Article details

Clause type anticipation based on prosody in Mandarin

Yang Yang,[1,2] Stella Gryllia,[2] Leticia Pablos,[2,3] and
Lisa Lai-Shen Cheng[2,3]
[1] Center for Linguistics and Applied Linguistics, Guangdong University
of Foreign Studies | [2] Leiden University Centre for Linguistics | [3] Leiden
Institute for Brain and Cognition

Mandarin *wh*-words such as *shénme* are *wh*-indeterminates, which can have
interrogative interpretations ('what') or non-interrogative interpretations
(i.e., 'something'), depending on the context and licensors. For example,
when *diǎnr* ('a little') appears right in front of a *wh*-word, the string can
have either a *wh*-question or a declarative interpretation (henceforth,
wh-declarative). Yang (2018) carried out a production study and the results
showed that *wh*-questions and *wh*-declaratives have different prosodic prop-
erties. To investigate whether and when listeners make use of prosody to
anticipate the clause type (i.e., question vs. declarative), we conducted a sen-
tence perception study and an audio-gating experiment. Results of the per-
ception study and the gating experiment show that (1) Participants can
make use of prosody to differentiate the two clause types; (2) Starting from
the onset of the first word of the target sentence
(*wh*-question/*wh*-declarative), participants already demonstrate a prefer-
ence for the clause type that was intended by the speaker. The current study
also sheds light on the clausal typing mechanism in Mandarin (e.g., how to
mark a clause as a *wh*-question) by providing evidence of the role of
prosody in marking clause types in Mandarin.

Keywords: Mandarin *wh*-declarative/*wh*-question, clause type anticipation,
prosody, perception and gating paradigm, clausal typing

1.   Introduction

1.1   Anticipation in language

Anticipation in general refers to the prediction of an action or of an event. Anticipation in language normally refers to how readers or listeners use a variety of syntactic, morphosyntactic, discourse-semantic and prosodic cues to predict upcoming linguistic material. For example, Staub and Clifton (2006) showed in an eye-tracking reading study that readers can predict an upcoming "*or*-clause" if they have read an "*either*-clause" before. In an earlier study that used the Visual World Paradigm, Altmann and Kamide (1999) found that listeners can predict the complement of a verb on the basis of its selectional properties.

1.2   Anticipation and prosody

A number of studies have examined the role that prosodic cues like duration and pitch accents play during processing in anticipation of upcoming linguistic material. For instance, Grosjean (1983) showed that listeners of English were able to predict the end of a sentence when basing themselves solely on prosodic cues, and Fo and duration cues are better predictors than amplitude. Swerts, Collier and Terken (1994) found that listeners of Dutch were able to predict the approaching end of a discourse using prosodic cues like pitch contour variation and degree of lengthening of the last word of the utterance. Furthermore, a number of studies reported anticipatory effects of pitch accents in reference resolution or information status interpretation (Terken & Hirschberg 1994; Dahan, Tanenhaus & Chambers 2002; Weber, Braun & Crocker 2006; Chen, den Os & de Ruiter 2007; Braun & Chen 2012, among others). For example, Weber, Braun and Crocker (2006) showed that contrastive pitch accents in German allow listeners to identify referents faster and to anticipate upcoming referents. It has also been shown that listeners use prosodic information to predict the upcoming syntactic structure (Beach 1991; Beach, Katz & Skowronski 1996; Kjelgaard & Speer 1999; Carlson, Clifton & Frazier 2001; Snedeker & Trueswell 2003, among others). For instance, Snedeker and Trueswell (2003) reported that listeners use prosodic cues such as prosodic breaks to predict the attachment site of an upcoming prepositional phrase which could attach to the VP or the NP.

1.3   Clause type anticipation and prosody

1.3.1   *Previous studies*

In many languages, the different clause types of sentences have been reported to be encoded by different prosody (Hermann 1942; Bolinger 1978; Ohala 1983, 1984, among others). For instance, Hermann (1942) has mentioned that the presence of a high pitch is often a property of question intonation cross-linguistically, differentiating questions from declaratives. In addition, a short duration also plays a role in marking questions (Lindsey 1985; Van Heuven & Van Zanten 2005). A number of studies have also examined whether listeners use prosodic cues to anticipate the upcoming clause type of an utterance (declarative or question) using the audio-gating paradigm (Face 2004; Vion & Colas 2006; Petrone & D'Imperio 2011; Petrone & Niebuhr 2014). In the typical audio-gating paradigm, fragmented audio-stimuli of different lengths are presented to listeners. After hearing each gate, they are instructed to decide whether the audio fragment is going to be a declarative or a question and to indicate their degree of confidence (cf. Grosjean 1980). Face (2004) showed that listeners of Castilian Spanish are able to differentiate a declarative from a string identical yes-no question after hearing the first pitch accent of the sentence, and thus they use the scaling of the initial F0 peak to anticipate the upcoming clause type. Similarly, Vion and Colas (2006) found that listeners of French use the down stepped pitch in the middle of the sentence in predicting that a yes-no question is upcoming. For Neapolitan Italian, Petrone and D'Imperio (2011) found that listeners use the H or L tone at the right edge of the Accentual Phrase (AP) in the pre-nuclear contour to predict whether the utterance will be a yes-no question or a declarative. Listeners of Northern Standard German (Petrone & Niebuhr 2014) can identify and predict the upcoming yes-no question or declarative based on the F0 shape, slope, and alignment of the pre-nuclear pitch accent. However, for a tonal language like Mandarin Chinese, there is not much research on the use of prosodic cues for clause type anticipation.

1.3.2   *The case of Mandarin*

Before addressing the issue of clause type anticipation in Mandarin, we present briefly some properties of Mandarin. Different from the languages mentioned in the discussion concerning clause type anticipation, Mandarin is a *wh*-in-situ language, in which *wh*-words remain at their base position just as their declarative counterparts do, as illustrated in (1a–b). (1a) is an ordinary information-seeking question, while (1b) is its corresponding declarative.

(1) a. Luó Wēi zuótiān mǎi-le   shénme?
          Luo Wei yesterday buy-PERF what
          'What did Luo Wei buy yesterday?'   [*wh*-question]
    b. Luó Wēi zuótiān mǎi-le   tízi.
          Luo Wei yesterday buy-PERF grapes
          'Luo Wei yesterday bought grapes.'   [declarative]

Diverging from *wh*-movement languages, *wh*-in-situ languages have no fronted *wh*-words to indicate the clause types. Moreover, Mandarin *wh*-words such as *shénme* are not only in-situ but are also known as *wh*-indeterminates, which can have interrogative ('what') and non-interrogative interpretations (e.g., existential interpretation 'something' or universal interpretation 'everything'), depending on the context and licensors as in Japanese and Korean (Huang 1982; Cheng 1991; Li 1992; Lin 1998). For example, when *diǎnr* ('a little') appears in front of a *wh*-word, the *wh*-word can have an interrogative or a declarative interpretation, as illustrated in (2). (2a) is a declarative sentence and the *wh*-word *shénme* is interpreted as an indefinite, meaning 'something'; we will use the term *wh*-declarative to refer to such declaratives. (2b) is the question counterpart of (2a), where the *wh*-word is interpreted as an interrogative, and we will call such sentences *wh*-questions. Both (2a) and (2b) contain the word *diǎnr* 'a little', which is a determiner with existential quantificational force (Tsai 2010) and licenses the existential reading of *shénme* ('something') in (2a). As we can see in (2), (2a) and (2b) are string identical, but differ in their interpretations.

(2) a. Zhāng Sān mǎi-le   diǎnr shénme.
          Zhang San buy-PERF a.little SHENME[1]
          'Zhang San bought a little of something.'   [*wh*-declarative]
    b. Zhāng Sān mǎi-le   diǎnr shénme?
          Zhang San buy-PERF a.little SHENME
          'What did Zhang San buy (a little of)?'   [*wh*-question]

When presented in written form without a punctuation mark and out of context, (2a) / (2b) is, in principle, ambiguous between a *wh*-declarative and a *wh*-question. Despite of the fact that the above *wh*-question and *wh*-declarative pair offers an ideal testing case for anticipating clause types based on prosody only, there has been no study of such pairs up to now. Previous studies that investigate clause type anticipation in Mandarin are restricted to yes-no questions and declaratives only. The only finding so far is that the sentence final syllable plays a critical role in differentiating and identifying yes-no questions and declaratives

---

1. From here on, we gloss *shénme* simply as SHENME, as it can have interrogative and non-interrogative interpretations.

(cf. Yuan 2004, 2006 for a detailed discussion). Although prosodic differences between Mandarin yes-no questions and declaratives are also manifested in non-final syllables, it remains unclear whether clause type anticipation takes place before the sentence final syllable given that no gating studies have been conducted in Yuan (2004, 2006).

Recapitulating, previous studies on clause type anticipation have shown that listeners use prosodic cues to predict an upcoming yes-no question or declarative. However, it remains unknown whether listeners also use prosodic cues to antic-ipate an upcoming *wh*-question or declarative. Mandarin is an ideal test case for clause type identification and anticipation, as *wh*-questions and *wh*-declaratives are string identical as shown in (2).

1.4   Mandarin clause type anticipation and the clausal typing mechanism

Examining the role of prosody in identifying and anticipating clause types in Mandarin can also shed light on a recalcitrant issue, namely, the clausal typing mechanism in Mandarin (e.g., how to mark a clause as a *wh*-question).

One central question concerning *wh*-in-situ languages is how these languages realize clausal typing. According to the Clausal Typing Hypothesis by Cheng (1991: 29), "Every clause needs to be typed. In the case of typing a *wh*-question, either a *wh*-particle in C° is used or else fronting of a *wh*-word to the Spec of C° is used, thereby typing a clause through C° by spec-head agreement." Hence *wh*-movement languages such as English type their *wh*-questions by fronting their *wh*-words (i.e., _What have you moved?_), while *wh*-in-situ languages achieve clausal typing by utilizing a *wh*-particle. For instance, in Japanese the presence of *ka* at clause final position (e.g., C°) types a *wh*-question. For Mandarin, clausal typing is less straightforward. As already discussed, Mandarin lacks *wh*-movement. Moreover, Mandarin lacks a *wh*-particle. Although previous stud-ies took the Mandarin particle *ne* (a so-called Sentence Final Particle (SFP)) as a *wh*-particle in *wh*-questions (Cheng 1991), others have argued that it is problem-atic to do so: First, *ne* is optionally used in *wh*-questions; second, *ne* appears in matrix *wh*-questions but it does not appear in embedded *wh*-questions; third, in addition to *wh*-questions, *ne* is also used in A-not-A questions (a type of yes-no question); moreover, *ne* appears not only in questions, with a different prosody, it can also appear in declaratives (Li 2006; Constant 2014). If *wh*-particles and the movement of *wh*-words are not present to type a *wh*-question, how is clausal typ-ing realized in a language like Mandarin?

Previous studies have shown that there is a correlation between SFPs and into-nation (Zhang 2014; Tang 2015; Wakefield 2016). For instance, SFPs and some spe-cific intonation are in complementary distribution. Although it is not clear how

sentence intonation is affected by the presence or absence of SFPs, especially in Mandarin, this line of thinking suggests an additional mechanism of clausal typing, in particular, using intonation/prosody. The current study on *wh*-questions and *wh*-declaratives takes up this line of thinking, by examining whether and how prosody marks and types Mandarin *wh*-questions and *wh*-declaratives and hence providing empirical evidence for the role of prosody in clausal typing in Mandarin.

1.5   Research questions

This study addresses the following research questions:

RQ1.   Can listeners classify correctly *wh*-questions and *wh*-declaratives[2] as in (2) on the basis of their prosodic properties?

RQ2.   If yes, at which point of the sentence can listeners anticipate the upcoming clause type?

RQ3.   What does the clause type anticipation tell us about the clausal typing mechanism in Mandarin?

Answering these questions, we not only come to understand the role of prosody in identifying and anticipating clause types, but also understand how prosody can be a reliable clausal typer for Mandarin *wh*-questions, just as the fronting of *wh*-words or the presence of *wh*-particles in other languages, hence enriching the Clausal Typing Hypothesis.

   The paper is organized as follows. In Section 2, we present the results of a perception experiment that addresses the first research question. Section 3 reports an audio-gating experiment that addresses the second research question and in Section 4 we present our conclusions, addressing the third research question.

---

2. One reviewer pointed out that by only investigating one case (*wh*-question and *wh*-declarative), it is too risky to give a conclusion or formulate an extended hypothesis. In the current study, in order to solely focus on the role of prosody in clause type identification and anticipation, we adopted *wh*-questions and declaratives that are string identical. Our other study (Gryllia, Doetjes, Yang & Cheng 2019) has investigated the production and perception/ gating of *wh*-questions and ordinary declaratives containing noun phrases (hence no *diǎnr shénme*), and the results are consistent with the findings in the current study. Taken together, the result does not only apply to a single case.

2.   Perception experiment

We ran a perception experiment to address the first research question. During the perception experiment, native speakers of Mandarin Chinese heard an audio stimulus (a *wh*-declarative or a *wh*-question as shown in (3)) and were asked to continue the discourse choosing one of two possible continuations, namely, a noun phrase or a *wh*-question, see Examples (4a–b). Choosing a noun phrase as a continuation signals that listeners interpreted the audio stimulus as a *wh*-question, while choosing a *wh*-question as a continuation signals that listeners interpreted the stimulus as a *wh*-declarative.

[Audio stimulus]
(3) Táo Wēi zuótiān ná-le   diǎnr shénme gěi Líu Gāng
     Tao Wei yesterday bring-PERF a.little SHENME to Liu Gang
     'Tao Wei brought a little of something to Liu Gang yesterday.'
     'What did Tao Wei bring (a little of) to Liu Gang yesterday?'

[Discourse continuations on screen]
(4) a.   Tízi.
           grapes
           Grapes.
     b.   Táo Wēi ná-le   shénme?
           Tao Wei bring-PERF what
           'What did TaoWei bring?'

2.1   Participants

Thirty-six native speakers of Beijing Mandarin (16 female, 20 male, $\bar{x}$ age = 19 years old, SD = 2.8) were paid to participate in the experiment. All of them came from the northern part of Mainland China and at the time of testing were students at Tsinghua University. Their study background is diversified and none of them have a background of Linguistics. None of them reported any hearing or vision disorders after correction. Prior to testing, informed written consent was obtained from each participant.

2.2   Audio stimuli

The stimuli consisted of two sets of clause types and each set contained 20 items. These 40 stimuli were selected from the productions of a female native speaker of

Beijing Mandarin[3] (age = 20 years' old) from a previous production study[4] (Yang, Gryllia & Cheng, 2019).

An example of a set of stimuli is given in (5). The word order of each stimulus was kept constant: Subject (proper name, e.g., *Táo Wēi* "Tao Wei"), Adverb (e.g., *zuótiān* "yesterday"), Verb (e.g., *ná* "bring") and Perfective marker (*le*), *diǎnr*, Direct Object (*shénme* "what/something"), Prepositional Phrase (e.g., *gěi Líu Gāng* "to/for Liu Gang"). Each stimulus consisted of 12 syllables and the stimulus length was kept constant across clause types and items. As is well known, Mandarin has four lexical tones as well as a neutral tone. We kept the combination of tones constant across clause types and items for all constituents, except for the verb. For the verb, we included all four lexical tones, to obtain more natural stimuli. The tonal composition of the stimuli is indicated in (5).

(5) a.   Táo Wēi zuótiān   ná-le   diǎnr shénme gěi   Líu Gāng?
          T2 T1 T2 T1   T2-T0   T3   T2 T0 T3   T2 T1
          Tao Wei yesterday bring-PERF a.little what   to/for Liu Gang
          'What did Tao Wei bring (a little) for Liu Gang yesterday?'   [*wh*-question]

---

3. In a previous study, we conducted a production experiment on 40 Mandarin speakers from Beijing (23 females and 17 males, with an average age 21 years old). We chose this female speaker (20 years old) and used her production results as stimuli for the current study for two reasons: (1) Although the speaker is not from a broadcasting major, she was active in taking part in the broadcasting and hosting events, according to her self-report. Her speech is clear, natural, and with a good pace, and no participants in the perception/gating experiment (36 participants in total) have complained about her speech. (2) The prosodic properties of the stimuli by this speaker are in general consistent with the average prosodic markings of *wh*-questions and *wh*-declaratives of the 40 speakers in the production experiment. We admit the limitations of choices made based on more subjective criteria, and for the future studies, we will conduct a judgement task when choosing a speaker.

4. The production experiment 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-1G), and Audacity software (sampling rate 44.1 kHz, 16 bit, mono). The stimuli were presented on screen without any preceding context using Praat (Boersma & Weenink 2016) 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 before uttering it as if they were talking with someone. After they uttered the sentence, a new stimulus appeared on screen. A pseudo-randomized list of stimuli was prepared for every participant to avoid any ordering effects.

b. Táo Wēi zuótiān ná-le   diǎnr shénme gěi   Líu Gāng.
    T2 T1 T2 T1   T2-T0   T3   T2 T0   T3   T2 T1
    Tao Wei yesterday bring-PERF a.little something to/for Liu Gang
    'Tao Wei brought a little something for Liu Gang yesterday.'

[*wh*-declarative]

2.2.1 *Prosodic properties of the stimuli*

Our aim is to investigate whether listeners can identify the different clause types based on the different prosodic cues of the stimuli, and therefore, we need to report the prosodic properties of the stimuli (questions and declaratives) and to test whether the prosodic properties of the stimuli between the two clause types are significantly different. To do so, we ran a series of linear mixed effects models using the *lmerTest* package (Kuznetsova, Brockhoff & Christensen 2013) in R (R Core Team 2017) to test whether the prosodic properties of the stimuli between the two clause types are significantly different. Specifically, for duration, we included the total duration of the stimulus or the duration of each syntactic constituent as a dependent variable, clause type as a fixed effect factor and item as a random factor. For F0, we included the specific F0 measurement as a dependent variable, clause type as a fixed effect factor and item as a random factor.

Duration
In general, *wh*-declaratives ($\bar{x} = 2197$ ms) are significantly longer than *wh*-questions ($\bar{x} = 2049$ ms), see Figure 1. When examining the duration of each syntactic constituent, we observe that the duration of the Subject, the Verb plus *le* and the Prepositional Phrase in *wh*-declaratives (Subject, $\bar{x} = 344$ ms, Verb-*le*, $\bar{x} = 296$ ms, Prepositional Phrase, $\bar{x} = 644$ ms) is significantly longer than the corresponding duration of these items in *wh*-questions (Subject, $\bar{x} = 325$ ms, Verb-*le*, $\bar{x} = 256$ ms, Prepositional Phrase, $\bar{x} = 571$ ms), see Figure 2. The *wh*-word *shénme* is longer in *wh*-questions ($\bar{x} = 319$ ms) than in *wh*-declaratives ($\bar{x} = 307$ ms), but the difference is not significant. The results of the linear mixed effects models are presented in Table 1.

Table 1. Summary of the linear mixed effects models on duration

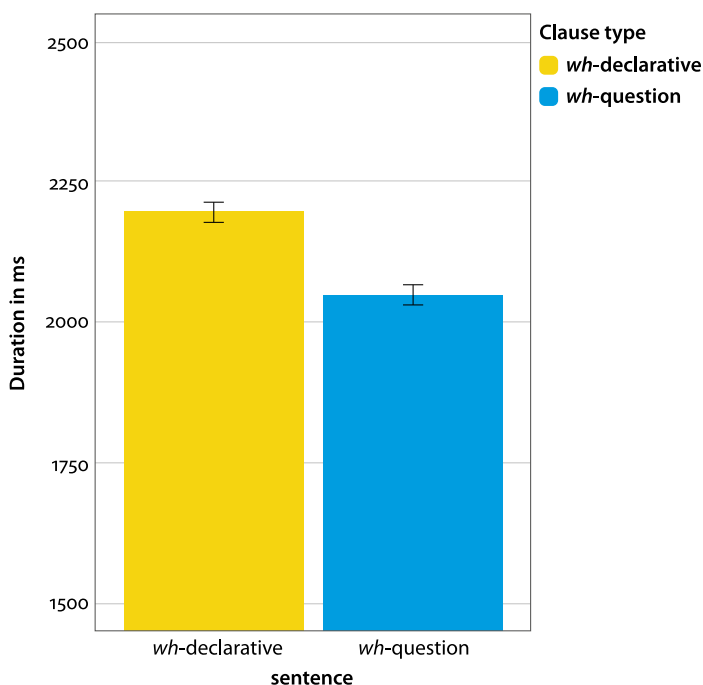|	Estimate β	Std. Error	*t*-value	*p*-value
the whole sentence	148.00	20.00	7.40	< 0.001
subject	18.45	5.36	3.44	< 0.01
verb-*le*	39.45	8.78	4.49	< 0.001
shénme	−9.30	5.97	−1.56	> 0.1
Prepositional Phrase	73.00	9.13	8.00	< 0.001

Figure 1. Mean sentence duration across clause types with error bars showing standard error

F0

We also examined the F0 of the stimuli comparing the F0 properties across the two clause types. Mandarin tones are dynamic pitch targets (Xu 2001; Xu & Wang 2001), but given that our research question does not tap into the realizations of each dynamic tonal contour, but rather focus on the prosodic differences between questions and declarative, we adopt representations with H(igh) and L(ow) following Duanmu (2004) (cf. van de Weijer & Sloos 2014) in describing 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 operationalized the phonological representation of the four lexical tones, measuring the F0-maximum for T1, the F0-minimum and the F0-maximum for T2, the F0 minimum for T3, and the F0-maximum and F0-minimum for T4. For the 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, while we measured first the F0-minimum and then the F0-maximum, when the preceding syllable bore T3. Figure 3 presents the F0 measurements for a *wh*-question. For the
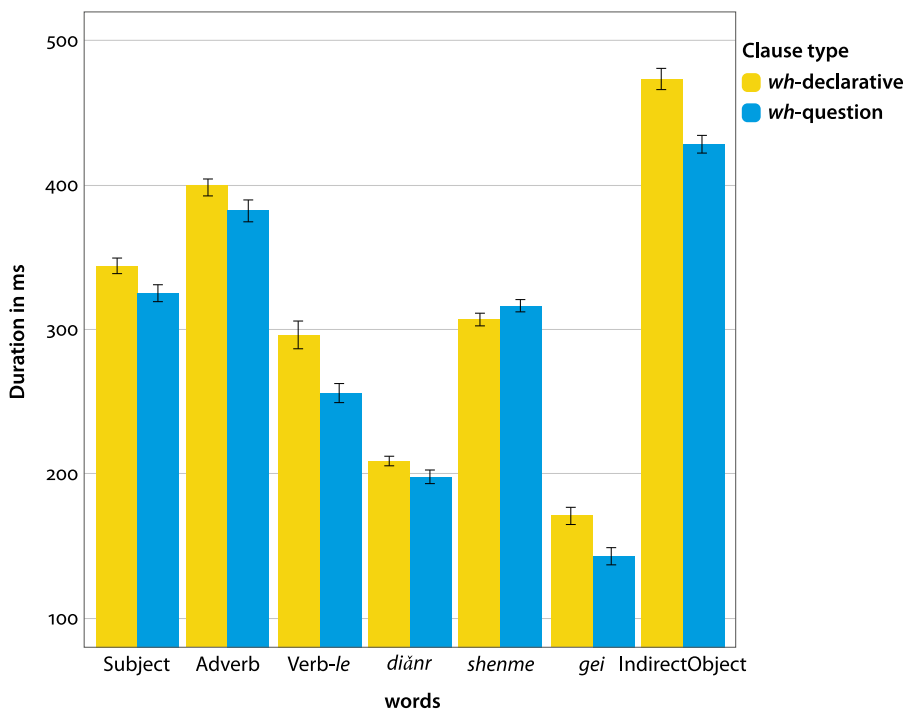
Figure 2. Mean word duration across clause types with error bars showing standard error

wh-word *shénme*, as it behaved like a rising contour in our data, we treated it as such, measuring the F0-minimum and then the F0-maximum. We also calculated the F0 range of the *wh*-word (*shénme*), following previous studies (Dong 2009; Liu, Li & Jia 2016). Given that the pitch contour at *shénme* is in general a rising contour, we calculated its F0 range as F0-maximum of *me* – F0-minimum of *shén*.

Figure 4 presents the mean F0 values of the stylized F0 curves for *wh*-questions and *wh*-declaratives split per verb tone. In general, the F0 contour of *wh*-questions is higher than the corresponding F0 contour of *wh*-declaratives. The most prominent difference between the two clause types is at the *wh*-word *shénme*, which shows a steep rise in *wh*-questions while it is relatively flat in *wh*-declaratives; the F0 range of the *wh*-word *shénme* in *wh*-questions ($\bar{x} = 90$ Hz) is significantly bigger than that in *wh*-declaratives ($\bar{x} = 6$ Hz). Table 2 presents the results of the linear mixed effects models.

2.3 Procedure

The experiment was set up with MFC Praat (Boersma & Weenik 2016) and took place in a sound-proof booth in a laboratory of Tsinghua University in Beijing.
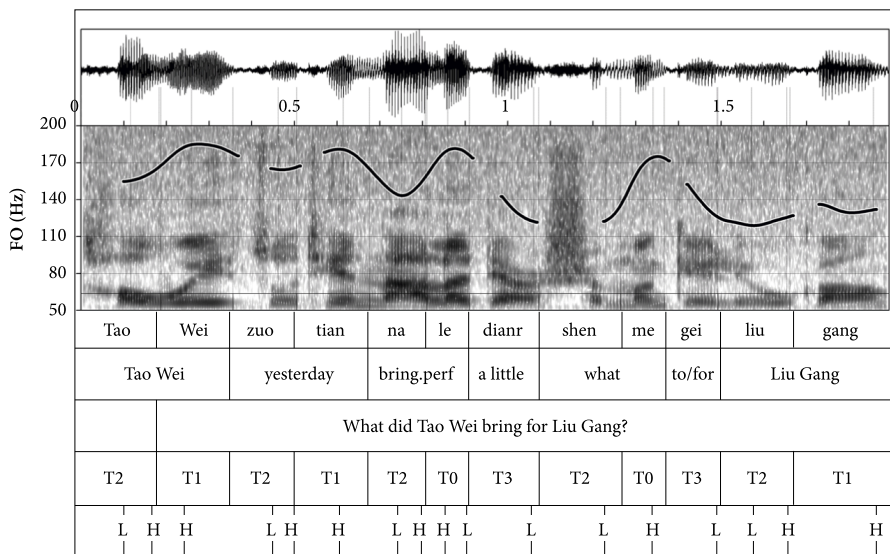
Tao	Wei	zuo	tian	na	le	dianr	shen	me	gei	liu	gang
Tao Wei		yesterday		bring.perf		a little	what		to/for	Liu Gang	
What did Tao Wei bring for Liu Gang?											
T2	T1	T2	T1	T2	T0	T3	T2	T0	T3	T2	T1
L H H		L H H		L H H L		L	L	H	L L H		H

Figure 3. The waveform and spectrogram with superimposed F0-contours, syllables, glosses and F0 measurement obtained based on the specific tones
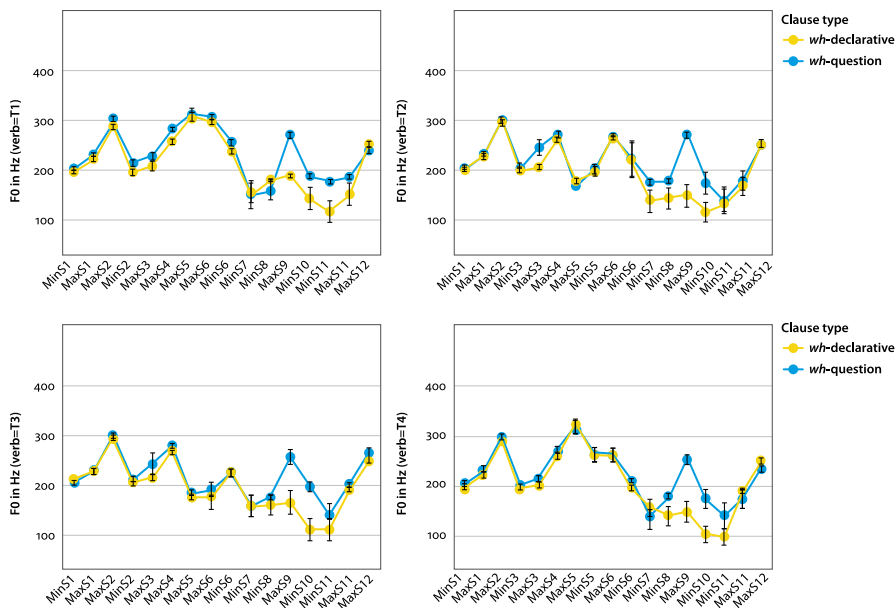

Figure 4. Stylized means of F0 curves across clause types and verb tones with error bar showing standard error (Min indicates the F0-minimum and Max the F0-maximum); the *wh*-word *shénme* is represented by S8 and S9

Table 2. Summary of the linear mixed effects models on the F0 measurement between clause types (significant differences were reported)

	Estimate β	Std. Error	t-value	p-value
F0-min adverb (1st syllable)	8.66	2.67	3.24	< 0.01
F0-max adverb (1st syllalbe)	24.65	7.95	3.10	< 0.01
F0-max adverb (2st syllalbe)	14.16	3.88	3.65	< 0.01
F0-min *le*	8.46	3.21	2.64	< 0.05
F0-max *me*	100.60	10.15	9.91	< 0.001
F0-min *gěi*	65.27	11.24	5.81	< 0.001
F0-min indirect object (1st syllable)	34.61	14.23	2.43	< 0.05
F0 range *shénme*	83.44	6.64	6.44	< 0.001

Participants were seated in front of a computer screen at an approximate distance of 50 cm and were asked to read silently the instructions that appeared on their screen. Once they were ready, they clicked on the OK button on their screen to start the experiment. The screen was empty and after 1.0 second, the audio stimulus was played. 0.3 seconds after the offset of the audio stimulus the two discourse continuations [i) a noun phrase and ii) a *wh*-question] appeared on the left and the right of the screen. Participants were asked to select one of the two continuations to complete the discourse on the basis of the audio stimulus they had heard. After their selection, participants had to click the OK button to submit their response and 1.0 second after clicking OK, the next audio stimulus was played. The audio stimuli were randomized for each participant to avoid a sequence effect; and the two continuations were counterbalanced to avoid any left/right preference among participants. There was no time pressure in indicating the responses; participants could spend as much time as they needed on this. The experiment lasted approximately 10 minutes.

2.4   Analysis and results

We obtained a total of 1440 responses (40 stimuli × 36 participants). When the audio stimulus was a *wh*-question, listeners chose a noun phrase as a discourse continuation at 93.9% of the time, thus correctly interpreting the stimulus as a *wh*-question. When the audio stimulus was a *wh*-declarative, listeners chose a *wh*-question as a discourse continuation at 95.0% of the time, thus correctly interpreting the stimulus as a *wh*-question, see Figure 5. The highest accuracy rate per participant was 100%, and the lowest accuracy rate was 70%. A chi-square analysis

showed that there was a significant association between clause types intended by the speaker and listeners' responses, $x^2 = 1137.92$ (1), $p < 0.001$.
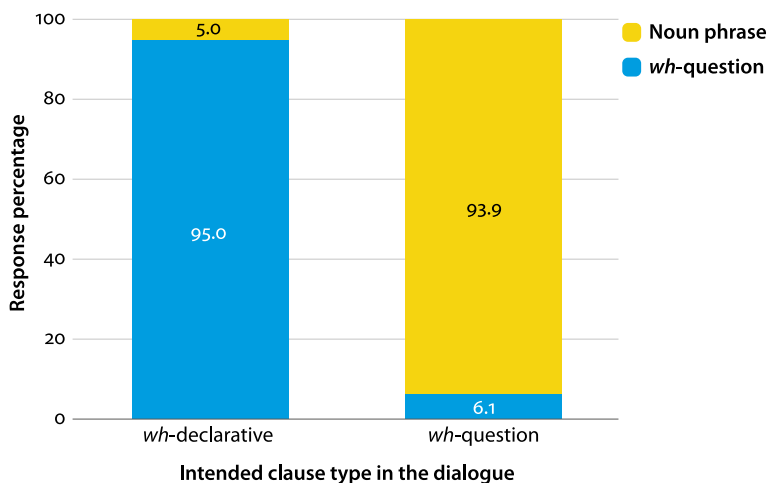

Figure 5. Listeners' responses in percentage (%) in the perception experiment

Moreover, we ran a mixed effects logistic regression using the *glmer* command in the *lme4* package (Bates, Maechler, Bolker & Walker 2015) in R (R Core Team 2017) to investigate whether participants' responses can be predicted on the basis of the intended clause type by the speaker. 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 intended clause type as a fixed effect factor to examine whether the model was improved. Finally, we ran models that included participants' responses as a dependent variable, the intended clause type as a fixed-effect factor, participants and items as random factors, allowing by-participant and by-item random intercepts, or/and by-participant and by-item random slopes for the intended clause type. Model fit was compared using the likelihood ratio test (Pinheiro & Bates 2000; Bolker, Brooks, Clark, Geange, Poulsen, Stevens & White 2009). The model which included both by-participant and by-item random intercepts, and by-participant and by-item random slopes for the intended clause type was found to perform best ($p < 0.001$). Here, we only consider the outcome of this maximal model. The results showed that the clause type intended by the speaker is a good predictor of participants' responses, see Table 3.

Table 3. Summary of the results of the mixed effects logistic regression between participants' responses and the intended clause type

Estimate β	Std. Error	z-value	p-value
8.57	0.81	10.57	< 0.001

2.5 Interim discussion

The results of the perception experiment indicate that native speakers of Mandarin differentiate *wh*-questions from string identical *wh*-declaratives on the basis of their prosody, and interpret them as *wh*-questions and *wh*-declaratives respectively. Furthermore, the results of the mixed effects logistic regression analysis show that the intended clause type by the speaker is a good predictor for listeners' responses. What remains unclear is whether listeners use early cues like the duration of the Subject to predict the upcoming clause type. This issue was tackled in an audio gating experiment.

3.   Audio-gating experiment

The audio gating experiment aimed at examining whether listeners of Mandarin use early prosodic cues to anticipate the upcoming clause type and addressed the question stated in R2. On the basis of the prosodic properties of the stimuli that were used in the perception experiment discussed in Section 2, we hypothesize that listeners of Mandarin use prosodic cues like duration in the beginning of the sentence to anticipate the upcoming clause type.

3.1 Participants

The participants who took part in the perception experiment discussed in Section 2 participated also in the audio-gating experiment. The audio gating experiment was run first to avoid any bias and to ensure that the prosodic information which was available to the participants was incremental. See Section 2.1 for the profile of the participants.

3.2 Stimuli

On the basis of the 40 stimuli that were used in the perception experiment, we generated three types of audio fragments. This resulted in a total of 120 fragments (40 stimuli×3 fragments). For reasons of naturalness, we decided to respect the

word boundaries and thus cut the fragments at the words' offset (cf. Petrone & Niebuhre 2014 for a similar reasoning). The audio fragments were presented in three gates: Type *a* audio fragment consisted of the Subject and was presented in Gate *a*, Type *b* consisted of the Subject and the Adverb and was presented in Gate *b*, and Type *c* consisted of the Subject, the Adverb, the Verb plus *le* and *diǎnr 'a little',* and was presented in Gate *c*. Note that we included *diǎnr* in Type *c* as *diǎnr* is described as a phonological clitic or a bound morpheme to the previous verb (Shih 1997; Chen 2000). For each gate, we created two types of sentence continuations, *wh*-questions and *wh*-declaratives. The two types of continuation were string identical and differed only with respect to the punctuation. Examples (6)–(8) illustrate the three types of audio fragments that were played to the participants and the corresponding sentence continuations that were presented on screen.

(6) audio Gate *a*   sentence continuations (visually presented)
Táo Wēi   – zuótiān ná-le   diǎnr shénme   gěi   Líu Gāng?
Tao Wei   yesterday bring-PERF a.little what   to/for Liu Gang
                   – zuótiān ná-le   diǎnr shénme   gěi   Líu Gāng.
                      yesterday bring-PERF a.little something to/for Liu Gang

(7) audio Gate *b*
Táo Wēi zuótiān   – ná-le   diǎnr shénme   gěi   Líu Gāng?
Tao Wei yesterday   bring-PERF a.little what   to/for Liu Gang
                         – ná-le   diǎnr shénme   gěi   Líu Gāng.
                            bring-PERF a.little something to/for Liu Gang

(8) audio Gate *c*
Táo Wēi zuótiān ná-le   diǎnr   – shénme   gěi   Líu Gāng?
Tao Wei yesterday bring-PERF a.little   what   to/for Liu Gang
                                  – shénme   gěi   Líu Gāng.
                                     something to/for Liu Gang

3.3 Procedure

The experiment was run using MFC Praat (Boersma & Weenink 2016). Participants were seated in front of a computer screen at an approximate distance of 50 cm and were asked to read silently the instructions that appeared on their screen. Participants had to click on the OK button on the screen to start the experiment. The screen was empty and after 2.0 seconds, the audio fragment was played. 0.5 seconds after the offset of the audio stimulus, two sentence continuations appeared on screen, one on the left and one on the right. Participants were instructed to listen to the audio fragment and then complete the sentence select-

ing one of the two continuations. After selecting one continuation, participants clicked the OK button to submit their response. 2.0 seconds after clicking the OK button, the next audio stimulus was played. The order of the two sentence continuations on the screen were counterbalanced to avoid any left/right preference among participants.

The audio fragments were presented in three consecutive gates, from Gate *a* to Gate *c*, to ensure that the information participants heard was incremental. Therefore, when the 40 audios of Gate *a* were finished, it went to Gate *b* and when the 40 audios of Gate *b* were finished, it went to Gate *c*. The stimuli in each gate were randomized for each participant to avoid a sequence effect. Participants could spend as much time as they wanted for submitting their responses; there was no time pressure. The whole experiment lasted about 20 minutes.

3.4   Results

We obtained a total of 4320 responses (3 gates × 40 stimuli × 36 participants). In general, participants were successful in correctly deciding which of the two clause types was upcoming, see Figure 6. In Gate *a*, where listeners heard only the Subject, they chose a question continuation 54.6% of the time when the fragment originated from a *wh*-question. When the fragment originated from a *wh*-declarative listeners chose a declarative continuation 59% of the times. A chi-square analysis showed that there was a significant association between the clause type intended by the speaker and the participants' responses, $x^2 = 26.73$ (1), $p < 0.001$. In Gate *b*, where participants heard the Subject and the Adverb, they chose a question continuation at 59.7% of the times, when the intended clause type was question, while they chose a declarative continuation at 64.6% when the intended clause type was a *wh*-declarative. Similar to the results of Gate a, a chi-square analysis showed that there was a significant association between the clause type intended by the speaker and the listeners' responses, $x^2 = 85.27$ (1), $p < 0.001$. In Gate *c*, where participants heard the Subject, the Adverb and the Verb-*le diǎnr*, they chose a question continuation 62.1% of the times when the fragmented originated from a *wh*-question. When the fragmented originated from a *wh*-declarative listeners chose for a declarative continuation at 72.1% of the times. A chi-square analysis showed that there was a significant association between the clause type intended by the speaker and participants' responses, $x^2 = 169.80$ (1), $p < 0.001$.

Similar to the perception experiment, in each gate we also ran a mixed effects logistic regression using the *glmer* command in the lme4 package (Bates, Maechler, Bolker & Walker 2015) in R (R Core Team 2017) to investigate whether participants' responses can be predicted on the basis of the intended clause type by the speaker. Specifically, we first ran a null model with participants' responses as a
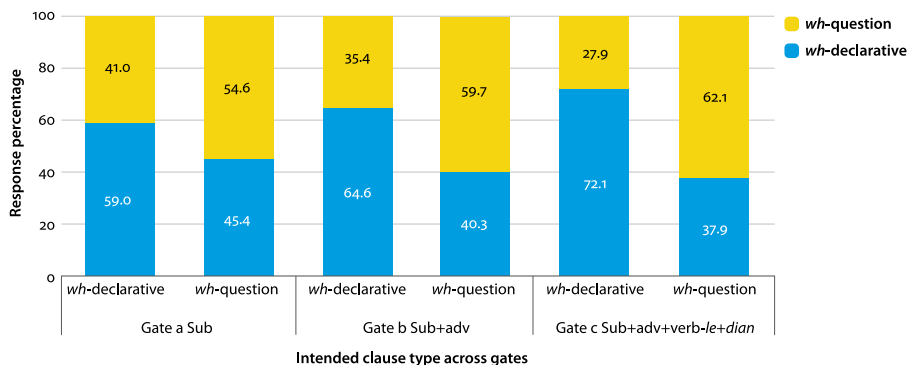
Figure 6. Listener's response in percentage (%) in Gate *a, b* and *c*

dependent variable, and participants and items as random factors. We then ran a model that included in addition the intended clause type as a fixed effect factor to see whether the model was improved. Finally, we ran models that included participants' responses as a dependent variable, the intended clause type as a fixed-effect factor, and participants and items as random factors, allowing by-participant and by-item random intercepts, or/and by-participant and by-item random slopes for the intended clause type. Again, model fit was compared using the likelihood ratio test (Pinheiro & Bates 2000; Bolker, Brooks, Clark, Geange, Poulsen, Stevens & White 2009). The model which included by-participant and by-item random intercepts, and by-item random slopes for the intended clause type was found to perform best in all three gates ($p < 0.001$). The mixed effects results in each gate are summarized in Table 4. In general, participants' response on the clause type in each gate can be predicted on the basis of the intended clause type by the speaker.

Table 4. Summary of the results of the mixed effects logistic regression between participants' responses and the intended clause type in each gate

	Estimate β	Std. Error	z-value	p-value
Gate *a*	0.62	0.30	2.10	< 0.05
Gate *b*	1.51	0.29	4.00	< 0.001
Gate *c*	1.58	0.17	9.35	< 0.001

The response accuracy increases from Gate *a* to Gate *c,* which seems to be consistent with the understanding that the more prosodic information that is heard by listeners, the more accurate their predictions with respect to the upcoming clause type are. In order to test whether this is true, assuming that gates roughly capture the amount of prosodic information, we ran a mixed effects logistic regression again, taking all the responses in all the gates together, different from previous

testing that is limited to each gate. The current mixed effects logistic regression analysis includes gates and clause types as fixed effect factors, and participants and items as random factors. The results show that there is a significant effect of the clause type that was intended by the speaker on the responses of the participants [$\beta = 0.644$, $SE = 0.039$, $p < 0.001$]. However, we did not find a significant effect of gates ($ps > 0.05$), see Table 5 for the mixed effects logistic regression effects of gates.

Table 5. The results of the mixed effects logistic regression taking gates and the intended clause types as predictors; Gate *b* and Gate *c* were compared with Gate *a*

	Estimate β	Std. Error	z-value	p-value
(Intercept)	−0.26	0.08	−3.31	< 0.01
Intended clause type	0.64	0.04	16.19	< 0.01
Gate *b*	0.01	0.05	0.12	> 0.05
Gate c	0.08	0.05	1.65	> 0.05

3.5   Interim discussion

The results of the audio gating experiment show that listeners of Mandarin use prosodic information to predict the upcoming clause type and thus argue for the use of prosody in clause type anticipation. Moreover, the results of Gate *a* suggest that listeners of Mandarin use early prosodic cues like the duration of the Subject to anticipate the upcoming clause type. It is worth noting that the mixed effects logistic regression analysis did show a significant effect of clause types but did not show a significant effect of gates, suggesting that it is not necessarily the case that the more cues available for the participants, the better they predict the clause type. Crucially, it indicates that even from the clause initial, namely at the subject, the clause type anticipation bias is already clear.

4.   General discussion

4.1   Clause type identification and anticipation based on prosody

The aim of this study was to examine the role of prosody in signaling clause type. Specifically, we investigate whether and at which point of the sentence Mandarin listeners can anticipate the clause type of a sentence soly based on prosodic cues, by conducting a case study on *wh*-questions and *wh*-declaratives which are string identical and which can be distinguished only by prosody. We conducted a perception experiment on *wh*-questions and *wh*-declaratives to inves-

tigate whether listeners can identify the right clause types on the basis of their prosodic differences. The results of the perception experiment demonstrated that listeners can identify the clause type correctly by listening to the prosodic cues of the clause type.

The audio-gating experiment directly addresses the question of clause type anticipation by testing different lengths of audio fragments of *wh*-questions and *wh*-declaratives. The results from Gates *a, b* and *c* demonstrate that listeners can make use of limited prosodic cues to anticipate clause types, distinguishing *wh*-questions and *wh*-declaratives before reaching the *wh*-word *shénme*. It should be noted that the fragments of *wh*-questions and *wh*-declaratives were not preceded by any prior contexts; in other words, prosodic information is the sole available cue for clause type anticipation. In Gate *a*, where only the subject of the sentence is heard, there is already a preference for the correct clause type.

We observe that as the gate goes from Gate *a* (where listeners can only hear the subject) to Gate *c* (subject, adverb and the verb plus *diǎnr*), the accuracy in anticipating clause types increases. However, the further statistical analysis taking both the intended clause types by the speaker and gates as predictors only show an effect of the intended clause type; we did not find any effect of gates on the participants' responses. It indicates that participants anticipate clause types based on the prosodic cues, but the clause type prediction itself is not shown to be accumulated.

With respect to clause type and prosody, our findings also suggest that, if clause types are marked prosodically in a language, these specific prosodic markings (i.e. duration and pitch in our study) are actively used as criteria / filters for listeners' identification of clause types (cf. Gérard & Clément 1998). Lastly, our findings also suggest that listeners are able to make an assessment of the limited prosodic cues they hear and anticipate the most plausible clause type or interpretation accordingly based on the existing prosody.

## 4.2   The theoretical implication on clausal typing mechanism

With respect to clausal typing, according to Chomsky and Lasnik (1977: 445), "each clause must be identified as declarative (a declarative or relative clause) or interrogative (a direct or indirect question)." The current study demonstrates the important and early role of prosody in differentiating and classifying clause types and hence provides empirical evidence for an additional clausal typing mechanism, namely, intonation/prosody. That is, prosody can be considered to be a clausal typing mechanism, which can type the clauses in Mandarin, on a par with the fronting of a *wh*-word (in English) and the presence of a *wh*-particle (in Japanese).

We can thus make an extension of the traditional Clausal Typing Hypothesis on *wh*-questions, "Every clause needs to be typed. In the case of typing a *wh*-question, either a *wh*-particle in C° is used or else fronting of a *wh*-word to the Spec of C° is used, thereby typing a clause through C° by spec-head agreement." (Cheng, 1991: 29). Our extended Clausal Typing Hypothesis is given in (9).

(9) Extended clausal typing hypothesis
     Every clause needs to be typed. In the case of typing a *wh*-question, languages can have different means to realize that. Languages can front a *wh*-word to the Spec of CP (*wh*-movement languages), or utilize a *wh*-particle (*wh*-in-situ languages with *wh*-particles), or/and utilize prosody to type the clause.

The above extended hypothesis raises another theoretical question. As prosody is often considered to be realized at PF, the question arises as to how it can type a clause in the narrow syntax, and in particular, before spell-out? Following Cheng and Rooryck (2000), we suggest that an "intonational Q-morpheme" is present in the narrow syntax, which gets "spelled-out" or "mapped" into a certain prosody at PF.[5] We hypothesize that the intonational Q-morpheme is inserted in C°, and has Spec-head agreement with the interrogative operator [+Q] at Spec-CP.

(10)
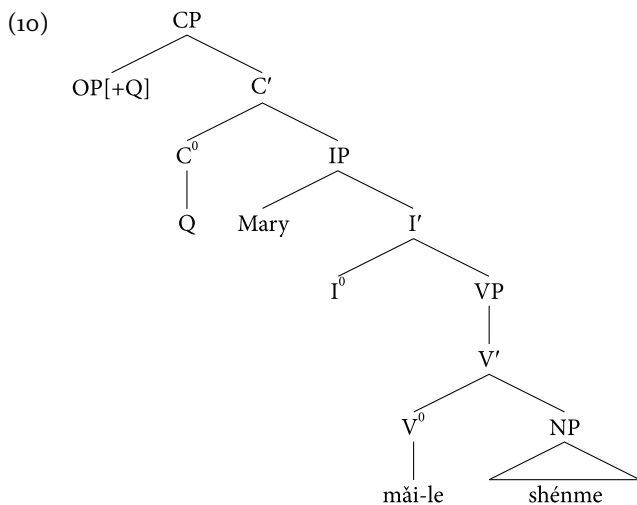
Further studies are needed for the complete theoretical analysis on the intonational Q-morpheme/intoneme Q before spell-out.

---

5. An alternative is to use the term "intoneme Q" following Feng (2015) and Tang (2015). It is unclear whether there is a difference between an intonational morpheme or an intoneme.

4.3   Conclusion

The results of the perception experiment that was presented in this paper show that prosodic cues are used by listeners of Mandarin to interpret the clause type. Moreover, the results of the audio gating experiment demonstrate that listeners of Mandarin use prosodic cues to anticipate the upcoming clause type. Furthermore, the current study sheds light on the clausal typing mechanism in Mandarin, providing evidence for the early and important role of prosody in typing clause types, when no *wh*-movement or *wh*-particle can type the *wh*-question in Mandarin. Based on the empirical results, we propose the Extended clausal typing hypothesis, taking prosody as a clausal typer in Mandarin and hence enriching the original clausal typing hypothesis.

Acknowledgements

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. *R package version* 1.1–7. 2014.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*. 30, 644–663. https://doi.org/10.1016/0749-596X(91)90030-N

Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *The Journal of the Acoustical Society of America*, 99(2), 1148–1160. https://doi.org/10.1121/1.414599

Boersma, P. & Weenink, D. (2016). Praat: doing phonetics computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from http://www.praat.org/

Bolinger, D. (1978). Intonation across languages. *Universals of Human Language*, 2, 471–524.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Braun, B., & Chen, A. (2012). Now for something completely different: Anticipatory effects of intonation. In O. Niebuhr (Ed.), *Understanding Prosody: The Role of Context, Function and Communication* (pp. 289–311). Berlin: de Gruyter. https://doi.org/10.1515/9783110301465.289

Carlson, K., Clifton Jr., C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45(1), 58–81. https://doi.org/10.1006/jmla.2000.2762

Chen, A., den Os, E., & de Ruiter, J. P. (2007): Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*, 24 (2–3), 317–344.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press. https://doi.org/10.1017/CBO9780511486364

Cheng, L. L.-S. (1991). On the Typology of WH-Questions, PhD dissertation, MIT.

Cheng, L. L.-S., & Rooryck, J. (2000). Licensing wh-in-situ. *Syntax*, 3(1), 1–19. https://doi.org/10.1111/1467-9612.00022

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, 8(3), 425–504.

Constant, N. (2014). Contrastive Topic: Meanings and Realizations. PhD dissertation, University of Massachusetts, Amherst.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002): Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*. 47, 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

Dong, H. (2009). Issues in the Semantics of Mandarin Questions. PhD dissertation, Cornell University.

Duanmu, S. (2004). Tone and non-tone languages: An alternative to language typology and parameters. *Language and Linguistics*, 5(4), 891–923.

Face, T. L. (2004). The intonation of absolute interrogatives in Castilian Spanish. *Southwest Journal of Linguistics*, 23(2), 65–80.

Feng, S. (2015). Tone, intonation and sentence final particles in Chinese. *Essays on Linguistics*, 51, 52–79.

Gérard, C., & Clément, J. (1998). The structure and development of French prosodic representations. *Language and Speech*, 41(2), 117–142. https://doi.org/10.1177/002383099804100201

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267–283. https://doi.org/10.3758/BF03204386

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, 21, 501–529. https://doi.org/10.1515/ling.1983.21.3.501

Gryllia, S., Doetjes, J., Yang, Y., & Cheng, L. L.-S. (2019). Prosody, clause typing and wh-in-situ Evidence from Mandarin. Manuscript submitted for publication. Leiden University.

Hermann, E. (1942). *Probleme der Frage*. Göttingen: Vandenhoeck & Ruprecht.

Huang, C. T. J. (1982). Move wh in a language without wh-movement. *Linguistic Review*. 1(4). 369–416. https://doi.org/10.1515/tlir.1982.1.4.369

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2), 153–194. https://doi.org/10.1006/jmla.1998.2620

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (Lmer Objects of lme4 Package). R Package Version 2.0–3 [Computer Software].

Li, A. (2002). Chinese prosody and prosodic labeling of spontaneous speech. In *Proceedings of Speech Prosody 2001*.

Li, B. (2006). Chinese Final Particles and the Syntax of the Periphery. PhD dissertation, Leiden University.

Li, Y. H. A. (1992). Indefinite wh in Mandarin Chinese. *Journal of East Asian Linguistics*, 1(2), 125–155. https://doi.org/10.1007/BF00130234

Lin, J. W. (1998). On Existential Polarity-WH-Phrases in Chinese. *Journal of East Asian Linguistics*, 7(3), 219–255. https://doi.org/10.1023/A:1008284513325

Lindsey, G. A. (1985). Intonation and Interrogation: Tonal Structure and the Expression of a Pragmatic Function in English and other Languages. PhD dissertation, UMI.

Liu, X., Li, A., & Jia, Y. (2016). How does prosody distinguish wh-statement from wh-question? A case study of standard Chinese. *Proceedings of Speech Prosody* 2016, 1076–1080. https://doi.org/10.21437/SpeechProsody.2016-221

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40 (1), 1–18. https://doi.org/10.1159/000261678

Ohala, J. J. (1984). An ethological perspective on common cross-Language utilization of Fo of Voice. *Phonetica*, 41(1), 1–16. https://doi.org/10.1159/000261706

Petrone, C., & D'Imperio, M. (2011). From tones to tunes: Effects of the Fo prenuclear region in the perception of Neapolitan statements and questions. In *Prosodic Categories: Production, Perception and Comprehension* (pp. 207–230). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0137-3_9

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, 57(1), 108–146. https://doi.org/10.1177/0023830913495651

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects Models in S and S-Plus*, 3–56. https://doi.org/10.1007/978-1-4419-0318-1_1

R Core Team (2017). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, 20, 81–124.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48(1), 103–130. https://doi.org/10.1016/S0749-596X(02)00519-3

Staub, A., & Clifton Jr., C. (2006). Syntactic prediction in language comprehension: Evidence from either… or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425.

Swerts, M., Collier, R., & Terken, J. (1994). Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, 15(1–2), 79–90. https://doi.org/10.1016/0167-6393(94)90043-4

Tang, S. W. (2015). A generalized syntactic schema for utterance particles in Chinese. *Lingua Sinica*, 1(1), 1–23. https://doi.org/10.1186/s40655-015-0005-5

Terken, J., & Hirschberg, J. (1994): Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position. *Language and Speech*, 37(2), 125–145. https://doi.org/10.1177/002383099403700202

Tsai, W. T. D. (2010). On the Syntax-Semantics Correspondences of Chinese Modals. *Zhongguo Yuwen*, 3, 208–221.

Van de Weijer, J., & Sloos, M. (2014). The four tones of Mandarin Chinese: Representation and acquisition. *Linguistics in the Netherlands*, 31(1), 180–191. https://doi.org/10.1075/avt.31.13wei

Van Heuven, V. J., & Van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, 47(1–2), 87–99. https://doi.org/10.1016/j.specom.2005.05.010

Vion, M., & Colas, A. (2006). Pitch cues for the recognition of yes-no questions in French. *Journal of Psycholinguistic Research*, 35(5), 427–445. https://doi.org/10.1007/s10936-006-9023-x

Wakefield, J. C. (2016). Sentence-final particles and intonation: two forms of the same thing. In *Proceedings of Speech Prosody 2016*. https://doi.org/10.21437/SpeechProsody.2016-179

Weber, A., Braun, B., & Crocker, M. (2006): Finding referents in time: eyetracking evidence for the role of contrastive accents. *Language and Speech*, 49(3), 367–392. https://doi.org/10.1177/00238309060490030301

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, 58(1–2), 26–52. https://doi.org/10.1159/000028487

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4), 319–337. https://doi.org/10.1016/S0167-6393(00)00063-7

Yang, Y. (2018). *The two sides of wh-indeterminates in Mandarin – a prosodic and processing account*. LOT Publication: the Netherlands.

Yang, Y., Gryllia, S., & Cheng, L. L.-S. (2019). *Wh-question or wh-declarative? Prosody makes the difference*. Manuscript submitted for publication.

Yuan, J. (2004). Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling. Unpublished doctoral thesis, Cornell University, NY.

Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In *Chinese Spoken Language Processing* (pp. 19–30). Springer Berlin Heidelberg. https://doi.org/10.1007/11939993_7

Zhang, L. (2014). Segmentless sentence-final particles in Cantonese: An experimental study. *Studies in Chinese Linguistics*, 35(2), 47–60.

Address for correspondence

Yang Yang
Center for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou, 510420
China

yangyanggw@gdufs.edu.cn

Co-author information

Stella Gryllia
Leiden University Centre for Linguistics
s.gryllia@hum.leidenuniv.nl

Leticia Pablos
Leiden Institute for Brain and Cognition
Leiden University Centre for Linguistics
l.pablos.robles@hum.leidenuniv.nl

Lisa Lai-Shen Cheng
Leiden Institute for Brain and Cognition
Leiden University Centre for Linguistics
l.l.cheng@hum.leidenuniv.nl