classified labels indicate the probability of the fitted class.*

For *wh*-questions, the initial split was caused by accent type, i.e., an intonational parameter. Further splits were due to accent type and speech rate. This suggests an interaction of phonological and phonetic cues for the classification of *wh*-questions. The interplay between phonetic and phonological parameters shows that *wh*-questions cannot be modeled based on intonational parameters alone [against 8, 16]. Recent perception studies point in the same direction. [32] tested the role of pitch accent type and voice quality in German *wh*-questions (L*+H L-% vs. H+!H* L-%, produced with breathy vs. 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%). Hence, pitch accent type and voice quality are additive cues. In future work, we plan to use the current findings from automatic classification to derive further hypotheses for perception. In particular, the classification results suggest that speech rate may be a useful discriminator for certain accent types in *wh*-questions. Furthermore, we plan to test the classifier on non-experimental data. We will also include data from other languages to probe the language-specificity of these parameters.

4. Acknowledgements

This project was funded by the German Research Foundation (DFG) with grants to Braun (BR 3428/4-1) and Dehé (DE 876/3-1) as part of research unit "Questions at the Interface".

5. References

[1] Caponigro, I. and Sprouse, J., "Rhetorical questions as questions," *Proceedings of Sinn und Bedeutung 11*, Puig-Waldmüller, E., Ed., Barcelona: Universitat Pompeu Fabra, 2007, pp. 121-133.

[2] Hudson, R. A., "The meaning of questions," *Language,* vol. 51, pp. 1-31, 1975.

[3] Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J., Eds., *A Comprehensive Grammar of the English Language*. New York, 1985.

[4] Wilson, D. and Sperber, D., "Mood and the analysis of non-declarative sentences," in *Human Agency: Language, Duty and Value*, Dancy, J., Moravcsik, J. M. E., and Taylor, C. C. W., Eds., Stanford, CA: Stanford University Press, 1988, pp. 77-101.

[5] Ilie, C., "The validity of rhetorical questions as arguments in the courtroom," in *Special Fields and Cases. Proceedings of the Third International Conference on Argumentation*, Amsterdam, 1995, pp. 73-88.

[6] Banuazizi, A. and Cresswell, C., "Is that a real question? Final rises, final falls, and discourse function in yes-no question intonation," *Proceedings of the 35th Annual Meeting of the Chicago Linguistics Society* Chicago, 1999, pp. 1-14.

[7] Biezma, M. and Rawlins, K., "Rhetorical questions: Severing asking from questioning," *Proceedings of SALT 27*, Burgdorf, D., Collard, J., Maspong, S., and Stefánsdóttir, B., Eds., 2017, pp. 302-322.

[8] Han, C.-H., "Interpreting interrogatives as rhetorical questions," *Lingua,* vol. 112, pp. 201-229, 2002.

[9] Ilie, C., "Rhetorical questions," in *The Pragmatics Encyclopedia*, Cummings, L., Ed., London, NY: Routledge, 2010, pp. 405-408.

[10] Sadock, J. M., *Toward a Linguistic Theory of Speech Acts*. New York, San Francisco, London: Academic Press, 1974.

[11] Sadock, J. M., "Queclaratives," in *Papers from the Seventh Regional Meeting, April 16-18, 1971*, Chicago Linguistics Society, 1971.

[12] Schaffer, D., "Can rhetorical questions function as retorts?: Is the Pope Catholic?," *Journal of Pragmatics,* vol. 37, pp. 433-460, 2005.

[13] Gutiérrez Rexach, J., "Rhetorical questions, relevance and scales," *Revista Alicantina de Estudios Ingleses,* vol. 11, pp. 139-155, 1998.

[14] Meibauer, J., *Rhetorische Fragen*. Tübingen: Nieymeyer, 1986.

[15] Wochner, D., Schlegel, J., Dehé, N., and Braun, B., "The prosodic marking of rhetorical questions in German," *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, 2015, pp. 987-991.

[16] Bartels, C., *The Intonation of English Statements and Questions. A Compositional Interpretation*. New York & London: Garland Publishing, 1999.

[17] Hedberg, N., Sosa, J. M., Gürgülü, E., and Mameni, M., "Prosody and pragmatics of wh-interrogatives," *Porceedings of the 2010 Meeting of the Canadian Linguistics Association*, 2010.

[18] Braun, B., Dehé, N., Neitsch, J., Wochner, D., and Zahner, K., "The prosody of rhetorical and information-seeking questions in German," submitted.

[19] Thurmair, M., *Zum Gebrauch der Modalpartikel 'denn' in Fragesätzen. Eine korpusbasierte Untersuchung*. Tübingen: Niemeyer, 1991.

[20] Thurmair, M., *Modalpartikeln und ihre Kombinationen*. Tübingen: Niemeyer, 1989.

[21] Turk, A., Satsuki, N., and Sugahara, M., "Acoustic segment durations in prosodic reserach: A practical guide," in *Methods in Empirical Prosody Research*, Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I*., et al.*, Eds., Berlin, New York: De Gruyter, 2006, pp. 1-28.

[22] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer," http://www.praat.org/, retrieved 11 May 2018, 2018.

[23] Grice, M. and Baumann, S., "Deutsche Intonation und GToBI," *Linguistische Berichte,* vol. 191, pp. 267-298, 2002.

[24] Grice, M., Baumann, S., and Benzmüller, R., "German intonation in autosegmental-metrical phonology," in *Prosodic Typology. The Phonology of Intonation and Phrasing*, Sun-Ah, J., Ed., Oxford: Oxford University Press, 2005, pp. 55-83.

[25] Landis, J. R. and Koch, G. G., "The measurement of observer agreement for categorical data," *Biometrics,* vol. 33, pp. 159-174, 1977.

[26] Benjamini, Y. and Hochberg, Y., "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society,* vol. 57, pp. 289-300, 1995.

[27] Therneau, T. and Atkinson, B., "rpart: Recursive partioning and regression trees," https://CRAN.R-project.org/package=rpart, 2018.

[28] de Krom, G., "Some spectral correlates of pathological breathy and rough voice quality for differnet types of vowel fragments," *Journal of Speech and Hearing Research,* vol. 38, pp. 794-811, 1995.

[29] Keating, P. A. and Esposito, A., "Linguistic voice quality," *UCLA Working Papers in Phonetics U6,* vol. 105, pp. 85-91, 2007.

[30] Baayen, H. R., *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press, 2008.

[31] Dehé, N. and Braun, B., "The prosody of rhetorical questions in English," submitted.

[32] Neitsch, J., Braun, B., and Dehé, N., "The role of prosody for the interpretation of rhetorical questions in German," *Proceedings of the 9th International Conference on Speech Prosody*, Poznan, Poland, 2018.