



The prosodic marking of rhetorical questions in Standard Chinese

Katharina Zahner¹, Manluolan Xu, Yiya Chen², Nicole Dehé¹, Bettina Braun¹

¹Department of Linguistics, University of Konstanz, Germany, ²Leiden University Center for Linguistics, the Netherlands

katharina.zahner | nicole.dehe | bettina.braun@uni-konstanz.de, xumanluolan@163.com, yiya.chen@hum.leidenuniv.nl

Abstract

We investigated the prosody of rhetorical questions (RQs) as compared to string-identical information-seeking questions (ISQs) in Standard Chinese – a language in which f₀ is considerably constrained by lexical tone. Our results show that overall RQs have a lower mean f₀ than ISQs. F₀ is also locally modified (on the first and last constituent) to mark illocution type. Additionally, RQs have longer durations than ISQs and show more instances of glottalized voice, mainly towards the end of the interrogative. Hence, similar to intonation languages, Standard Chinese uses prosody to distinguish between these two illocution types. Our findings hence suggest f₀, duration, and voice quality to be cross-linguistic signals of rhetorical meaning, with their implementation being language-specific.

Index Terms: lexical tone, prosody, intonation, rhetorical questions, Standard Chinese

1. Introduction

This paper studies the prosodic differences between rhetorical questions (RQs) and information-seeking questions (ISQs) in Standard Chinese. ISQs are interrogatives that request information from the addressee in order to close a knowledge gap (e.g., [1, 2]). RQs share their syntactic form with ISQs while clearly serving a different purpose: RQs do not necessarily require an answer [3, 4] but function to make a point [5]. RQs have further been defined as committing the interlocutor to the answer that is presupposed in the RQ [3] – often of opposite polarity from what was asked [5]. Signals to rhetorical meaning are, among others, syntactic or lexical cues (e.g., discourse particles like German *schon* [6] or Standard Chinese *nandao* [7]).

Importantly, prosody has also been shown to distinguish between the two illocution types. Specifically, recent studies, which systematically investigated the prosodic differences between string-identical ISQs and RQs (polar and *wh*-questions) in various languages, showed that RQs differ from ISQs in regard to duration, voice quality, and intonation ([8] on German, [9] on English, [10, 11] on Icelandic, [12] on French, [13] on Japanese): RQs are longer in duration [8-10, 12, 13] and show more instances of breathy voice [8, 9]. In terms of intonation, RQs may differ from ISQs in the position of the pitch accent [9], the type of accent [8, 10], as well as the type of edge tone [8, 9, 12].

The present paper extends this line of research to Standard Chinese (as spoken in Beijing). Standard Chinese provides an interesting test case in regard to f₀ marking of illocution type, because f₀ is considerably constrained by lexical tone. In Standard Chinese, every syllable carries one of four lexical

tones: Tone 1 (T55, high-level), Tone 2 (T35, rising), Tone 3 (T214, low-rising) and Tone 4 (T51, falling) – higher numbers indicate a higher pitch level –, or the neutral tone [14].

While f₀ is used to distinguish lexical meaning in Standard Chinese, it also serves other purposes, such as the marking of information-structure and interrogativity, cf. [15, 16] for recent overviews. Under focus, tones are commonly realized with a larger f₀ range; if not focussed, they are compressed, e.g., [17-20]. Beyond f₀, longer durations [21, 22] and hyperarticulated segmental contrasts [23] also mark focus in Standard Chinese. In terms of f₀ marking of interrogativity, polar-ISQs have been reported to be produced with higher f₀ than string-identical declaratives [24-26]. Moreover, a question-induced final rise has been discussed in the literature (see [16] for a summary). The phonetic implementation of the final rise preserves the overall shape of the lexical tones, but for instance reduces the range in the falling Tone 4 or enhances the range in the rising Tone 2 [16]. Furthermore, polar ISQs have shorter durations than declaratives [24-26]. Likewise, *wh*-ISQs (where *shénme* means 'what') have been shown to exhibit higher f₀ compared to string-identical declaratives (where *shénme* means 'a little bit of' or 'whatever'), mostly towards the end [27, 28]. [28] further showed an increased f₀ range in *shénme* for interrogatives. *Wh*-ISQs are also shorter than declaratives [27, 28]. Taken together, we know that Standard Chinese uses f₀ and other prosodic cues to signal sentence-level linguistic functions – despite f₀ also serving a lexical function. Thus far, however, no study directly compares the prosodic encoding of ISQs and RQs in a tone language. Given that accent placement and accent type as well as edge tones are important for non-tonal languages to differentiate the two illocution types, the remaining question is how a tonal language like Standard Chinese encodes the contrast between RQs and ISQs. We predict that:

H1: Standard Chinese ISQs and RQs differ in their prosodic realization with respect to f₀-related features, both globally (overall f₀ contour) and locally (f₀ modifications on specific constituents).

H2: Given the constraints of lexical tone on possible f₀ modifications for illocutionary acts, Standard Chinese ISQs vs. RQs are also distinguished by other prosodic cues (i.e., duration and voice quality).

2. Production experiment

To test H1 and H2, a production experiment was designed (cf. [8]) and adapted to Chinese. Participants read short contexts presented on screen (triggering either an ISQ reading or an RQ reading). These contexts were followed on screen by the target interrogatives (polar and *wh*-questions), which the participants produced. The experiment was run in Beijing in spring 2018.

2.1. Methods

2.1.1. Materials

Twenty-two polar and 22 *wh*-questions were constructed, along with two contexts for each question (one context eliciting an ISQ reading, the other an RQ reading), resulting in 22 context-question quadruplets, see Tab. 1. Quadruplets were translated from [8] by a native speaker of Standard Chinese (2nd author) and adapted where necessary. In the ISQ version of the context, the answer was not known to the speaker but would have filled a knowledge gap (*you would like to know*). By contrast, in the RQ version of the context, the answer was obvious from the context (*it is known that*). The interrogatives were felicitous in both illocutions and contained an object noun that was non-constraining as to one of the readings (e.g., *lemons*), as verified by a pre-test. The particle *me* is commonly used in polar questions [29] and was thus included in both polar ISQ and RQ.

Table 1: Exemplar question-context quadruplet. For convenience, contexts are given in English.

Polar question			
Context for ISQ		Context for RQ	
At a party, you offer cake made with lemons. <i>You would like to know</i> which of the guests like this fruit and whether they would like some or not.		Your aunt offers lemons to her guests. However, <i>it is known that</i> this fruit is too sour to be eaten on its own.	
You say to your guests:		You say to your cousin:	
有人	吃	柠檬	么?
Yǒurén	chī	níngméng	me?
Anyone	eat	lemon	PRT?
‘Does anyone eat lemons?’			
Wh-question			
Context for ISQ		Context for RQ	
At a party, you offer cake made with lemons. <i>You would like to know</i> which of the guests like this fruit and would like some of it.		Your aunt offers lemons to her guests. However, <i>it is known that</i> this fruit is too sour to be eaten on its own.	
You say to your guests:		You say to your cousin:	
谁	吃	柠檬	
Shéi	chī	níngméng	
Who	eat	lemon	
‘Who eats lemons?’			

The first constituent (i.e., *yǒurén* ‘anyone’ in polar and *shéi* ‘who’ in *wh*-questions) was the same across the board, but verbs and object nouns varied – also in terms of number of syllables and lexical tones. Moreover, the same predication (e.g., *eating lemons*) was used for polar and *wh*-questions. For the object noun, lexical tone was distributed such that all four tones occurred in the last syllable of the noun (6 times Tone 1, 6 times Tone 2, 4 times Tone 3, and 6 times Tone 4, in both polar and *wh*-questions). Additionally, 34 fillers (structurally ambiguous declaratives, exclamatives, alternative questions and neutral polar questions), and their contexts, were translated from [8].

2.1.2. Procedure

Two experimental lists were constructed, each containing both question types and both illocution types. Each list contained half of the polar questions ($N = 22$; 11 in an ISQ and 11 in an RQ reading) and half of the *wh*-questions ($N = 22$; 11 in an ISQ and 11 in an RQ reading) and all 34 fillers. One of the two lists was randomly assigned to each participant. Each participant received a randomized order of items, with the constraint of separating the same question (in the 2 readings) by at least 4

other items. Three practice trials preceded 78 trials (44 experimental and 34 fillers). Participants received oral instructions by the experimenter (2nd author). The experiment was controlled in *Presentation* [30] and lasted 25 to 30 minutes.

On each trial, participants silently read a context displayed on screen. Upon button press, the target interrogative appeared on screen and the recording started. Participants were instructed to read each context carefully and to produce the subsequent interrogatives in a way that they were suitable in the given context. In case of a mistake, they were allowed to produce the sentence again. Upon another button press, a new trial started and the recording for the previous target stopped. The productions were recorded using a headset microphone (Shure SM10A) and digitized on a computer (44.1 kHz, 16 Bit, stereo).

2.1.3. Participants

Ten native speakers of Standard Chinese (all female, av. age = 26.5 years; $SD = 2.0$), born and raised in Beijing, participated. The data of two additional speakers was not considered as they did not speak the Beijing variety of Standard Chinese. This was done to minimize the influence of dialectal variation.

2.1.4. Data treatment

In total, 440 target interrogatives were produced (44 contexts x 10 participants). 17 interrogatives (3.9%) were excluded from the analysis because of technical errors (2), mispronunciation (3), or pauses between the constituents (12). The final data set comprised 216 polar questions (107 ISQs, 109 RQs) and 207 *wh*-questions (103 ISQs, 104 RQs). All interrogatives were annotated in Praat [31], see Fig. 2 (ISQ top, RQ bottom panel).

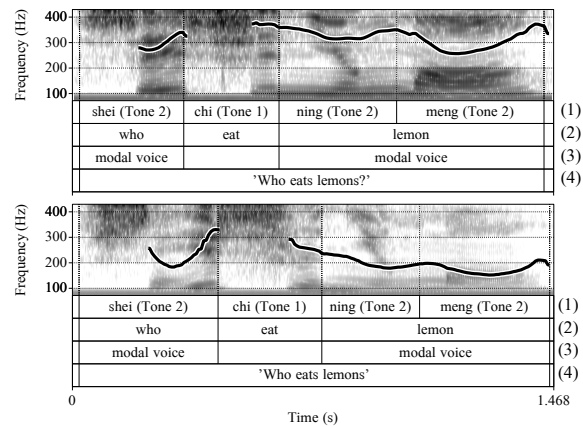


Figure 2: Example *wh*-question pair (by participant09) and annotation (ISQ top panel, RQ bottom panel).

Word and syllable boundaries were manually set based on standard segmentation criteria [32] by a native speaker of Standard Chinese (2nd author), see tiers 1 and 2 in Fig. 2. Another native speaker annotated voice quality on the first and last constituent in the interrogative (tier 3), based on perceptual classification (modal, breathy, glottalized voice), cf. [8]. Reliability checks for voice quality analysis showed “almost perfect” agreement ($\kappa = 0.94$; 98% agreement) [33–35] (based on 50% of data additionally annotated by the 1st author). Pitch tracking errors were manually corrected (1st author), i.e., erroneous pitch points were removed in the Praat Manipulation editor [31] and the corrected Manipulation object was saved as a wav-file (Pitch overlap-add). From these files, mean f_0 , constituent durations, and voice quality labels were

automatically extracted using a Praat script. Furthermore, f0 values (10 per syllable) were extracted using Prosody Pro [36].

Mean f0 and durations were statistically analysed using linear mixed effects models (*lmers*) in R [37]. Each model contained *question type* and *illocution type* as fixed effects (interaction term and main effects) and *subjects* and *items* as crossed random factors [38]. Random slopes were added and retained if the model fit improved [39] (based on a comparison of LogLikelihoods). Voice quality labels were analysed using logistic mixed effects models (*glmers*) coding glottalized voice as 1 and modal voice as 0 (breathy voice did not occur). The *glmer* modelling procedure was the same as for *lmers*. In both types of models, p-values were obtained using the Satterthwaite approximation implemented in lmerTest [40] and subsequently adjusted based on the Benjamini-Hochberg correction [41]. Below we will report both original and adjusted p-values (p_{adj}). Note that interactions between *illocution type* and *question type* were not significant in any of the models, which is why we report main effects only.

For the investigation of local modifications in the f0 contour, we fitted general additive mixed models (GAMMs) [42, 43] for the first and last constituents. GAMMs can account for the continuous nature of the f0 contour via smooth functions, which predict the data as closely as possible. The visualization of the predicted values by the GAMMs gives the time period in which two contours differ as a function of a predictor. In our GAMMs, the response variable was the raw f0 value at different time points (10 measurements per syllable extracted by ProsodyPro [36]). The models included *illocution type* as a parametric effect, along with a factor smooth for the interaction of *illocution type* over time, $s(\text{Normtime}, \text{by} = \text{illocution type})$. A smooth for subjects (random intercept) was also included. All models were corrected for auto-correlation in the data using the `acf_resid()` function from the package *itsadug* [44]. For model fitting of the GAMMs, we used the R package *mgcv* [42, 45], for visualization of the model *itsadug* [44].

2.2. Results

2.2.1. Global and local f0 characteristics

Globally, RQs were produced with significantly lower mean f0 in both *question types* ($\beta = -25.14$, $SE = 5.26$, $df = 15.26$, $t = -4.79$, $p < 0.001$, $p_{adj} < 0.01$): polar-ISQs: 261.2Hz (SD = 26.0) vs. polar-RQs: 239.3Hz (SD = 30.7); *wh*-ISQs: 262.1Hz (SD = 28.3) vs. *wh*-RQs: 233.5Hz (SD = 31.7).

Regarding local f0 modifications, we concentrate on the first constituent, i.e., *yǒurén* 'anyone' in polar and *shéi* 'who' in *wh*-questions, which has the same tone, and the last syllable in the object noun, in which all four lexical tones were equally distributed. For the first constituent (*yǒurén* in polar, *shéi* in *wh*-questions), Fig. 3 shows the estimated f0 values in the two illocution types (left panel) and the estimated difference in f0 (RQ minus ISQ, right panel), as predicted by the GAMMs. The grey band indicates the 95% confidence interval (CI) of the mean difference. Values below 0 indicate lower f0 in RQs. The difference in f0 between RQs and ISQs is significant if the 95% CI does not include 0 (see red interval(s) on the x-axis). The difference curves (Fig. 3, right) reveal that the contours differ significantly for the whole constituent. The predicted values (Fig. 3, left) suggest that RQs are significantly lower and show a greater f0 range than ISQs for both *yǒurén* and *shéi*. The increased range seems to be due to a lowering of the low target (\sim Normtime 10 for polar and Normtime 7 for *wh*-questions).

For the final part of the interrogatives, we analysed the last syllable of the noun together with the particle *me* for polar questions, and the last syllable of the noun in *wh*-questions – split by lexical tone in the final syllable of the noun, see Fig. 4 for GAMM results. The analyses revealed an interaction between *lexical tone* and *illocution type* for both question types (all $p < 0.01$), suggesting that the difference in f0 differed depending on tone. The predicted values (Fig. 4, left panels) suggest RQs to be overall lower, with the shape of the lexical tone being preserved. The difference in f0 across illocution types is smallest for Tone 3.

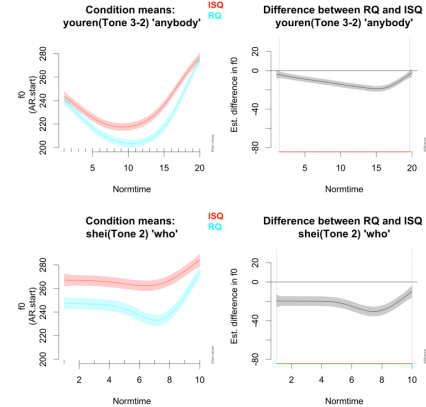


Figure 3: Predicted f0 values by the GAMMs (left columns) and predicted difference (RQ-ISQ, right columns) for *yǒurén* in polar (top) and *shéi* in *wh*-questions (bottom).

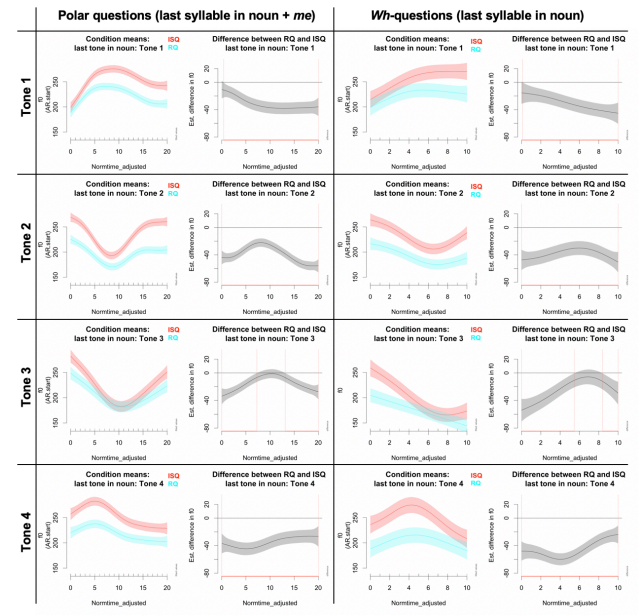


Figure 4: Predicted f0 values by the GAMM (left figure columns) and predicted difference in f0 (RQ-ISQ, right figure columns) for the last two syllables in polar questions (final syllable in noun together with particle *me*) and the final syllable in *wh*-questions.

2.2.2. Duration

There was an effect of *illocution type* on overall interrogative duration ($\beta = 163$, $SE = 0.03$, $df = 8.75$, $t = 5.17$, $p < 0.001$, $p_{adj} < 0.01$), with RQs being longer than ISQs (polar-ISQs: 1337ms (SD = 265) vs. polar-RQs: 1501ms (SD = 309); *wh*-ISQs: 1114ms (SD = 248) vs. *wh*-RQs: 1281ms (SD = 302)), which

was independent of the last lexical tone in the object noun (interaction *tone* x *illocution type*: $p = 0.58$; $p_{adj} = 0.73$). All constituents were longer for RQs than for ISQs (all $p < 0.05$; $p_{adj} < 0.05$) except for the sentence-final particle *me* in polar questions ($p = 0.23$, $p_{adj} = 0.36$), see Fig. 6. In both question types, the relative difference across illocution type was biggest for the sentence-initial constituent, with *yǒurén* 'anyone' being 24% longer in polar-RQs, and *shéi* 'who' being 28% longer in *wh*-RQs (lengthening in other constituents for RQs $< 11\%$).

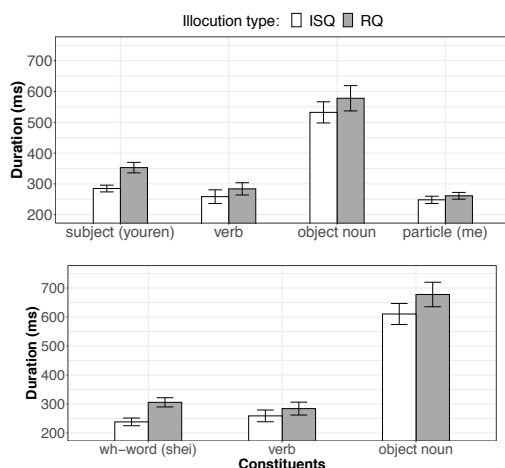


Figure 6: Average constituent durations in polar questions (top) and *wh*-questions (bottom) in the two illocution types.

2.2.3. Voice quality

The first constituent (*yǒurén* in polar and *shéi* in *wh*-questions) was realized with glottalized voice in 7% of the cases in polar-ISQs, in 14% of the cases in polar-RQs, in 3% in *wh*-ISQs, and in 8% in *wh*-RQs. All other instances were realized with modal voice. This difference in glottalized instances across *illocution type* was significant ($\beta = -1.04$, $SE = 0.44$, $z = -2.36$, $p < 0.05$, $p_{adj} < 0.05$). The final constituent (*me* in polar, object noun in *wh*-questions), in turn, was realized with glottalized voice in 5% of the cases in polar-ISQs, in 18% of the cases in polar-RQs, in 21% in *wh*-ISQs, and in 43% in *wh*-RQs. All other instances were modal. This difference in glottalized instances across *illocution type* was also significant ($\beta = -1.67$, $SE = 0.54$, $z = -3.08$, $p < 0.01$, $p_{adj} < 0.01$). Figure 7 shows the distribution of voice quality labels in the last constituent for *wh*-questions, split by the final lexical tone in the object noun. In the last constituent in *wh*-questions, glottalized voice is more frequent in Tone 3 and Tone 4, as compared to Tone 1 or 2 (all $p < 0.01$, $p_{adj} < 0.01$). The difference in glottalization across *illocution type* is smallest for Tone 3 ($p = 0.42$, $p_{adj} = 0.58$); it is significant for Tone 4 ($p < 0.01$, $p_{adj} < 0.05$) and approaches significance for Tone 1 ($p = 0.09$, $p_{adj} = 0.17$). For Tone 2 it could not be calculated as in ISQs there were no glottalized instances.

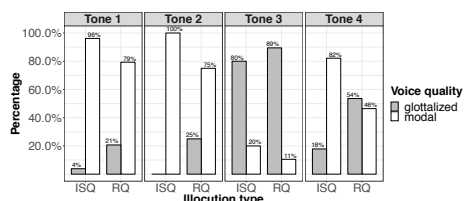


Figure 7: Distribution of voice quality for the last constituent in *wh*-questions, split by final lexical tone in the object noun.

3. Discussion and Conclusions

We studied the prosodic marking of illocution type (ISQ vs. RQ) in Standard Chinese. Confirming H1, RQs were globally realized with lower f_0 than ISQs. Given that such a difference is also found between Standard Chinese declaratives and interrogatives [24-28, 46], it will be interesting to compare Standard Chinese RQs to string-identical declaratives in future studies. Along with the difference in overall f_0 , there were local f_0 modifications, further confirming H1: The first constituent showed a wider pitch range in RQs than in ISQs, mainly due to a lowering in the low tonal target. At the end of the interrogative (last two syllables in polar questions and last syllable in *wh*-questions), RQs also showed lower pitch while preserving the overall shape of the tone (cf. [47]). For interrogatives ending in Tone 3, the difference in f_0 in the last syllable(s) across illocution type was smallest. This lack of f_0 demarcation for Tone 3 might have to do with the shape of the tone itself, which is a) the most complex and b) involves the lowest tonal target. It might hence be more difficult to lower f_0 even further in RQs. The increased occurrence of glottalized voice in this tone might have enhanced the perceptual low tone and in that way represent a strategy to encode illocution type (cf. [21] on the smaller capacity for f_0 range change under focus for Tone 3). From the perspective of perception, it will be interesting to test whether the identification of illocution type in Standard Chinese differs across lexical tones. Supporting H2, we found that Standard Chinese uses both duration and voice quality to distinguish between illocution types: Replicating findings on duration in other languages [8-13], RQs were produced with longer duration, with the first constituent showing the largest relative difference across illocutions. RQs also showed more instances of glottalized voice, especially for the final constituent (particle *me* in polar and object in *wh*-questions). Future research will clarify whether this phenomenon is caused by the lowered f_0 in RQs and thus a byproduct of f_0 modifications [48]. In sum, Standard Chinese clearly uses f_0 and other prosodic cues to encode illocution type, similar to the marking of focus [17-22] and interrogativity [24-28]. Gender-specific differences have not been found (or explicitly tested) in previous studies [24-28]. Future research will have to include male speakers and examine potential gender-specific differences for the realization of RQs.

The current findings, together with those from other languages [8-13], allow conclusions about the use of prosodic cues to RQs in typologically different languages. Specifically, longer duration seems to be a stable cue to RQs. F_0 and voice quality are also consistently used to distinguish between ISQ and RQ, but their implementation seems to underlie language-specific principles (for f_0 : pitch accent type and edge tone in intonation languages vs. global and local f_0 modifications in a tone language; for voice quality: breathy voice in some, glottalized voice in other languages to signal RQs). Clearly, research on other tone languages is needed to corroborate our findings. Taken together, on the basis of RQs, our study contributes to our understanding of the prosody-pragmatics interface from a cross-linguistic perspective.

4. Acknowledgements

We are grateful to Meiqiao Li for help with annotation, Jasmin Rimpler for data preparation, and Haiping Long, Tian Li, Jun Liu, Mengzhu Yan, and the audience at P&P 2019 in Düsseldorf for discussion. The research was funded by the DFG as part of research unit 'Questions at the Interfaces' (FOR 2111, P6), grant numbers BR 3428/4-1,2 and DE 876/3-1,2.

5. References

- [1] Groenendijk, J.A.G. and M.J.B. Stokhof, *Studies on the semantics of questions and the pragmatics of answers*. 1984, University of Amsterdam.
- [2] Meibauer, J., *Rhetorische Fragen*. 1986, Tübingen, Germany: Niemeyer.
- [3] Biezma, M. and K. Rawlins. *Rhetorical questions: Severing asking from questioning*. SALT 27. 2017.
- [4] Caponigro, I. and J. Sprouse. *Rhetorical questions as questions*. Sinn und Bedeutung. 2007. Barcelona, Spain.
- [5] Han, C.-H., *Interpreting interrogatives as rhetorical questions*. Lingua, 2002. 112: p. 201–229.
- [6] Bayer, J. and H.-G. Obenauer, *Discourse particles, clause structure, and question types*. The Ling Review, 2011. 28.
- [7] Xu, B. *Nandao-questions as a special kind of rhetorical question*. Semantics and Linguistic Theory (SALT). 2013.
- [8] Braun, B., et al., *The prosody of rhetorical and information-seeking questions in German*. Language and Speech, 2019. 62: p. 779-807.
- [9] Dehé, N. and B. Braun, *The prosody of rhetorical questions in English*. English Lang and Linguistics (onlinefirst), 2019.
- [10] Dehé, N. and B. Braun, *The intonation of information seeking and rhetorical questions in Icelandic*. Journal of Germanic Linguistics, To Appear 2020. 32.
- [11] Dehé, N., B. Braun, and D. Wochner. *The prosody of rhetorical vs. information-seeking questions in Icelandic*. Proc. of Speech Prosody, 2018. Poznań, Poland.
- [12] Delais-Roussarie, E. and C. Beyssade, *Information-seeking vs. rhetorical questions: From gradience to categoricity*. Paper at Phonetics and Phonology in Europe (PaPE). 2019: Lecce, Italy.
- [13] Miura, I. and N. Hara, *Production and perception of rhetorical questions in Osaka Japanese*. JPhon, 1995. 2: p. 291-303.
- [14] Chao, Y.R., *Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese*, in *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*, M. Halle, et al.. 1956, Mouton Publishers: The Hague, The Netherlands. p. 52–59.
- [15] Xu, Y., *Prosody, Tone and Intonation*, in *The Routledge Handbook of Phonetics*, W.F. Katz and P.F. Assmann, Editors. 2019, Routledge: New York. p. 314-356.
- [16] Chen, Y., *Tone and intonation*, in *Cambridge Handbook of Chinese Linguistics*, C.-R. Huang, Y.-H. Lin, and I.-H. Chen. Forthcoming, Cambridge UP: Cambridge.
- [17] Xu, Y. and C. Xu, *Phonetic realization of focus in English declarative intonation*. JPhon, 2005. 33: p. 159-197.
- [18] Xu, Y., *Effects of tone and focus on the formation and alignment of f(0) contours*. JPhon, 1999. 27(1): p. 55–105.
- [19] Chen, Y. and B. Braun. *Prosodic realization in information structure categories in Standard Chinese*. Proc. of Speech Prosody. 2006. Dresden, Germany.
- [20] Jin, S., *An acoustic study of sentence stress in Mandarin Chinese*. 1996: Columbus: OSU.
- [21] Chen, Y. and C. Gussenhoven, *Emphasis and tonal implementation in Standard Chinese*. JPhon, 2008. 36: p. 724-746.
- [22] Chen, Y., *Durational adjustment under corrective focus in Standard Chinese*. JPhon, 2006. 34: p. 176-201.
- [23] Chen, Y., *The acoustic realization of vowels of Shanghai Chinese*. JPhon, 2008. 36: p. 629-648.
- [24] Yuan, J., *Mechanisms of question intonation in Mandarin*, in *Chinese Spoken Language Processing*, Q. Huo, et al. 2006, Springer. p. 19-30.
- [25] Liu, F. and Y. Xu, *Parallel encoding of focus and interrogative meaning in Mandarin Intonation*. Phonetica, 2005. 62: p. 70-87.
- [26] Lee, O.J., *The prosody of questions in Beijing Mandarin*. 2005, The Ohio State University.
- [27] Liu, X., A. Li, and Y. Jia. *How does prosody distinguish wh-statement from wh-question? A case study of standard Chinese*. Proc. of Speech Prosody. 2016. Boston, USA.
- [28] Yang, Y., *The two sides of wh-indeterminates in Mandarin: A prosodic and processing account*. 2018, Utrecht: LOT.
- [29] Liing, W.-J., *How to ask questions in Mandarin Chinese*, in *Graduate Center*. 2014, City University of New York: CUNY Academic Works.
- [30] Presentation, N.S. 2000: Berkeley, CA.
- [31] Boersma, P. and D. Weenink. *Praat: doing phonetics by computer [Computer program]*. Version 5.3.84, 2014.
- [32] Turk, A.E., S. Nakai, and M. Sugahara, *Acoustic segment durations in prosodic research: A practical guide*, in *Methods in Empirical Prosody Research*, S. Sudhoff, et al. 2006, Walter de Gruyter: Berlin, New York.
- [33] Cohen, J., *A coefficient of agreement for nominal scales*. Educa. and Psycholog. Measurement, 1960. 20: p. 37-46.
- [34] Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data*. Biomet., 1977: p. 159-174.
- [35] Gamer, M., et al., *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84. 2012.
- [36] Xu, Y. *ProsodyPro - A tool for large-scale systematic prosody analysis*. Tools and Resources for the Analysis of Speech Prosody. 2013. Aix-en-Provence, France.
- [37] R Development Core Team, R., *R: A language and environment for statistical computing*. 2015, R Foundation for Statistical Computing: Vienna.
- [38] Baayen, R.H., *Analyzing Linguistic Data. A Practical Introduction to Statistics*. 2008, Cambridge: CUP.
- [39] Matuschek, H., et al., *Balancing type I error and power in linear mixed models*. JML, 2017. 94: p. 305-315.
- [40] Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen. *lmerTest: Tests for random and fixed effects for linear mixed effect models*. 2013
- [41] Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society, 1995. 57: p. 289-300.
- [42] Wood, S.N., *Generalized additive models: An introduction with R*. 2nd ed., 2017, Boca Raton [u.a.]: CRC press.
- [43] Wieling, M., *Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English*. Journal of Phonetics, 2018. 70: p. 86-116.
- [44] van Rij, J., et al., *itsadug: Interpreting time series and autocorrelated data using GAMMs*. 2017.
- [45] Wood, S.N., *Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models*. Journal of the Royal Statistical Society (Statistical Methodology), 2011. 73: p. 3-36.
- [46] Yang, Y., et al., *Clause type anticipation based on prosody in Mandarin*. Int. Journal of Chin. Linguistics 2019. 6: p. 1-26.
- [47] Yuan, J., *Intonation in Mandarin Chinese: acoustics, perception, and computational modeling*. 2004, Cornell University.
- [48] Kuang, J.J., *Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice*. JASA, 2017. 142: p. 1693-1706.