Wh-question or wh-declarative? Prosody makes the difference

Yang Yang\textsuperscript{a,b,*}, Stella Gryllia\textsuperscript{b}, Lisa Lai-Shen Cheng\textsuperscript{b}

\textsuperscript{a} Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, China
\textsuperscript{b} Leiden University Centre for Linguistics, The Netherlands

\section*{A R T I C L E   I N F O}

\textbf{Keywords:}
Wh-question
Wh-declarative
Prosody
Clause type
Ambiguity

\section*{A B S T R A C T}

Mandarin wh-words can have question or non-question (e.g., existential, universal quantificational) interpretations. Their interpretations in a sentence are usually not ambiguous, as the distinct interpretations need to be licensed by particular items/contexts. The starting point of our study concerns a case which allows the wh-words to remain ambiguous in a sentence: wh-words such as \textit{shénme} appearing with \textit{dùō}. After empirically confirming that such sentences are indeed ambiguous (Study 1), we turn to the question of whether and how prosody helps disambiguate such sentences. Our production experiment (Study 2) shows that wh-declaratives differ from wh-questions in terms of prosodic properties already from the clause onset. Wh-declaratives are longer than wh-questions starting from the subject and the pattern reverses at the wh-word; wh-declaratives are lower in F0 and smaller in F0 range than wh-questions at the wh-word and there is a F0 range compression in the post-wh-word region in wh-questions; wh-declaratives show larger intensity range than wh-questions at the verb and the pattern reverses at the wh-word. An implication of this study concerns the focal status of wh-words in wh-questions and wh-declaratives: wh-words are foci in wh-questions but cannot be foci in wh-declaratives.

\section*{1. Introduction}

\subsection*{1.1. Wh-words in Mandarin Chinese}

Mandarin Chinese is a wh-in-situ language (Huang, 1982), in which question words (i.e., wh-words) remain at their base position, as illustrated in (1). (1a) is an ordinary information-seeking question, where the wh-word \textit{shénme} ‘what’ corresponds to the direct object of the sentence. The wh-word \textit{shénme} occupies the same position as the direct object \textit{shū} ‘book’ in the declarative counterpart in (1b).

\begin{enumerate}
  \item [(1a).] Zhang Sān mài-le shénme?
  \begin{tabular}{llll}
  Zhang San & buy-PERF & what
  \end{tabular}
  \begin{itemize}
  \item [\textit{What did Zhang San buy}?] [wh-question]
  \end{itemize}
  \item [(1b).] Zhang Sān mài-le shū.
  \begin{tabular}{llll}
  Zhang San & buy-PERF & book
  \end{tabular}
  \begin{itemize}
  \item [\textit{Zhang San bought a book}.] [declarative]
  \end{itemize}
\end{enumerate}

Besides the interrogative interpretation, wh-words in Mandarin (as in Japanese and Korean) can receive a universal (‘everything’) or existential (‘something’) interpretation, subject to different licensors/quantifiers, as shown in (2a) and (2b). In (2a), the universal interpretation is licensed by the maximality operator \textit{dùō} (see Giannakidou and Cheng, 2006; Cheng, 2009 for the analysis of \textit{dùō} as a maximality operator), while in (2b), the existential interpretation is triggered by the yes-no question particle \textit{ma}.

\begin{enumerate}
  \item [(2a).] Zhang Sān shénme dōu mài.
  \begin{tabular}{llll}
  Zhang San & SHENME & all & buy
  \end{tabular}
  \begin{itemize}
  \item [\textit{Zhang San buys everything}.] [universal]
  \end{itemize}
  \item [(2b).] Zhang Sān mài-le shénme ma?
  \begin{tabular}{llll}
  Zhang San & buy-PERF & SHENME & ma?
  \end{tabular}
  \begin{itemize}
  \item [\textit{Did Zhang San buy anything}?] [existential]
  \end{itemize}
\end{enumerate}

Although the wh-words can have different interpretations, the interpretation of wh-words in Mandarin in each sentence is in general unambiguous, as the different interpretations are connected to different licensors. In the absence of any overt licensors, a wh-word like \textit{shénme} is typically interpreted as an interrogative word (‘what’) and the sentence is a wh-question, as (1a). In the case most relevant for the current study, i.e., the existential interpretation of wh-words (e.g., ‘something’ for \textit{shénme}), aside from the yes-no question environment (2b), the existential interpretation of the wh-word (‘something’) has been shown to be licensed by sentences containing negation, conditionals, and epistemic modalities. These sentences are referred to as ‘nonveridical sen-

\begin{itemize}
  \item [\textsuperscript{1}] From here on, we gloss \textit{shénme} simply as \textit{SHENME}, as it can have interrogative and non-interrogative interpretations.
  \item [\textsuperscript{2}] We gloss \textit{ma} as yes-no particle, the particle for yes-no questions.
\end{itemize}
tences/contexts’. That is, in these sentences/contexts, the truth of a proposition (e.g. ‘Zhang San bought things’) cannot be entailed, as in (2b) (see Li, 1992; Lin, 1998, 2014; Huang, 2018 for details).

1.2. An ambiguous case in interpreting wh-words

Though most cases of wh-words are unambiguous, there are a few instances where the wh-word is in fact ambiguous between a declarative and question interpretation. Consider the sentence in (3), which contains an adverb zuōtān ‘yesterday’ and the perfective marker le. This is a typical verticalic sentence, in which the truth of the proposition (e.g. ‘Zhang San bought things’) is entailed. Following previous analyses of existential interpretation of wh-words (as mentioned above), where the existential interpretation is only licensed in nonveridical contexts, we expect the wh-word in (3) to only have an interrogative interpretation (‘what’) as in (3i).

(3) 赤|gān zuō|tān mài-le diān|fr 3 shénme
gān zuō tān mài le diān 3 shénme

1. ‘What did Zhang San buy a little of yesterday?’ (wh-question)
2. ‘Zhang San bought a little of something yesterday.’ (wh-declarative)

Nonetheless, our observation in (3) runs counter to the existing literature. In particular, native speaker informants consider the wh-word in (3) to also have an existential reading (i.e., ‘something’) as in (3ii). This means that the sentence in (3) is in fact ambiguous; it can be interpreted as a wh-question or as a declarative. In this paper we use the term wh-declarative to refer to a declarative sentence which contains a wh-word. To determine whether (3) is indeed ambiguous between a wh-question interpretation and a wh-declarative interpretation, a solid empirical study is needed, which we conducted, and is reported in Section 2.

1.3. The role of prosody in distinguishing questions and declaratives

With ambiguous sentences like (3), the question also arises whether and how the two interpretations or clause types (wh-questions and wh-declaratives) are disambiguated by speakers.

It is known that context can play a role in resolving ambiguities or differentiating interpretations or clause types (Spivey-Knowlton et al., 1993). Nonetheless, when an utterance such as (3) is used out of context, how can the wh-question interpretation be distinguished from the wh-declarative interpretation? Previous studies have shown that prosody can function as the defining feature of questions in the absence of an overt syntactic interrogativity marker (e.g. Bolinger, 1978; Ohala, 1983, 1984). In other words, when declaratives are string identical to questions, prosody plays a critical role in marking the different clause types (Jun and Oh, 1996; Frota, 2002; Face, 2004; Vion and Colas, 2006; Baltazani, 2007, among others). In Mandarin, prosodic markings are often investigated from either a global perspective like the sentence F0 curve/contour or local prosodic features like duration, F0 (range) or intensity (range) on the syllable or word level or a combination of both (Shi, 1980; Shen, 1994; Yuan, 2004; Liu, 2009; Jiang and Chen, 2011). In Mandarin, polar questions can also be string identical to declaratives, and the former are marked with a higher sentence F0 curve as compared with their declarative counterparts (Shi, 1980; Shen, 1990, 1994; Yuan, 2004, 2006; Jiang and Chen, 2011, among others) and the biggest F0 difference between the two clause types often lies in the final syllable (Yuan, 2004, 2006).

The above studies on the role of prosody in marking different clause types are centered on polar questions. However, it remains unclear whether and how wh-declaratives and wh-questions as in (3) are prosodically marked and distinguished. Studies have shown that Korean (also a wh-in-situ language with wh-words with similar properties as Mandarin) utilizes pitch accent and prosodic phrasing to differentiate wh-questions and declaratives containing wh-words: normally, a wh-interrogative bears a high-pitch accent and a wh-existential bears a low-pitch accent; wh-questions are characterized by a post-wh-word de-phrasing, namely, a deletion of accentual phrasings following the wh-word (Jun and Oh, 1996; Shen, 2005; Yun, 2012). In Mandarin, so far, only two studies investigate the prosody of the two relevant clause types (Dong, 2009; Liu et al., 2016). Below we review them one by one.

Dong (2009) compares wh-questions and wh-declaratives in terms of F0, although not on string identical cases. As illustrated in (4), the wh-word shéi is interpreted as a question word ‘who’ in (4a) and as an existential in (4b) meaning ‘someone’. Dong’s results show that wh-word is the most prosodically prominent item in wh-questions as shown by its expanded pitch range, while in contrast it has a compressed pitch contour in wh-declaratives; instead, the verb bears the prosodic prominence with expanded pitch range in wh-declaratives.

(4) a. Méi yěyǔ-le
gān fù|lèi-ming diān|fr 3 shénme
gān jī|fēng zhī
gān jī|fēng zhī

Mei ridicule-PERF
Who did Mei ridicule?’
Who did Mei ridicule?’

b. Hāoxiāng Méi yěyǔ-le
gān fù|lèi-ming diān|fr 3 shénme
gān jī|fēng zhī
gān jī|fēng zhī

Seem Mei ridicule- PERF someone
Shéi
diān 3

‘It seems that Mei ridiculed someone.’

Dong’s study focuses on F0. Nonetheless, duration (and intensity) can also be informative prosodic cues to investigate (Shen, 1993; Chuang and Fon, 2010).

Liu et al. (2016) investigate the prosodic marking of wh-questions with the string identical wh-declaratives. They compare the two clause types from the perspective of both local prosodic features (prosodic properties of each word) and global prosodic features (sentence F0 curve). The examples (5a-b) illustrate the comparisons between the two clause types.

(5) a. 脩|éng sān dā|shūn chī diān 3 shénme gāo?
gān fù|lèi-ming diān|fr 3 shénme
gān jī|fēng zhī
gān jī|fēng zhī

Dong San intend.to eat a little SHENME cake
‘What kind of cake does Zhang San intend to eat?’

b. 脩|éng sān dā|shūn chī diān shénme gāo.
Zhang San intend.to eat a little SHENME cake
‘Zhang San intends to eat whatever cake.’

Similar to Dong (2009), Liu et al. found that the wh-word in wh-questions bears extended pitch range and higher pitch register than its counterpart in wh-declaratives. However, unlike Dong (2009), Liu et al. do not find any prosodic differences at the verb between the two clause types in (5), and the prominent word in both clause types is mostly at the sentence final position (e.g. gāo). Furthermore, it should be noted that, in Liu et al.’s study, the wh-word shénme is a modifier of the object noun ‘cake’, as shown in (5), while Dong uses the wh-word as the object, as shown in (4). In addition, Liu et al.’s results are based on a pilot study with a small sample of speakers (8 speakers). Hence Liu et al.’s results should be considered cautiously.

Based on these very limited studies, we can only conclude that wh-words have higher pitch and expanded pitch range when used as question words as compared with their existential counterparts. A knowledge gap remains as we cannot draw a clear picture with respect to how prosody can distinguish the two clause types, namely wh-questions and wh-declaratives. In particular, at which point in the sentence does a wh-
declarative begin to differ from a wh-question in terms of prosody, and in which prosodic properties (F0, duration or intensity) do they differ?

1.4. Research questions

In this paper, we scrutinize the prosody of wh-declaratives and wh-questions, addressing the following research questions:

R(1). In sentences containing wh-words, are the two interpretations (wh-declarative, wh-question) both available?
R(2). Is there a prosodic distinction between wh-questions from wh-declaratives? If yes, how do they differ?

The paper is organized as follows. In Section 2, we report the results of a reading study that investigated the availability of both a declarative and a question interpretation in sentences like (3), answering research question R(1). Section 3 presents the results of a production experiment on wh-questions and wh-declaratives, answering research question R(2). Section 4 concludes and examines the implications of the results found in the reading study and the production study.

2. A reading study on wh-questions and wh-declaratives

We have reported above that the veridical sentence in (3) containing a wh-word preceded by the quantifier diān (‘a little’), is observed to be ambiguous between a question interpretation and a declarative interpretation (wh-word shénme interpreted as an existential ‘something’), which challenges previous analyses where the existential interpretation is licensed in nonveridical sentences only. To investigate whether our observation is empirically supported, we conducted a reading study where participants were asked to read silently wh-sentences containing wh-words preceded by diān (‘a little’) as in example (3), and to finish them using a question mark or a full stop. Since no prosodic cues or previous contexts are given, we predict that participants will identify the wh-sentences as both questions and declaratives, due to the ambiguous status of the wh-word plus diān.

2.1. Participants

Eighty-four native speakers of Beijing Mandarin (53 females and 31 males, T age = 20 years old, SD = 2.3) were paid to participate in the reading experiment. All of them came from the northern part of Mainland China and at the time of testing were students at Tsinghua University. None of them reported speech problems or visual impairment. Prior to testing, informed written consent was obtained from each participant.

2.2. Experimental materials

We created a total of 40 stimuli. Half of the stimuli were ambiguous between an interrogative and a declarative interpretation (target stimuli), see (6a). The other half of the stimuli were unambiguous; they could be interpreted only as interrogatives (control stimuli), see (6b). As shown in (6), the target and control stimuli were string identical except for diān ‘a little’, which was present in the target stimuli but absent in the control. To construct the stimuli we used the basic word order in Mandarin (Li, 1990): Subject (proper name, e.g. Féng Tāo), Adverb (e.g. zìtītān ‘yesterday’), Verb (e.g. zú ‘make’) + Perfective marker (le), (diān), Direct Object (shénme ‘what/something’), and Prepositional Phrase (e.g. gěi Wáng Ying ‘to/for Wang Ying’). For the subject and the indirect object we used the commonly used disyllabic Chinese proper names. The length (the number of syllables) of all constituents was kept constant across stimuli.


The target and control stimuli were intermingled with 120 fillers. The fillers consisted of 20 polar questions with ma (i.e., yes-no question particle), 20 A-not-A questions and 80 declaratives. All stimuli lacked punctuation and were randomized for each participant.

2.3. Procedure

Participants were tested individually in a quiet room in Tsinghua University in Beijing, and the whole experiment lasted about 15 min. The participants’ were instructed to first read the stimulus on screen silently, and then to complete the sentence choosing a question mark or a full stop. The experiment was run on E-prime 2.0 (Psychology Software Tools) and was semi-self-paced. The procedure was as follows. A stimulus appeared on the computer screen and participants had a maximum of 5 s to read it silently. If participants read the stimulus faster than 5 s, they could press the space bar and move to the next screen to indicate their response. If participants exceeded the 5 s time limit, the next screen appeared automatically. This next screen showed the two punctuation marks: a question mark (?) and a full stop (.) and participants had again 5 s to indicate their response. If participants exceeded the time limit of 5 s, the next stimulus appeared automatically on screen. The question mark and the full stop were counter-balanced on screen to avoid any left or right preference by participants.

2.4. Statistical analysis and results

We obtained a total of 3350 responses (1673 responses for target stimuli, 1677 responses for control stimuli; there were 7 missing responses for target stimuli and 3 missing responses for control stimuli). We also extracted participants’ stimuli reading time and response time respectively.

2.4.1. Responses

Participants interpreted the control stimuli (wh-sentences without diān) as questions 93.1% of the time, while they interpreted the target stimuli (wh-sentences with diān) as questions 59.7% of the time, and as declaratives 40.3% of the time, see Fig. 1. We found a significant association between the stimulus type (target or control) and the participants’ responses, \(x^2 = 518.91 (1), p < 0.001\).

Moreover, we ran a mixed effects logistic regression using the glmer command in the lme4 package (Bates et al., 2015) in R (Core Team, 2017) to investigate whether participants’ responses can be predicted on the basis of the stimulus type. Specifically, we first ran a null model with participants’ responses as a dependent variable, and participants and items as random factors. We then ran a model that included in addition the stimulus type as a fixed effect factor to see whether the model was improved. Finally, on the basis of the first two models that included participants’ responses as a dependent variable, the stimulus type as a fixed-effect factor, and participants and items as random factors, allowing by-participant and by-item random intercepts, we ran another three models that in addition allowed by-participant or/and by-item random slopes for the stimulus type. Model fit was compared based on the likelihood ratio test (Pinheiro and Bates, 2000; Bolker et al., 2009) and the model which included by-participant and by-item random intercepts,

6 A-not-A question is a type of polar question in Mandarin offering a choice between an affirmative sentence and its negative counterpart.
and by-participant random slopes for the stimulus type was found to perform best \( \chi^2 = 69.481, df = 2, p < 0.001 \). The results showed that the stimulus type (target or control) is a reliable predictor for participants’ response, see Table 1.

### 2.4.2. Reading and response time

Participants spent more time reading the target stimuli \( \bar{\tau} = 2364.5 \text{ ms} \) as well as indicating their responses \( \bar{\tau} = 559.7 \text{ ms} \) than the control stimuli \( \bar{\tau} = 2024.1 \text{ ms} \); response time: \( \bar{\tau} = 445.7 \text{ ms} \), see Figs. 2 and 3.

To investigate the effect of stimulus type (target or control) on participants’ reading and response time, we ran a series of linear mixed-effects models using the lmerTest package \( \text{Kuznetsova et al., 2013} \) in R \( \text{R Core Team, 2017} \). For the reading time we first ran a null model with participants’ stimuli reading time as a dependent variable, and participants and items as random factors. We then ran a model that included in addition the stimulus type as a fixed effect factor to examine whether the model was improved. Finally, on the basis of the first two models that included participants’ stimuli reading time as a dependent variable, the stimulus type as a fixed-effect factor, and participants and items as random factors, allowing by-participant and by-item random intercepts, we ran another three models that in addition allowed by-participant or/and by-item random slopes for the stimulus type. For the response time, we followed a similar procedure to the stimuli reading time. Model fit was compared based on the likelihood ratio test \( \text{Pinheiro and Bates, 2000; Bolker et al., 2009} \) and for both the stimuli reading time and the response time, the model which included by-participant and by-item random intercepts, and by-participant random slopes for the stimulus type was found to perform best \( \chi^2 = 37.384, df = 2, p < 0.001 \); for response time, \( \chi^2 = 35.825, df = 2, p < 0.001 \).

The results showed that participants spent significantly longer time in reading the target stimuli that contained \( \text{diänr} \) than the control stimuli.

---

**Table 1**

Summary of the results of the mixed effects logistic regression between participants’ responses and the stimulus type.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.261</td>
<td>0.286</td>
<td>-7.916</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Participants’ responses in percentage (%) in the reading study.

**Fig. 2.** Participants’ stimuli reading time with error bars showing standard errors.
without diǎnr. Participants spent also significantly longer time in indicating the clause type for the target stimuli than the control stimuli, see Table 2.

To summarize, the results show that wh-sentences with diǎnr like (6a), are ambiguous between a declarative and a question interpretation in the absence of punctuation, context or prosody, as supported by participants’ responses in choosing the punctuations. The longer reading time and response time in the target sentences than the control sentences offer further support for the claim that sentences containing diǎnr followed by a wh-word are ambiguous between wh-questions and wh-declaratives (see Ferreira and Clifton, 1986, among others, for the discussion of a longer processing time on ambiguous cases than unambiguous cases).

After establishing that wh-sentences with diǎnr are indeed ambiguous in terms of clause types, we conducted a production experiment to investigate the research question (R2) of whether and how prosody is employed to distinguish the two clause types.

3. Production experiment

3.1. Method

3.1.1. Participants

Forty native speakers of Beijing Mandarin (23 females and 17 males, \( \bar{\text{age}} = 21 \) years old, SD = 2.53) who were born and raised in Beijing were paid to participate in the experiment. None of these participants have participated in the reading study or any other similar studies. At the time of recording these participants were students at Tsinghua University. None of them reported speech problems or visual impairment. Prior to recording, informed written consent was obtained from each participant.

Table 2
Summary of the results of the linear mixed-effects models on participants’ stimuli reading time and response time.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimuli reading time</td>
<td>−341.093</td>
<td>41.654</td>
<td>−8.189</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Response time</td>
<td>−114.384</td>
<td>18.234</td>
<td>−6.273</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

3.1.2. Materials

We created a total of 56 stimuli; half of the stimuli were wh-declaratives (see example in 7a), while the other half were wh-questions (see example in 7b). Wh-declaratives were string identical to their corresponding wh-questions except for the punctuation at the end of the sentence. As shown in (7), we used the word order which is essentially the same as the reading study: Subject (proper name, e.g. Tāo Wēi), Adverb (e.g. zhuòtiān ‘yesterday’), Verb (e.g. ná ‘bring’) + Perfective marker (le), (diǎnr), Direct Object (shènme ‘what/something’), and Prepositional Phrase (e.g. gěi Líu Gāng ‘to/for Liu Gang’). Each stimulus was 12 syllables long and the stimulus length was kept constant across clause types and items. We also kept constant the combination of tones across clause types and items for all the constituents, but the verb. For the verb, we included all four lexical tones, to obtain enough natural stimuli for the experiment. An example of a stimulus set is given in (7).

[wh-declarative]
(7) a. Tāo Wēi zhuòtiān ná-le diǎnr shènme gěi Líu Gāng.
T2 T1 T2 T1 T2-T0 T3 T2 T0 T3 T2 T1
Tao Wei yesterday bring-perf a little something to Liu Gang

‘Tao Wei brought a little something for Liu Gang yesterday.’

[wh-question]
b. Tāo Wēi zhuòtiān ná-le diǎnr shènme gěi Líu Gāng?
T2 T1 T2 T1 T2-T0 T3 T2 T0 T3 T2 T1
Tao Wei yesterday bring-perf a little what for Liu Gang

‘What did Tao Wei bring (a little) for Liu Gang yesterday?’

3.1.3. Procedure

The recordings took place in a sound-proof booth in a lab of the Department of Foreign Languages and Literatures at Tsinghua University in Beijing. For the recordings, we used a head-worn unidirectional dynamic microphone (Shure SM10A) which was connected to an external sound card (UA-1 G), and Audacity software (sampling rate 44.1 kHz, 16 bit, mono). The stimuli were presented on screen without any preceding context using Praat (Boersma and Weenink, 2017) and the presentation pace of each stimulus was controlled by the experimenter. Participants were instructed to silently read the stimulus on screen to understand its meaning, and then to utter it as if they were talking with someone. Participants were not forced to read the sentences under any time pressure. Once they had uttered the sentence, the new stimulus appeared
on screen. A pseudo-randomized list of stimuli was prepared for every participant to avoid any ordering effects.

3.1.4. Acoustic analysis

We recorded a total of 2240 utterances (40 participants × 56 stimuli). 338 utterances were excluded from any further analysis due to slips of the tongue, disfluencies and unnatural pausing. The remaining 1902 stimuli were manually annotated using Praat (Boersma and Weenink, 2017), as shown in Fig. 4. We obtained the following measurements using a number of Praat scripts.

3.1.5. Duration

(i) Utterance duration in ms
(ii) Word duration in ms; this was calculated based on the syllable duration, see Fig. 4. When the initial was a stop, we measured the stop from the beginning of the release burst.

3.1.6. F0

In Mandarin, a lexical tone is encoded by a pitch pattern (Xu, 2001; Xu and Wang, 2001). Following Duanmu (2004) (see also van de Weijer and Sloos, 2014) we used H(igh) and L(low) to operationalize the representation of the four lexical tones (T) in Mandarin. Hence, T1 is a high level tone and is represented as H, T2 is a rising tone and is represented as LH, T3 is a low tone represented as L and T4 is a falling tone represented as HL. We measured the following F0 points.

(iii) F0-maximum (H) of the syllable that bore T1 (high level tone)
(iv) F0-minimum and then F0-maximum (LH) of the syllable that bore T2 (rising tone)
(v) F0-minimum (L) of the syllable that bore T3 (low tone)
(vi) F0-maximum and then F0-minimum (HL) of the syllable that bore T4 (falling tone)
(vii) For T0 (neutral tone) of the perfective marker le, following Li (2002), we measured first the F0-maximum and then the F0-minimum, when the preceding syllable (verb) bore T1, T2 or T4.

On the other hand, we measured first the F0-minimum and then the F0-maximum, when the preceding syllable bore Tone 3. For the second syllable of the wh-word shénme, namely, me, when found in isolation it bears Tone 0. However, in our data, it behaved like a rising tone (T2); thus we treated it as such, measuring the F0-minimum and then the F0-maximum.

The obtained F0 values in Hz were converted into semitones (ST) to reduce variation across speakers; following Li and Chen (2012), for female speakers we used formula (i) ST = 12log2 (Hz/100), while for male speakers we used formula (ii) ST = 12log2 (Hz/50).

3.1.7. F0 range

We also calculated the F0 range in ST of the wh-word (shénme), and the post wh-word region, following previous studies (Dong, 2009; Liu et al., 2016).

(viii) F0 range of shén and me: Given that shén is a rising tone and that me also behaves like a rising tone, we calculated the F0 range of shén and me respectively, shén as F0-maximum of shén – F0-minimum of shén and me as F0-maximum of me – F0-minimum of me.

(ix) F0 range of post-wh-word region, namely, the prepositional phrase (PP) (e.g. gěi plus indirect object Liu Gāng “to/for Liu Gang”). Given that gěi carries a low tone, and that the first syllable of the indirect object bears a rising tone and the second syllable a high tone, the pitch contour in the whole prepositional phrase is in general a rising contour. Hence we calculated the F0 range of the prepositional phrase as F0-maximum of the second syllable of the indirect object (e.g. Gāng) – F0-minimum of gěi. See Fig. 5.

31.8. Intensity range

(x) Intensity range of each syllable defined as Maximum-Intensity—(minus) Minimum-Intensity (Chen, 2005; Ouyang and Kaiser, 2015), see Fig. 6.
3.2. Statistical analysis

We ran a series of linear mixed-effects models using the lmerTest package (Kuznetsova et al., 2013) in R (R Core Team, 2017). Specifically, for every measurement, we first ran a null model with the relevant measurement as the dependent variable, and participants and items as random factors. A second model included in addition clause type as a fixed effect factor. Finally, for each measurement, on the basis of the first two models that included the relevant measurement as a dependent variable, the clause type as a fixed-effect factor, and participants and items as random factors, allowing by-participant and by-item random intercepts, we ran another three models that in addition allowed by-participant or/and by-item random slopes for clause type. Model fit was compared based on the likelihood ratio test for each measurement (Pinheiro and Bates, 2000; Bolker et al., 2009); see Appendix A for the details of the fitting models in each measurement and their model fit comparison results.

3.3. Results

3.3.1. Utterance duration

We found an effect of clause type on utterance duration; wh-declaratives ($\bar{\tau} = 2050$ ms) are significantly longer than wh-questions ($\bar{\tau} = 2020$ ms); see Fig. 7.

3.3.2. Word duration

Fig. 8 presents the mean duration of all words in the utterance. We found an effect of clause type on the duration of the Subject, of the Verb plus the perfective marker le, of dīnār and of the wh-word. Specifically, the duration of the Subject, Verb plus le and dīnār in wh-declaratives ($\bar{\tau} = 347$ ms, $\bar{\tau} = 289$ ms and $\bar{\tau} = 171$ ms respectively) are significantly longer than those in wh-questions ($\bar{\tau} = 341$ ms, $\bar{\tau} = 261$ ms and $\bar{\tau} = 166$ ms respectively). The pattern changes when examining the duration of the wh-word shēnme. Shēnme in wh-declaratives ($\bar{\tau} = 294$ ms) is significantly shorter than in wh-questions ($\bar{\tau} = 305$ ms). When looking at the post-wh-word region, the prepositional phrase (gěi plus indirect object) in wh-declaratives ($\bar{\tau} = 572$ ms) does not differ from wh-questions ($\bar{\tau} = 572$ ms).

In short, wh-declaratives are in general longer in terms of duration than the corresponding wh-questions with an exception at the wh-word: the wh-word shēnme is longer in duration when it is a question word than when it is an existential meaning something. The detailed results of the mixed effects model can be found in Table 3.

3.3.3. F0

Fig. 9 presents the stylized means of F0 curves of the two clause types broken per verb tone (from T1 to T4). As is shown, the most striking F0 difference between the two clause types is at the wh-word shēnme (syllable (S) 8 and (S) 9), which shows a steep rise in wh-questions but is relatively flat in wh-declaratives, and the F0 in wh-questions remains higher than that in wh-declaratives until the end of sentence. To be specific, shēnme in wh-declaratives has lower F0 at the F0-minimum of

<table>
<thead>
<tr>
<th>Table 3: Summary of the linear mixed effects models on the durations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate $\beta$, Std. Error, $t$ value, $p$ value</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>Adverb</td>
</tr>
<tr>
<td>verb-le</td>
</tr>
<tr>
<td>dīnār</td>
</tr>
<tr>
<td>shēnme</td>
</tr>
<tr>
<td>Preposition phrase</td>
</tr>
</tbody>
</table>
The F0-minimum of me \((\bar{x} = 10.462 \text{ ST})\) and the F0-maximum of me \((\bar{x} = 12.392 \text{ ST})\) are smaller than that in wh-questions \((\bar{x} = 10.579 \text{ ST})\) for F0-minimum of shèn, \(\bar{x} = 12.008 \text{ ST}\) for F0-minimum of me and \(\bar{x} = 17.625 \text{ ST}\) for F0-maximum of me.

In the pre-wh-word region, we also found F0 differences at the verb when it bears T2. The F0-minimum and F0-maximum of the T2 verb in wh-declaratives \((\bar{x} = 11.947 \text{ ST}, \bar{x} = 13.929 \text{ ST})\) are lower than that of wh-questions \((\bar{x} = 13.274 \text{ ST}, \bar{x} = 15.465 \text{ ST})\) respectively. In the post-wh-word region, wh-declaratives are continuously lower in F0 than in wh-questions. To be specific, the F0-minimum of gěi in wh-declaratives \((\bar{x} = 8.820 \text{ ST})\) is lower than that in wh-questions \((\bar{x} = 12.036 \text{ ST})\); the F0-minimum and F0-maximum of the first syllable of the indirect object in wh-declaratives \((\bar{x} = 7.601 \text{ ST}, \bar{x} = 10.262 \text{ ST})\) are also lower than that in wh-questions \((\bar{x} = 10.328 \text{ ST}, \bar{x} = 12.017 \text{ ST})\); finally, the F0-maximum of the second syllable of the indirect object in wh-declaratives \((\bar{x} = 13.327 \text{ ST})\) is again lower that in wh-questions \((\bar{x} = 14.560 \text{ ST})\). The detailed results of the mixed effects model can be found in Table 4.

### Table 4

<table>
<thead>
<tr>
<th>Verb</th>
<th>Estimate (\beta)</th>
<th>Std. Error</th>
<th>(t) value</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0-min shèn</td>
<td>-0.714</td>
<td>0.171</td>
<td>-4.167</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-min me</td>
<td>-1.630</td>
<td>0.263</td>
<td>-6.199</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-max me</td>
<td>-5.298</td>
<td>0.393</td>
<td>-13.468</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-min verb (T2)</td>
<td>-1.351</td>
<td>0.541</td>
<td>-2.497</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>F0-max verb (T2)</td>
<td>-1.561</td>
<td>0.265</td>
<td>-5.894</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-min gěi</td>
<td>-3.355</td>
<td>0.438</td>
<td>-7.661</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-max indirect object (1st syllable)</td>
<td>-2.714</td>
<td>0.426</td>
<td>-6.365</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-max indirect object (1st syllable)</td>
<td>-1.817</td>
<td>0.237</td>
<td>-7.659</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0-max indirect object (2nd syllable)</td>
<td>-1.364</td>
<td>0.310</td>
<td>-4.401</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### 3.3.5. Intensity range

Fig. 10 shows the mean intensity range of each syllable (S) across the two clause types. As illustrated in the Fig. 10, the main differences between wh-declaratives and wh-questions are at the verb-le (S5 and S6), the wh-word shènme (S8 and S9) and gěi (S10). Wh-declaratives have a bigger intensity range at the verb \((\bar{x} = 20.210 \text{ dB})\) and at le \((\bar{x} = 7.701 \text{ dB})\) than wh-questions \((\bar{x} = 19.079 \text{ dB}, \bar{x} = 6.865 \text{ dB})\). The direction changes when looking at the wh-word. Shèn and me in wh-declaratives are significantly smaller in intensity range \((\bar{x} = 9.477 \text{ dB}, \bar{x} = 7.745 \text{ dB})\) than those in wh-questions \((\bar{x} = 10.995 \text{ dB}, \bar{x} = 9.470 \text{ dB})\). Finally, for the prepositional phrase after the wh-word, we only find intensity range differences at gěi, which is smaller in wh-declaratives \((\bar{x} = 15.315 \text{ dB})\) than that in wh-questions \((\bar{x} = 17.966 \text{ dB})\). The detailed results of the linear mixed effects model are summarized in Table 6.

### Table 5

Summary of the linear mixed effects models on F0 range with significant differences between clause types.

<table>
<thead>
<tr>
<th>Clause type</th>
<th>Estimate (\beta)</th>
<th>Std. Error</th>
<th>(t) value</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 range me</td>
<td>-3.653</td>
<td>0.274</td>
<td>-13.146</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F0 range preposition phrase</td>
<td>1.975</td>
<td>0.487</td>
<td>4.054</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Table 6

Summary of the linear mixed effects models on the intensity range with significant differences between clause types.

<table>
<thead>
<tr>
<th>Verb</th>
<th>Estimate (\beta)</th>
<th>Std. Error</th>
<th>(t) value</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb</td>
<td>1.119</td>
<td>0.179</td>
<td>6.243</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>le</td>
<td>0.833</td>
<td>0.224</td>
<td>3.713</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>shèn</td>
<td>-0.620</td>
<td>0.157</td>
<td>-3.963</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>me</td>
<td>-1.729</td>
<td>0.351</td>
<td>-4.927</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>gěi</td>
<td>-2.662</td>
<td>0.461</td>
<td>-5.767</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
4. Discussion and conclusions

The starting point of our studies is a new and interesting observation that wh-sentences containing *dānr* is ambiguous between wh-questions and wh-declaratives (wh-word has an existential interpretation), challenging previous studies which claim that the existential interpretation of a wh-word is available only in nonveridical sentences. To empirically test this observation, we conducted a reading study on wh-sentences containing *dānr*. Another related question concerns how string identical wh-questions and wh-declaratives are distinguished prosodically. By reviewing previous studies, we found that an important knowledge gap still remains with respect to the detailed prosodic differences between Mandarin wh-declaratives and wh-questions. We therefore conducted a production study on wh-declaratives and wh-questions, examining the prosodic properties of each clause type.

The results of the reading study show that in Mandarin, wh-sentences that contain *dānr* are ambiguous and can be interpreted as wh-questions or wh-declaratives. In the former case the wh-word is interpreted as a question word, while in the latter case the wh-word is interpreted as an existential 'something’. This empirical finding challenges the state-of-the-art assumption that the existential interpretation of a wh-word is licensed only in nonveridical sentences (negation, questions, conditionals and epistemic modalities) (see Lin, 1998; Xie, 2007; Lin et al., 2014; Huang, 2018), and calls for a reconsideration of the licensing conditions of existential interpretations of wh-words in Mandarin (see Yang, 2018).

The results of the production study show that prosody indeed marks the clause type (declarative vs. question) and differentiates wh-declaratives from string identical wh-questions. Specifically, wh-declaratives differ from wh-questions in duration, F0 and F0 range as well as intensity range. Wh-declaratives are uttered with a longer du-
ration than the corresponding wh-questions in general. As for F0, consistent with previous studies, wh-words in wh-questions exhibit a steep F0 rise and expanded F0 range (Hu, 2002; Lee, 2005; Dong, 2009; Liu, 2009, 2016). The findings on F0 are also consistent with previous studies claiming that the presence of a high pitch is often a property of question intonation cross-linguistically (van Heuven and van Zanten, 2005). In addition to F0, wh-words present a longer duration and a bigger intensity range in wh-questions than in words used in wh-declaratives.

Moreover, we find that duration not only marks the clause type, but also marks it “early”. As shown in the results of the production experiment, it is mainly duration that provides an early cue to differentiate the two interpretations from the onset of the utterance (sentence subject), with wh-declaratives always longer than wh-questions and at the verb-le the difference reaches the peak. This early durational property (a shorter duration in wh-questions as opposed to declaratives) can be considered as another important feature of Mandarin wh-questions, in addition to the commonly known higher F0. This finding of the duration property is in general consistent with studies of other languages reporting that duration plays a role in marking questions (Lindsey, 1985; van Heuven and van Zanten, 2005; Cangemi and D’Imperio, 2013). In the post-wh-word region, wh-questions show a long-lasting higher pitch and their F0 range is smaller than that in wh-declaratives, indicating a typical F0 compression after wh-words in wh-questions. In addition to the above findings on the different prosodic markings of the two clause types, our production results also shed light on the discussion of the focal property of wh-words and their prosodic realizations. Following the approach to focus proposed by Jacobs (1984, 1991), the partition of content into focus and ground is related to illocutionary semantics: focus is defined as the part of the sentence that is specifically affected by the sentence’s illocutionary operator. According to this, in a wh-question, which is an information-seeking question, the interrogative wh-word is typically a focus (see Cho, 1990; Lambrecht and Michaelis, 1998; Deguchi and Kitagawa, 2002; Ishihara, 2002, among others), while in a wh-declarative, where the wh-word is a narrow scope indefinite ‘something’ and normally cannot be affected by an illocutionary operator, is not a focus. Our production results provide further support to identifying the different focus status of wh-words in wh-questions in contrast with wh-words in wh-declaratives in Mandarin.

It should be noted that different languages may have different ways to realize focus, ranging from morpho-syntactic markers to prosodic means. For those languages that have a general correspondence between focus and the prosodic marking, focused constituents are typically characterized with an expanded pitch range and the post-focal regions typically show a compressed F0 range (for Germanic languages, see Cruttenden, 2006, Fery and Kügler, 2008, among others; for Mandarin studies, see Xu, 1999; Yuan, 2004; Li, 2009; Chen, 2010; Xu et al., 2012, among others). In addition, a focused element quintessentially has a longer duration and greater intensity (Xu, 1999; Chen et al., 2009; Li, 2009). In our production study, we found that wh-questions have a raised and expanded F0 range, lengthened duration and greater intensity range at the interrogative wh-word shénme and an F0 range compression after shénme, consistent with the prosodic markings of focus in Mandarin. In contrast, our study reveals that non-interrogative wh-words in wh-declaratives are largely suppressed with an almost flat pitch, a short duration and a small intensity range, inconsistent with the prosodic realizations of a focus in Mandarin. The different prosodic realizations of the wh-words in wh-questions and wh-declaratives serve as additional evidence that wh-words have different focal properties as a question word in wh-questions as compared with an indefinite in wh-declaratives. To conclude, in this study we have established that sentences containing diānr followed by a wh-word are indeed ambiguous between wh-questions and wh-declaratives, counterintuitive to previous studies. Furthermore, we have uncovered that wh-questions and wh-declaratives are distinguished by prosody, supporting the role of prosody in marking clause types in the absence of an overt syntactic interrogativity marker (Boëling, 1978; Ohala, 1983, 1984; Jun and Oh, 1996; Frota, 2002; Face, 2004; Vion and Colas, 2006; Baltazani, 2007). Our production results have demonstrated that wh-questions and wh-declaratives differ in various prosodic properties, not limited to F0; it is duration that marks the two clause types early as wh-declaratives are continuously longer than wh-questions in terms of word duration since the onset of the sentence (subject). The current study also indicates that wh-questions and wh-declaratives differ in terms of the information structure; in particular, a wh-word in Mandarin is a focus when it is a question word (e.g., ‘what’), but it is not a focus when it is an indefinite (interpreted as an existential ‘something’) used in a wh-declarative.

Declaration of Competing Interest

None.

Acknowledgments

We thank Xiaolu Yang for assistance in conducting the recording experiment and Jos Pacilli for assistance in writing Praat scripts. We also thank Leticia Pablos, Jenny Doetjes and Aliza Glasbergen-Plas for their comments and discussions during the whole process of experimental design and data analysis. The study was presented in the 2nd International
Conference on Chinese Prosodic Grammar and we thank Laura Downing, San Duanmu, Eric Zee and Dylan Tsai for their useful comments.

**Funding**

This study was supported by a project granted by the Philosophy and Social Science Foundation of Guangdong Province [grant number GD19YY030] and a project granted by Guangzhou Philosophy and Social Science Foundation [grant number 2019GZGJ66] on the first author and was also funded by the NWO project Understanding Questions [project number 360-70-480] on the second author.

**Appendix A**

Summary of all the fitting models on prosodic measurements where the clause type has an effect and their model fit comparison results.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>The fitting model</th>
<th>Model fit comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utterance speaking rate</td>
<td>Utterance ~ speaker + (condition</td>
<td>subject) + (1</td>
</tr>
<tr>
<td>Utterance duration</td>
<td>Utterance ~ Utterance ~ condition + (condition</td>
<td>subject) + (1</td>
</tr>
<tr>
<td>Word duration</td>
<td>Word ~ (duration</td>
<td>condition + (1</td>
</tr>
<tr>
<td></td>
<td>Word ~ (duration</td>
<td>condition + (1</td>
</tr>
<tr>
<td></td>
<td>shēnme ~ (duration</td>
<td>condition + (1</td>
</tr>
<tr>
<td></td>
<td>F0 ~ (duration</td>
<td>condition + (1</td>
</tr>
<tr>
<td></td>
<td>F0 ~ (duration</td>
<td>condition + (1</td>
</tr>
<tr>
<td></td>
<td>F0 ~ (duration</td>
<td>condition + (1</td>
</tr>
</tbody>
</table>

**References**


