Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Prosodic realization of rhetorical and information-seeking questions in Persian

Mortaza Taheri-Ardali[1], *Tina Ghaemi*[2], *Tina Bögel*[3], *Bettina Braun*[3]

[1]Shahrekord University, Iran
[2]Technical University of Dortmund, Germany
[3]University of Konstanz, Germany

taheri@sku.ac.ir, tina.ghaemi@tu-dortmund.de, tina.boegel@uni-konstanz.de, bettina.braun@uni-konstanz.de

Abstract

This paper investigates the prosodic realization of rhetorical questions (RQs) in comparison to information-seeking questions (ISQs) in Persian, an Iranian language. In a lab setting, we recorded polar questions (word order: S-O-V) and wh-questions (word order: wh-O-V) in information-seeking and rhetorical contexts. We then analyzed constituent durations, voice quality (breathy, modal, glottalized), and the intonational realization. The results showed that all constituents, except the verb in wh-questions, were significantly lengthened in RQs compared to ISQs. Furthermore, RQs were more often realized with breathy voice quality than ISQs. Intonationally, RQs had lower f0-values than ISQs, in particular towards the end of the questions. This was caused by phonological differences: Polar questions, which were primarily realized with high or downstepped high boundary tones (H%, !H%), differed significantly in whether there was an accent on the verb (ISQ) or not (RQ). In wh-questions, in which the object and the verb were typically deaccented, differed in boundary tone: wh-RQs most often had L% boundary tones and wh-ISQs !H%. The results for Persian align with findings for other, typologically different, languages. Furthermore, they provide data that have not been discussed for Persian yet (e.g., !H% in wh-questions).

Index Terms: Rhetorical question, information-seeking question, intonation, voice quality, duration, Persian

1. Introduction

Rhetorical questions are an interesting figure of speech because they do not request information, but serve to make a point [1]. They are intended to commit the interlocutor to the information presupposed in the questions: For instance, the RQ *Who lifted a finger to help?* intends to make the point that no one did [2].

While there are linguistic cues to rhetorical questions (e.g., negative polarity items such as *lifted a finger*), the same string of words may be used to ask a genuine question or a rhetorical question (e.g., *Who knows syntax?*). Over the past years, there have been a number of studies that investigated the prosodic differences of such string-identical questions [3, 4, 5, 6]. A cross-linguistic comparison of prosodic differences between RQs and ISQs (German, Western Canadian English, Icelandic, Italian, Standard Chinese, Cantonese, Japanese, and French) identified major cross-linguistic phonological and phonetic cues that distinguish RQs from ISQs [7]. The generalization is that RQs differ reliably from ISQs in that RQs have longer duration, lower pitch excursion, and more cases of non-modal voice quality than ISQs. The present study investigated whether this pattern can be extended to Persian, a member of the Iranian languages hitherto not discussed in this context. These languages are interesting from a linguistic point of view because they have a different syntactic and prosodic structure than the languages discussed in [7].

Persian is an official/state language in the three Iranophone countries of Iran, Afghanistan, and Tajikistan, also referred to as *Fārsi*, *Dari*, and *Tojiki*, respectively [8]. Here, we investigated modern colloquial Persian or *Fārsi* as spoken in Tehran. Colloquial Persian demonstrates a flexible word order, diverging from the standard written Persian, which conventionally follows the SOV word order. Prosodically, Persian has been claimed to be a pitch accent language by [9], [10], and [11], who assume a lexically assigned pitch accent (L+)H* to signal word-level prominence. However, [12] showed that, in the absence of pitch, Persian signals word-level prominence through durational means as well, confirming [13]'s original classification of Persian as a stress accent language. Traditionally, word-level prominence in Persian is usually ascribed to the last syllable of the word, although some suffixes and, in particular, the verbal prefixes can change this prevalent pattern. To put it differently, Persian nouns, adjectives, and most adverbs exhibit prominence on the final syllable, while prominence on verbal elements is observed towards the beginning of the verbs [14, 15].

[16] used four pitch accents (H*, L*, L+H*, and L*+H), two phrase accents (L- and H-, Intermediate Phrases), and two boundary tones (L% and H%) for Persian. In more recent work, [17] assumed two higher prosodic units: the Accentual Phrase (AP, marked by l and h) and the Intonational Phrase (IP, marked by L% and H%). The AP usually consists of a content word carrying an (L+)H* pitch accent on the prominent syllable of the word, with L only surfacing in polysyllabic words, and ending with one of two AP tones (l or h). In contrast, [18] suggested that the most suitable representation for the pre-nuclear rising accent in Persian is a sequence devoid of a star, composed of a low and a high tone, specifically LH. In the current paper, we largely follow [17] for the configuration of Persian question intonation, where he suggests that the pitch accent (L+)H* is consistently found in Persian APs across all question types.

Research on Persian interrogatives has mainly focused on ISQs [17, 19, 20, 21, 22, 23, 24]. Polar questions are characterized by an H% boundary tone [17, 19, 24], while wh-questions have a falling intonation resembling declaratives, with the nuclear pitch accent on the wh-constituent [17, 19]. In a more recent study, [24] conducted a comparison of syntactically identical polar questions and declaratives. His results showed that pre-nuclear and nuclear peaks are higher and located earlier in questions compared to statements.

To enhance investigations into Persian interrogatives in general, and to enrich the research on RQs with a typologically different language, we conducted a production experiment following the protocol established in [3].

10.21437/SpeechProsody.2024-181

2. Experiment

2.1. Methods

2.1.1. Materials

The materials were translated from German [3] into Persian and were adapted to account for culture-specific differences. They consisted of 21 context-pairs that triggered either an ISQ (1a, 1b) or an RQ (1c) and 21 questions that appeared in two question types: as a polar question (2a) or as a wh-question (2b). The polar question contained an open element (*anyone*) to match the open wh-element in wh-questions. All contexts explicitly mentioned the proposition of the question (e.g., studying math) to keep information-structure identical.

(1) a. **polar-ISQ**: You meet some former schoolmates who were from your math class, and you want to know if one of them is studying math. You say to your former schoolmates:

b. **wh-ISQ**: You meet some former schoolmates from your math class and you want to know which of them is studying math. You say to your former schoolmates:

c. **(polar/wh)-RQ**: Your neighbour mistakenly thinks that you study math, although everyone knows that you are very bad at math. You say to your neighbour:

(2) a. **polar question**: *kasi* *riāzi* *mixune?*
anyone math study
'Does anyone study math?'

b. **wh-question**: *ki* *riāzi* *mixune?*
who math study
'Who studies math?'

The verbs were all trisyllabic with the word-level prominence on the initial syllable. The object nouns were monosyllabic (N = 2), disyllabic (N= 14), trisyllabic (N=5) or foursyllabic (N=1). In addition to the experimental items, we translated the 28 filler items and 6 familiarization items (each consisting of contexts and a target utterance). The fillers included questions with more syntactic constituents and exclamations.

The 21 questions were pretested to establish that each of them could be used as a potential ISQ or RQ. Twenty native speakers of Persian, different from the ones in the main experiment, indicated whether they agreed or disagreed with the proposition in the question, e.g., whether they studied math in the example above. The results confirmed that all of the target structures resulted in responses and could, therefore, be used as ISQ and RQ.

2.1.2. Participants

Twelve native speakers of Persian (4 males, 8 females) voluntarily participated in the online experiment (average age 22.6 years, range 18-25). All participants grew up monolingually in Iran and learned a second language only after the age of 6 years.

2.1.3. Procedure

The experimental material was divided into two basic lists, each containing one half of the polar questions in both illocution types (ISQ and RQ) and one half of the wh-questions in both illocution types (ISQ and RQ), cf. [3]. This resulted in 42 items per list. Illocution type was hence manipulated within-subjects and within-items. The fillers were interspersed with the experimental items so that two experimental items were separated by at least two filler items. The trials were pseudo-randomized so that information-seeking and rhetorical questions of an item were separated by at least six other items. Each list was presented in its original or reversed order. The six familiarization trials were put at the start of each list. Lists were randomly assigned to participants.

The experiment was programmed in SoSci Survey [25] and participants were tested online. Participants were instructed to do the experiment in a quiet room and to use headset microphones, if possible. Before starting the experiment, they gave written consent. Each trial started with a screen that showed one context, triggering either an RQ or an ISQ interpretation. The contexts were presented in Perso-Arabic script, black on white screen. The target utterances were presented in blue font. When participants finished reading the contexts, they pressed a button to record the stimulus. They stopped the recording with another button. The productions were recorded directly and transformed anonymously onto SoSci.

2.1.4. Analysis

The constituents (subject/wh-word, object, verb) were segmented manually by student assistants in Konstanz following standard segmentation criteria [26] in Praat [27]. We excluded 39 utterances that contained extra/wrong words or laughter: 34 RQs (24 polar, 10 wh) and 5 ISQs (3 polar, 2 wh). The most frequent reasons for exclusion were wrong or extra words or syllables (N = 19) or the change from the question word *kasi* to *ki* in RQs, which turns a polar question into a wh-question. For the remaining 423 utterances, constituent durations were extracted via a Praat script. The f0 over time was extracted with ProsodyPro [28] with two different f0 settings for male and female speakers (male: 75Hz to 350Hz, female: 110Hz to 500Hz). Ten f0 points were extracted per constituent and z-scored by participants to minimize speaker-specific differences.

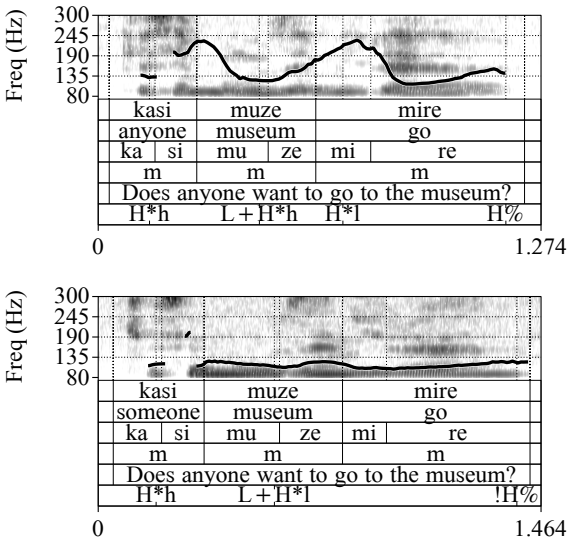

Figure 1: *Example polar ISQ-RQ pairs from the same speaker. The nuclear pitch accent is on the verb in the upper panel (ISQ), while it shifts to the object in the RQ (lower panel). Boundary tones differ and so does the pitch range.*

Intonation and voice quality were annotated perceptually. Two of the authors (BB, MTA) annotated one batch of files together (3 items from all speakers), both in terms of voice quality and intonation. The annotations were compared and differences dis-

cussed and resolved. Voice quality was annotated as modal, breathy, or glottalized (acoustic analyses were not possible due to the differences in recording devices). For the analysis of intonation, the accented words were labelled with either H* or L+H* followed by an AP tone (h or l). Narrow focus on a constituent resulted in post-focal deaccentuation. Boundary tones were L%, H% and !H% (a descriptive label for a boundary tone that was neither low nor high). In some cases, the peak of the initial constituent *ki* was delayed and realized on the first syllable of the object (which sounded very prominent), but which is supposedly unstressed/unaccented. There was then a drop to a low(er) tone, mostly in the object. In these cases, we decided to mark an "X" at the low elbow for descriptive purposes (Figure 2 bottom). The two authors labelled half of the data each and consulted each other in case there were deviant contours. Interrator agreement on one of the batches (n = 40) was 93%.

To corroborate the analyses, the most frequent contours with the same phonological patterns (N >20) were stored into one folder and the authors checked whether the contours in one folder sounded the same and consulted each other in case they detected "outliers". These outliers were labelled anew. The results reported in Section 2.2.3 below are based on this final analysis. Example contours are shown in Figures 1 and 2.

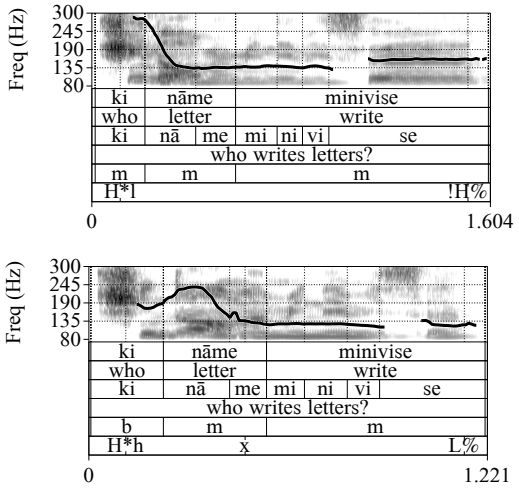

Figure 2: *Example wh-ISQ-RQ pair from the same speaker. In the upper panel, the nuclear pitch accent is on the wh-word and concludes with !H%. In the lower panel, the nuclear pitch accent advances and is placed on the initial syllable of the object and finally ends in L%.*

2.2. Results

2.2.1. Duration

The average constituent durations are shown in Figure 3. We calculated linear mixed-effects with question type (polar, wh), illocution type (ISQ, RQ), and position as fixed factors and participants and items as crossed random factors [29]. Random slopes for within-group factors were added if this improved the fit of the model. To estimate the degrees of freedom (and arrive at p-values), we used the Satterthwaite approximation implemented in the R-library lmerTest [30].

Results showed a significant effect of illocution type (β = 68.60, SE = 12.54, df = 1257.50 , t 5.47, p <0.0001). The average lengthening for RQs was 86ms (15.4% relative to ISQs). Furthermore there was a three-way interaction between illocution type, question type and position (p <0.0001): The verb in wh-questions was the only constituent that was not lengthened.

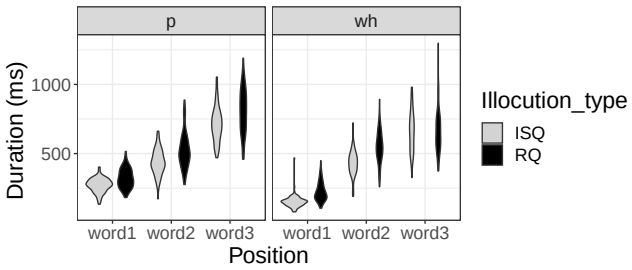

Figure 3: *Constituent duration.*

2.2.2. Voice quality

Figure 4 shows the distribution of breathy voice. For statistical analysis, we coded voice quality as breathy (1) or non-breathy (0) and analyzed the data using mixed-effects logistic regression models (glmer) with the same modeling as described above.

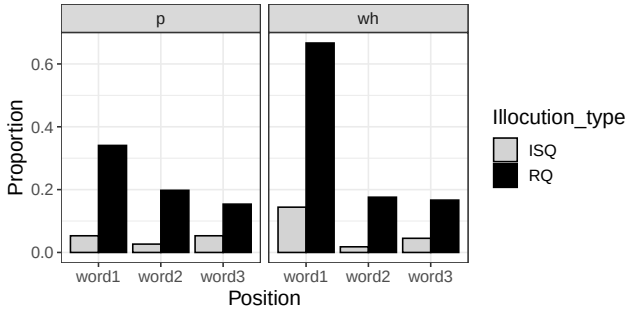

Figure 4: *Distribution of breathy voice quality.*

Results showed effects of position (more breathy voice in word 1 than in words 2/3), question type (more breathy voice in wh-questions), and illocution type (more breathy voice in RQs), all p <0.0005. The model with a three-way interaction did not converge; there were no two-way interactions (all p > 0.3).

2.2.3. Intonation

Figure 5 shows the z-scored averaged f0-contours. There is a perceptual rise on the first word in all conditions. Given the monosyllabic nature of the wh-word *ki*, this rise is phonetically not always visible. There are clear differences in f0 between ISQ and RQs, which grow stronger towards the end of the utterance.

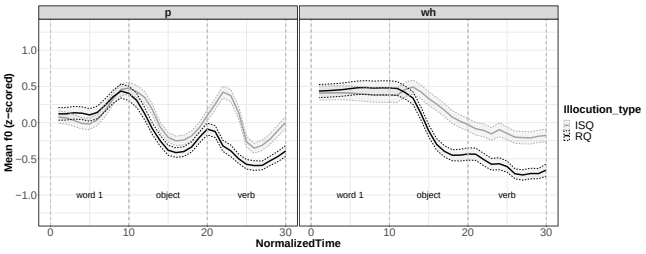

Figure 5: *Z-scored mean f0 over time as extracted by ProsodyPro. The band indicates +/- 1 standard error.*

Regarding the phonological structure, we focus first on the most frequent contours (>10%) and their distribution across question type and illocution type and then analysed the whole data set. The most frequent contours are shown in Table 1. For **polar questions**, ISQ were mostly realized with an H*h on *anyone*, a rising accent on the object noun (L+H*h) and a fall-rise on the verb (H*l H%), see line 1 of Table 1. This contour was never attested in RQs. The second most frequent contour was similar except for a downstepped high boundary tone (!H%), a contour that occurred also in RQs (but less frequently, line 2 in Table 1). The most frequent contours in RQs had a deaccented verb, ending in !H% or H% (lines 3 and 4). This suggests that – for polar questions – illocution type is distinguished by the presence or absence of an accent on the verb (mostly accented verb for ISQs, mostly deaccented verb for RQs).

For **wh-questions**, the most frequent ISQ tune had H*h on the wh-word and a shallow decline to a !H%; the object and the verb were both unaccented (line 1). The second most frequent contour in ISQ was the contour with the delayed peak in the first syllable(s) of the object noun (that we descriptively labelled as X, line 2). This contour occurred with even higher frequency in RQs. The second very frequent contour in RQs had an H*l on the first constituent, followed by a low boundary tone (L%, line 4), the same boundary tone as in the last line. Wh-questions hence seem to distinguish illocution type by boundary tone (mostly !H% for ISQs, mostly L% for RQs).

Table 1: *Distribution of frequent tunes in %.*

type	word 1	object	verb	bt	ISQ	RQ
polar	H*h	L+H*h	H*l	H%	38	0
	H*h	L+H*h	H*l	!H%	35	19
	H*h	L+H*l	none	!H%	0	26
	H*h	L+H*l	none	H%	0	19
wh	H*h	none	none	!H%	32	0
	H*h	X	none	L%	22	31
	H*l	none	none	!H%	15	0
	H*l	none	none	L%	0	28
	H*h	L+H*l	none	L%	0	11

To corroborate the generalizations derived from the most frequent contours, we calculated separate glmers for all polar and wh-questions. Specifically, we investigated whether illocution type significantly influences the presence/absence of an accent on the verb and boundary tone (for polar questions) and the kind of boundary tone (for wh-questions). The glmers were fitted as described above for voice quality. For polar questions, illocution type had a significant effect on the presence of the accent of the verb (β = 3.6, SE = 0.5051, z= -7.1, p <0.0001) but no effect on boundary tone (p >0.1) For wh-questions, the effect of illocution type on final boundary tone (low vs. non-low) was also significant (β=2.9, SE= 0.5, z= 6.4, p <0.0001).

3. Discussion and conclusion

We tested the prosody of Persian polar and wh questions in two illocution types: rhetorical and information-seeking. In summary, all constituents except the verb in wh-questions were significantly longer in RQs than in ISQs. RQs were more often realized with breathy voice than ISQs; also, there was a strong effect of position: The first word had a higher proportion of breathy voice than the second or final words. Regarding intonation, the average f0 was phonetically lower in RQs than in ISQs, in particular toward the end of the utterance. Phonologically,

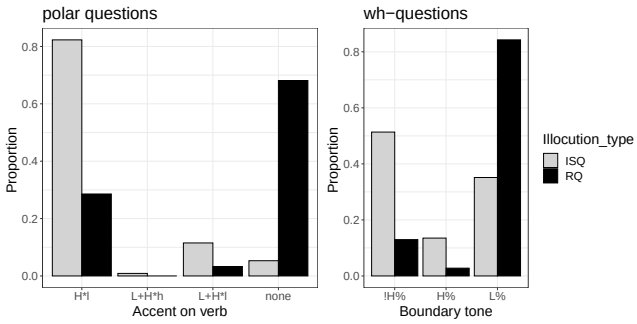

Figure 6: *Distribution of accents on the verb for polar questions (left panel) and boundary tones for wh-questions (right panel).*

this was achieved by deaccenting the verb in polar questions and by a difference in final boundary tone for wh-questions: wh-ISQs mostly ended in !H%, wh-RQs mostly in L%. The results for boundary tones in wh-questions were unexpected for two reasons: First, the !H% boundary tone, in which most wh-ISQs conclude with, have not been previously reported in the Persian literature. Conversely, the L% boundary tone, the most frequent boundary tone in RQs in our wh-questions, has previously been noted for wh-ISQs. It should be highlighted that when the contour ends in L% in wh-ISQ, the peak is remarkably delayed in the first word and peaks on the first or second syllable of the object. Future research will investigate longer wh-phrases (e.g. *who of you* instead of *who*) to determine the generalisability of the findings (e.g. whether the monosyllabic status of the first word contributes to the peak delay). Also, we plan to test the perceptual relevance of the identified cues (cf. [31]).

Note that speakers more frequently added extra words (particles) in RQs than in ISQs and changed polar to wh-questions. This finding resembles the behaviour of German participants in rhetorical contexts in a fragment completion task [32].

Cross-linguistically, the prosodic differences between ISQs and RQs in Persian resemble those in Germanic and Romance languages, as well as Mandarin Chinese. The common cues are longer durations (with the uncommon exception of the final verb in wh-questions) and more frequent non-modal voice quality. Similar to Chinese, Persian RQs are characterized by a lower f0, particularly towards the end of the utterance. In contrast to intonation languages, Persian exhibits a smaller set of contours, with certain contours prevalent in one illocution type and nonexistent in the other. This suggests a highly specific mapping from tonal realization to pragmatic function. Finally, our phonological analysis of intonational realization has revealed that the documented phonetic differences in f0, as e.g., already presented in [33], indicating lower f0 in RQs towards the end of utterances, are phonologically governed (deaccentuation of the verb in polar questions and variations in boundary tones in wh questions). This sets Persian apart from Mandarin Chinese, where tonal differences between ISQs and RQs appeared otherwise similar.

4. Acknowledgements

The research was funded by the DFG as part of the research unit "Questions at the Interfaces" (FOR 2111, BR 3428/4), the Ontario-Baden-Württemberg faculty mobility program, and Carleton University, Ottawa, Ca. (visiting professorship, MTA).

5. References

[1] M. Biezma and K. Rawlins, "Rhetorical questions: Severing asking from questioning," in *Proceedings of SALT 27 (2017)*, 2017, pp. 302–322.

[2] C.-H. Han, "Interpreting interrogatives as rhetorical questions," *Lingua*, vol. 112, no. 3, pp. 201–229, 2002.

[3] B. Braun, N. Dehé, J. Neitsch, D. Wochner, and K. Zahner, "The prosody of rhetorical and information-seeking questions in German," *Language and Speech*, vol. 62, no. 4, pp. 779–807, 2019.

[4] K. Zahner-Ritter, Y. Chen, N. Dehé, and B. Braun, "The prosodic marking of rhetorical questions in Standard Chinese," *Journal of Phonetics*, vol. 95, pp. 101–119, 2022.

[5] N. Dehé and B. Braun, "The intonation of information-seeking and rhetorical questions in Icelandic," *Journal of Germanic Linguistics*, vol. 32, no. 1, pp. 1–42, 2020.

[6] ——, "The prosody of rhetorical questions in English," *English Language and Linguistics*, vol. 24, no. 4, pp. 607–635, 2020.

[7] N. Dehé, B. Braun, M. Einfeldt, D. Wochner, and K. Zahner-Ritter, "The prosody of rhetorical questions: a cross linguistic view," *Linguistische Berichte*, vol. 269, pp. 3–42, 2022.

[8] L. Paul, "Persian," in *The Languages and Linguistics of Western Asia: An Areal Perspective*, G. Haig and G. Khan, Eds. Berlin: De Gruyter Mouton, 2018, pp. 569–624.

[9] V. Abolhasanizadeh, M. Bijankhan, and C. Gussenhoven, "The Persian pitch accent and its retention after the focus," *Lingua*, vol. 122, pp. 1380–1394, 2012.

[10] H. Rahmani, T. Rietveld, and C. Gussenhoven, "Stress 'deafness' reveals absence of lexical marking of stress or tone in the adult grammar," *PloS ONE*, vol. 10, no. 12, pp. 1–16, 2015.

[11] ——, "Post-focal and factive deaccentuation in Persian," *Glossa: A Journal of General Linguistics*, vol. 3, no. 1, pp. 1–36, 2018.

[12] V. Sadeghi, "Word-level prominence in Persian: An experimental study," *Language and Speech*, vol. 60, no. 4, pp. 571–596, 2017.

[13] S.-A. Jun, "Prosodic typology," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. New York: Oxford University Press, 2005, pp. 430–458.

[14] T. Vahidian-Kamyar, "Suprasegmental features in Persian: Stress, intonation and a description of pause," Ph.D. dissertation, University of Tehran, Tehran, 1972.

[15] G. Lazard, *A grammar of contemporary Persian (English translation)*. Mazda Publishers, 1992.

[16] M. Eslami, "The prosody of the Persian language and its application in computer-aided speech recognition," Ph.D. dissertation, University of Tehran, Tehran, 2007.

[17] N. Sadat-Tehrani, "The intonational grammar of Persian," Ph.D. dissertation, University of Manitoba, Winnipeg, 2007.

[18] V. Sadeghi, "The timing of pre-nuclear pitch accents in Persian," *Journal of the International Phonetic Association*, vol. 49, no. 3, pp. 305–329, 2019.

[19] N. Sadat-Tehrani, "The intonation patterns of interrogatives in Persian," *Linguistic Discovery*, vol. 9, no. 1, pp. 105–136, 2011.

[20] Z. Shiamizadeh, J. Caspers, and N. O. Schiller, "The role of prosody in the identification of Persian sentence types: Declarative or wh-questions," *Linguistics Vanguard*, vol. 3, no. 1, pp. 1–11, 2017.

[21] ——, "The role of f0 and duration in the identification of wh-in-situ questions in Persian," *Speech Communication*, vol. 93, pp. 11–19, 2017.

[22] ——, "When is a wh-in-situ question identified in Persian?" *Language, Cognition and Neuroscience*, vol. 33, no. 9, pp. 1168–1183, 2018.

[23] ——, "Do Persian native speakers prosodically mark wh-in-situ questions?" *Language and Speech*, vol. 62, no. 2, pp. 229–249, 2019.

[24] V. Sadeghi, "The intonation of Persian polar questions," *Speech Communication*, vol. 145, pp. 47–70, 2022.

[25] D. J. Leiner, "SoSci survey (version 3.1. 06) [computer software]," 2019.

[26] A. Turk, S. Nakai, and M. Sugahara, *Acoustic segment durations in prosodic research: a practical guide*, ser. Language, Context and Cognition. Mouton de Gruyter, Jul. 2006, pp. 1–28.

[27] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]," 2023.

[28] Y. Xu, "ProsodyPro — a tool for large-scale systematic prosody analysis," in *Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, 2013.

[29] R. Baayen, D. Davidson, and D. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008, special issue: Emerging Data Analysis.

[30] A. Kuznetsova, P. Brockhoff, and R. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[31] M. Kharaman, M. Xu, C. Eulitz, and B. Braun, "The Processing of Prosodic Cues to Rhetorical Question Interpretation: Psycholinguistic and Neurolinguistics Evidence," in *Proceedings of Interspeech*, 2019, pp. 1218–1222.

[32] N. Dehé, D. Wochner, and M. Einfeldt, "The interaction of discourse markers and prosody in rhetorical questions in German," *Journal of Linguistics*, vol. 60, no. 1, pp. 103–127, 2022.

[33] B. Braun and T. Bögel, "Rhetorical questions in Persian," in *Phonetics and Phonology in German-speaking Countries*, poster presentation, Bielefeld, 2022.