Open Library of Humanities

No prosody-syntax trade-offs: Prosody marks focus in Mandarin cleft constructions

Anja Arnhold, Department of Linguistics, University of Alberta, Edmonton, Alberta, CA, arnhold@ualberta.ca

In line with the idea that language has evolved to be efficient and to avoid redundancy, syntactic means of marking information structure have been derived from prosodic ones, and vice versa, for many languages. On the basis of crosslinguistic comparisons, prosody-syntax trade-offs have frequently been described for clefts. The present study investigated whether such trade-offs can also be observed language-internally, testing whether clefting reduced prosodic focus marking in production or its effects on perception in Mandarin. A production study found that clefts showed prosodic focus marking equal to or exceeding that found in syntactically unmarked equivalents. In both syntactic conditions, focused constituents had larger f0 ranges, higher f0 maxima and longer durations compared to a broad focus baseline, while post-focal constituents showed lower f0 maxima and minima, lower intensity and, for clefts, shorter durations (28 participants, 937 utterances containing 4466 syllables analyzed in total). A rating study likewise found that the effect of prosody on the perception of information structure was not modulated by clefting, which neither affected ratings nor reaction times (102 participants, 2448 responses analyzed in total). These findings suggest that prosody is integral for marking focus in cleft constructions instead of constituting a redundant cue.

1. Introduction

Language is commonly assumed to be efficient, based on the idea that it has evolved to be optimally suited for human communication (overviews in Coupé et al., 2019; Fedzechkina, 2014; Gibson et al., 2019; Haspelmath, 2021; Hawkins, 2014). Many authors argue that efficient communication avoids redundancy, even though this is somewhat controversial and the degree of redundancy found may depend on the population studied (e.g., children vs. adults, Tal & Arnon, 2022) and the definition of redundancy (e.g., compare Aylett & Turk, 2004; Beekhuizen et al., 2013; Levshina, 2021). This article concentrates on redundancy in the sense of a language producer using more signals than necessary for the perceiver to understand the intended message, i.e., the use of multiple linguistic cues to the same meaning (see Regier et al., 2015, for an overview of the literature on the trade-off between the producer's and perceiver's needs). In addition to the language producer and perceiver, linguists have also described this issue from the point of view of the linguistic system, which may be considered inefficient when maintaining unnecessary complexity (Mollica et al., 2021; Regier et al., 2015).

Typological surveys often show tendencies to avoid redundancy and unnecessary complexity. For example, many studies have found a significant negative correlation between a language having fixed word order and it having case, two means of encoding core arguments (Koplenig et al., 2017; Siewierska, 1998; Sinnemäki, 2008, 2010, 2014), and this correlation is also supported by evidence from artificial language learning (Fedzechkina, 2014; Fedzechkina et al., 2017; Roberts & Fedzechkina, 2018; but see Levshina, 2021). The marking of speech acts also shows evidence of trade-offs; for example, prosodic marking of interrogativity is weaker in *wh*-questions, which differ from statements morpho-syntactically (e.g., *When are you going?*), than in declarative questions, where prosody is the only marker of interrogativity (e.g., *You are going?*), both phonologically and phonetically (see overview in van Heuven, 2017a). This has been explained both via a functional hypothesis and in terms of generative syntax (Haan, 2001; van Heuven, 2017b; De Clercq, 2017).

Overall, whether languages systematically avoid redundant encoding and instead show trade-offs between or within their linguistic components has been debated controversially, with empirical studies showing mixed results (Fenk-Oczlon & Pilz, 2021; Maddieson, 2005; McWhorter, 2001; Nichols, 2009; Pimentel et al., 2020; Shosted, 2006; Yadav et al., 2020). However, research suggests that trade-offs are more likely to appear between linguistic components, such as phonology or morphology, than within them (Fenk-Oczlon & Fenk, 2008), as well as being more likely to appear within one functional domain, such as core argument marking, rather than across unrelated linguistic features (Miestamo, 2008; Sinnemäki, 2008, 2014).

Heeding these suggestions, the present contribution will look for trade-offs between prosody and morpho-syntax in focus marking, a functional domain that has not previously featured prominently in the literature on language efficiency, redundancy and complexity (except for the

central role of prominence in the smooth signal redundancy hypothesis; see Aylett & Turk, 2004). It will do so by studying the role of prosody in the production and perception of cleft constructions in Mandarin Chinese. In a nutshell, the idea is as follows: Since clefts are already morpho-syntactically marked, additional prosodic focus marking is potentially redundant. Therefore, it could be hypothesized that prosodic focus marking will be less extensive in production and less effective in perception in clefts compared to syntactically unmarked equivalents. This would be an instance of a trade-off between linguistic components.

1.1 Prosody-syntax interactions in focus marking and clefts

The literature on information structure frequently describes interactions between prosody and morpho-syntax in its marking. In line with the assumption that the language system is efficient (Reinhart, 2006), researchers often derive prosodic focus marking from morpho-syntactic focus marking or vice versa (e.g., Büring, 2010; Büring & Gutiérrez-Bravo, 2002; Féry et al., 2007; Hamlaoui, 2007; Szendrői, 2017; Vallduví & Vilkuna, 1998; Vander Klok et al., 2018). For example, Kratzer and Selkirk (2020) describe prosodic phrasing as influenced by a syntactic focus feature. Conversely, Féry (2013) motivates the prosodic, syntactic and morphological focus marking strategies of various languages as a universal preference for focus to be aligned with the edge of a prosodic phrase. Samek-Lodovici (2005) suggests that languages differ in whether they preferentially use prosody or syntax for marking information structure, depending on whether syntactic or prosodic constraints are ranked higher.

While these accounts concentrate on modelling the availability of different means of focus marking–whether or not a language is able to use changes in the position of accents or constituent order at all–cross-linguistic differences in the *frequency* with which means of focus marking are used have mostly been discussed with respect to cleft constructions. Clefts, as illustrated in (1) below, can mark the clefted constituent as the focus, as indicated by the square brackets, and the rest of the sentence as background (although they do not have to do so, this is often considered their prototypical use, see Akmajian, 1970; Atlas & Levinson, 1981; Cassarà et al., 2022; Chomsky, 1971; Hedberg, 2000; Jackendoff, 1972; Karssenberg et al., 2019; Kiss, 1999; Lambrecht, 2001; Rochemont, 1986). As already suggested by Jespersen (1937, p. 85), it is often observed that clefts are more frequent in languages with fixed constituent order, such as English and French, than in languages which can use their flexible constituent order to mark focus, such as German and Italian (de Cesare, 2014; Di Tullio, 2006; Gundel, 2008; Lambrecht, 2001; Skopeteas & Fanselow, 2010). Moreover, languages with less flexible prosody, like French and Spanish, use clefts more often than languages like English, whose flexible prosody includes a large inventory of pitch-accents that can be placed variably within a sentence to express pragmatic contrasts (Gundel, 2006, 2008; Lambrecht, 2001; Sánchez-Alvarado, 2020; Skopeteas & Fanselow, 2010; van Valin & LaPolla, 1997; Wehr, 2005; but see Dufter, 2009 and Gundel, 2006, 2008, who argue

that this cannot completely explain cross-linguistic differences in cleft frequency). Similarly, clefts are more frequent in written language, where prosodic focus marking is unavailable, than in spoken language (Collins, 1991; Declerck, 1988, p. 226; Tönnis et al., 2016).[1]

(1)   It is [my aunt]$_F$ who is steaming mushrooms.

A question that has not been considered with respect to clefts and prosody-syntax interactions is whether using a cleft together with prosodic focus marking constitutes a redundancy that speakers try to avoid or attenuate. Since clefting can already mark the information structure by itself (as it does in written language), additional prosodic marking may be redundant.[2] If speakers avoid redundancy, they could be expected to either not use prosodic focus marking in clefts or to use it less extensively than in morpho-syntactically unmarked utterances, where prosody is the only means of marking information structure. Whether this kind of prosody-syntax trade-off can indeed be observed has not been investigated previously (with the exception of Arnhold, 2021).

Instead, several authors have used observations about the prosodic realization of clefts as arguments for particular syntactic, information structural or discourse analyses, frequently based on corpora of spontaneous speech (Cassarà et al., 2022; Collins, 2006; Delin, 1995; Frascarelli & Ramaglia, 2013; Geluykens, 1984; Hedberg, 1990; Huber, 2006; Pinelli et al., 2020; Prince, 1978; Van Praet & O'Grady, 2018). Experimental studies on prosody-syntax interactions in clefts have mostly focused on cases where prosody and morpho-syntax do not align to mark the clefted constituent as focused: A series of experiments by Calhoun, Yan and colleagues tested cases where prosody and morpho-syntactic focus marking are in conflict and compared them to equivalents where they do align or where only prosodic marking is present, finding that languages differ in whether prosody or morpho-syntax takes precedence (Calhoun et al., 2021; Yan et al., 2020; Yan & Calhoun, 2019, 2020). Greif and Skopeteas (2021) tested if the ability of four languages to prosodically mark object focus in subject cleft sentences predicts the acceptability of subject clefts with object focus in a rating task with written stimuli. They found that such

---

[1] This is true of *it*-clefts as illustrated in (1), which are usually referred to simply as 'clefts' and taken to be the equivalent of the Mandarin clefts investigated here. So-called pseudo-clefts (e.g., *What my aunt steamed was mushrooms*) are more common in spoken than in written English (Collins, 1991). Of course, it is not that prosody plays no role in reading, since the reader will supply implicit prosody (e.g., Jun & Bishop, 2015), but the writer cannot determine prosody the way a speaker or signer can. Note also that differences in the frequency of clefts in written vs. spoken language have, to my knowledge, not yet been discussed for any language outside the Indo-European family. However, for written language, Lee (2005) observes that *shi…(de)* cleft constructions were much more frequent in the Chinese translation of a novel than clefts were in the English original. This could indicate that the generalization of more frequent cleft use in written than spoken language also holds for Mandarin.

[2] Note that information structure marking is not the only function of clefts. For example, much recent work on clefts has been devoted to exhaustivity (see overview in Onea, 2019). This will be revisited in the discussion. For now, it suffices to say that as long as it can be assumed that clefts mark information structure, even if that is not their only or even primary function, additional prosodic information structure marking could be argued to be redundant.

clefts with focus following the clefted constituent (e.g., *It is John that bought the [bicycle]$_F$*) are more acceptable in English and German, which have flexible nuclear-accent placement, than in French and Mandarin. Finally, Pinelli et al. (2020) showed very similar prosodic realizations for two syntactically different cleft types in Italian. More in line with the present study's concerns is the approach taken by Kember et al. (2019), who tested participants' memory of words that were clefted, made prominent prosodically, both or neither. For Korean listeners, the authors found a stronger memory advantage of clefting than of prosodic marking, while for English, both effects were equivalent and additive if combined, i.e., there were no trade-offs between prosody and syntax (also see the rating study in Arnhold, 2021). In contrast, Blything et al. (2021) found no additional effect of clefting on pronoun resolution in English once the effect of prosodic focus marking was accounted for.

With the exception of the most recent studies, research considering the prosody of clefts has entirely concentrated on Germanic and Romance languages.

1.2 Mandarin clefts and prosodic focus marking

Mandarin clefts, see (2), below, are also called *shi...de* constructions, since they contain the copula 是 *shi4* and the toneless syllable 的 *de0*, which are not present in syntactically unmarked equivalents; compare (2) to (3). These constructions occur in different variants, with *shi4* and *de0* both being able to appear by themselves, whose syntactic and semantic analysis has been debated (Cheng, 2008; Hole, 2011; Hole & Zimmermann, 2013; Lee, 2005; Liu & Shi, 2022; Paul & Whitman, 2008; Simpson & Wu, 2002; Xie, 2012). As indicated in (2), in transitive cleft sentences *de0* can appear either between the verb and the object or sentence-finally, conditioned by dialect. Speakers of southern Mandarin dialects exclusively place *de0* sentence-finally, whereas in northern dialects, both placements are possible, but only verb-adjacent *de0* is compatible with a narrow focus interpretation (Paul & Whitman, 2008). Moreover, Simpson and Wu argue that *de0* is undergoing a re-analysis as a past tense morpheme in northern dialects (also see Lee, 2005). For these reasons, clefts without *de0*, also called bare *shi* constructions, were used in the studies presented here (as also done in Greif & Skopeteas, 2021; Liu & Yang, 2016).

(2) 是　　　姑妈　　蒸　　　(的)　　冬菇　　　(的)。
　　　shi4　　gu1ma1　zheng1　(de0)　　dong1gu1　(de0)
　　　COPULA　aunt　　steam　　(DE)　　mushroom　(DE)
　　　'It was aunt who steamed mushrooms.'

(3) 姑妈　　蒸　　　冬菇。
　　　gu1ma1　zheng1　dong1gu1
　　　aunt　　steam　　mushroom
　　　'Aunt steams/steamed mushrooms.'

The prosodic realization of Mandarin clefts is not very well researched. The literature discussing semantic and syntactic analysis contains some remarks about possible locations of prominence, see especially Paul and Whitman (2008). In particular, Paul and Whitman state that in bare *shi* constructions, the clefted subject is only focused when it is made prominent prosodically; otherwise, "the truth of the entire sentence is strongly asserted, with a meaning comparable to 'it is (really) that S' or 'it is because S'" (p. 424). The latter meaning seems to correspond to verum focus, i.e., focus on the truth of the proposition (Höhle, 1992; Lohnstein, 2016). Liu and Shi (2022) also show that Mandarin clefts can convey verum focus, but state that in these cases, the copula *shi4* "receives a focal pitch accent" (p. 4). Both descriptions are based on speaker intuition and provide no further discussion of the acoustic correlates of prominence in either subject-focus or verum focus-clefts. Note, however, that Liu and Shi's description parallels verum focus prosody in many other languages, e.g., English: *My aunt IS steaming mushrooms* (with capitals marking the location of nuclear accent; also see e.g., Han & Romero, 2014; Féry & Arnhold, 2019; and Gutzmann et al., 2020, on prosody of verum focus in various languages). Crucially, verum focus notably differs from broad focus, to which subject focus is compared in the present study, also semantically: Broad focus is focus on the entire proposition, whereas verum focus is focus specifically on its truth value. As the present experiments systematically induced either subject focus or broad focus, but not verum focus, the potential verum focus reading of Mandarin clefts should not be relevant here, but we will return to this issue in discussing the results of the perception study (Section 3.3).

There are no corpus studies considering prosody of Mandarin clefts and only one production experiment published so far: In Greif and Skopeteas (2021), 16 speakers produced four transitive sentence items in four conditions each: unmarked syntax with subject focus, subject cleft with subject focus, unmarked syntax with object focus, subject cleft with object focus (256 utterances total). Statistical analysis evaluated f0 measurements, looking for interactions between the factors focus (subject vs. object focus), construction (subject cleft vs. unmarked) and time (multiple measurements per syllable). It found an interaction between focus and time, indicating more rapid f0 rises for focused constituents than for pre- or post-focal ones (most target syllables carried the rising lexical tone 2). It also found an interaction between construction and time. The authors explain this as a delay in the implementation of tonal targets on subjects when preceded by *shi4* rather than suggesting a difference in prosodic focus marking between clefts and unmarked syntax.

Other acoustic measures were not analyzed, but the f0 results are in line with previous research on (Mainland) Mandarin prosodic focus marking, finding expanded f0 ranges and raised f0 on focused constituents, as well as post-focal compression (Y. Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; B. Wang et al., 2017; B. Wang & Xu, 2011; T. Wang et al., 2020;

Xu, 1999). Previous research also found focused constituents to have higher intensity and longer durations compared to pre- and especially post-focal constituents (Y. Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; B. Wang et al., 2017; T. Wang et al., 2020; Xu, 1999), but this has not been confirmed with a production study for clefts yet. In their perception studies, Yan and colleagues used stimuli produced by a single speaker, which included higher mean f0, larger f0 range, longer duration and higher intensity for focused constituents than for pre- and post-focal ones (Yan et al., 2020; Yan & Calhoun, 2019, 2020).[3] Like Greif and Skopeteas (2021), they did not include a broad focus baseline. Finally, there is evidence that voice quality plays a role in Mandarin focus marking. Chen and Gussenhoven report that some of their participants produced more creak for the low-dipping tone 3 in focus than pre-focally (creak is associated with low f0 and, thereby, with tone 3 more than with the other lexical tones, e.g., Huang, 2020; Kuang, 2017). Similarly, Cao and Zhang (2008) found that creaky voice contributes to the perception of tone 3 words as focused. Zheng (2006) showed that focus leads to more realizations with non-modal voice quality, but only for tones 2 and 3, based on visual inspection of waveform and spectrogram. Likewise using manual classification, Huang et al. (2018) found an increase in creaky voice only for tone 4. However, this finding did not replicate with the acoustic measures they analyzed (harmonic-to-noise ratio, cepstral peak prominence and two spectral tilt measures). Wang et al. likewise did not find an effect of focus on the two phonation measures they investigated—cepstral peak prominence and H1–H2—for any of the four lexical tones. None of these studies on voice quality included cleft sentences or other syntactic focus marking.

Regarding the role of prosody in the perception of Mandarin clefts, studies by Yan and colleagues showed that when prosody conflicts with syntax, as in a subject cleft with prosodic focus marking on the object, prosody wins: Prosodic focus marking, not clefting, determines appropriateness ratings of answers to questions inducing either subject or object focus (Yan et al., 2020). Similarly, only prosodic focus marking is able to prime alternatives (Yan & Calhoun, 2019) and facilitate the rejection of false alternatives (Yan & Calhoun, 2020; also see S. H. Chen et al., 2012). Interestingly, clefting the focused constituent never improved appropriateness ratings compared to the unmarked syntax equivalent when the focused constituent was marked prosodically. However, clefting a background constituent lowered ratings (Yan et al., 2020). This is only partially in line with the findings of Greif and Skopeteas (2021). Their participants also rated subject clefts lower than unmarked syntax in object focus, i.e., where the clefted constituent was part of the background. However, in subject focus, subject clefts received higher ratings than unmarked equivalents. This difference between the

---

[3] The experiment reported in Yan et al. (2020) is also published in Yan (2020) and Yan et al. (2022).

studies is likely due to Greif and Skopeteas using written stimuli. As they point out, subjects in unmarked SVO sentences are typically interpreted as topics in Mandarin, in line with a cross-linguistic tendency which explains why non-minimal marking, such as clefts, is used more often for subject than for non-subject focus. It is therefore likely that their participants applied the default topic-comment interpretation to the unmarked sentences, which conflicts with subject focus contexts, whereas prosodic focus marking on the subject would have prevented this default interpretation in the studies by Yan and colleagues. Again, none of these studies used a broad focus baseline.

1.3 Hypotheses

The present study will investigate cleft sentences and syntactically unmarked equivalents in Mandarin to test for syntax-prosody trade-offs in focus marking. In line with the common assumption that language is efficient and avoids redundancy, such trade-offs are hypothesized to occur in both production and perception. With respect to production, the hypothesis is that prosodic focus marking will be less pervasive in clefts than in syntactically unmarked equivalents (vs. simultaneous and additive use of prosodic and syntactic focus marking that should occur in the absence of trade-offs). In other words, the effect of focus on acoustic measures associated with prosodic focus marking is hypothesized to either be present only with unmarked syntax (strong version) or be smaller in clefts than with unmarked syntax (weak version), see (4), below. The strong version of the hypothesis could be predicted based on generative accounts deriving prosodic focus marking from syntactic focus marking, or vice versa–based on the assumption that grammar avoids redundancy, and that deriving one from the other is more efficient than computing prosodic and syntactic structures with focus marking independently and then having to relate them to each other (as spelled out explicitly e.g., by Reinhart, 2006). A functional approach could derive both the strong and the weak version of the hypothesis, paralleling van Heuven's (2017a,b) discussion of phonological (categorical) and phonetic (gradual) differences between syntactic sentence types.

(4)   **Hypothesis for production experiment**
    a. <u>Strong version: Prosodic focus marking is absent in clefts</u>. Compared to a broad focus baseline, in subject focus only sentences with unmarked syntax show significant prosodic focus marking. Cleft sentences with subject focus do not show significant prosodic differences from the broad focus baseline.
        (Subject focus with unmarked syntax > subject focus with clefts, broad focus)
    b. <u>Weak version: Prosodic focus marking is weaker in clefts than with unmarked syntax.</u> Both clefts and sentences with unmarked syntax differ significantly from the broad focus baseline, but also from each other. Sentences with unmarked syntax show significantly stronger prosodic focus marking than cleft sentences.
        (Subject focus with unmarked syntax > subject focus with clefts > broad focus)

Both versions of the hypothesis will be tested in a production experiment measuring f0 range, f0 maximum, f0 minimum, duration, intensity and the use of non-modal voice quality (operationalized as a binary variable based on manual annotation following Zheng, 2006, and Huang et al., 2018, the only previous production studies of Mandarin finding an effect of focus on voice quality). Based on the existing research summarized in Section 1.2, prosodic focus marking can maximally be expected to manifest as subject focus showing larger f0 range, higher f0 maximum and minimum, longer duration, higher intensity and more non-modal voice quality on the narrow focus subject (focus expansion) and conversely a smaller f0 range, lower f0 maximum and minimum, shorter duration, lower intensity and less non-modal voice quality on the verb and object (post-focal reduction of prominence) than in broad focus. The hypothesis in (4) can reasonably be considered confirmed if at least half of the evaluated measures, i.e., three out of six, show the expected prosodic focus marking in terms of focus expansion, post-focal reduction or both. Note also that the hypothesis is more tentative for voice quality, since most previous studies have shown effects of focus on voice quality only for words with tones 3 and 2 (except Huang et al., see Section 1.2), whereas the present study included only tones 1 and 4 (see Section 2.1.2).

A second experiment will test the corresponding hypothesis of trade-offs in perception, see (5), below, by asking participants to rate the fit between a question inducing subject focus or broad focus on the one hand and, on the other hand, an answer with subject focus marked via prosody, clefting, both or neither. The hypothesis assumes that participants will rate the answer to a question as more appropriate when they judge it to have suitable focus marking, whereas they will give lower ratings in the absence of the focus marking they expected. Similarly, an absence of suitable focus marking could lead to processing delays, resulting in longer reaction times. The strong version of the trade-off hypothesis predicts that when prosodic and syntactic focus marking are combined, either one of them is redundant and will therefore have no effect. Thus, either prosodic or syntactic focus marking is enough to make an answer fit the context, and adding a second type of concurrent marking will not make a difference. The weak version of the hypothesis predicts that the redundant effect is still significant, but also significantly smaller than when information structure is marked only by either prosody or syntax. That is, using two concurrent types of focus marking has a stronger effect than using only one type of marking, but this combined effect is not as strong as simply adding together the effect of both occurring separately.

(5)   **Hypothesis for perception experiment**
    a. <u>Strong version</u>: There is a trade-off between prosodic and syntactic marking of subject focus, such that either one needs to be present, but when both are present, only one of them has an effect.
    b. <u>Weak version</u>: There is a trade-off between prosodic and syntactic marking of subject focus, such that when both are present, one of them is reduced.

The hypothesis of trade-offs as presented in (5) thus contrasts with the alternative hypothesis that prosodic and syntactic focus marking are additive, which would predict that the combined effect of concurrent syntactic and prosodic focus marking would be at least the same as adding together the individual effects of prosodic focus marking and clefting. For example, in subject focus contexts, the use of subject focus prosody alone should increase ratings and speed up reaction times by a certain amount x compared to the use of broad focus prosody (unmarked syntax with subject focus prosody vs. unmarked syntax with broad focus prosody), while the use of subject clefts by itself should increase ratings and speed up reaction times by a certain amount y (clefts with broad focus prosody vs. unmarked syntax with broad focus prosody). Under the assumption that prosodic and syntactic focus marking are additive, their combined effect in terms of improvement in ratings and reduced reaction times (clefts with subject focus prosody vs. unmarked syntax with broad focus prosody) should correspond to at least $x+y$, whereas the hypothesis of trade-offs as given in (5) predicts that the combined effect is smaller than $x+y$.

2. Production experiment

2.1 Methods

2.2.1 Participants

Twenty-nine native speakers of Mandarin, all undergraduate students at the University of Alberta, participated. One participant's data was discarded because she did not follow the instructions. Results are based on the remaining 28 participants (24 female, 4 male; age 18–25, mean 20.71; for further information see supplementary materials, section 7).

All participants received partial course credit for an introduction to linguistics course as compensation. The study was approved by the Research Ethics Board 2 of the University of Alberta (study ID Pro00069812).

2.1.2 Materials

The experimental materials contained 24 target sentences in two syntactic conditions, as clefts and in unmarked syntax. Target sentences were presented as answers in question-answer pairs, each preceded by a short context. Context and question induced one of two information structure conditions: narrow focus on the subject or broad focus (see **Table 1** for an example). Note that the factors syntax (unmarked vs. cleft) and information structure (broad focus vs. subject focus) were not fully crossed, since participants were not asked to produce clefts in broad focus contexts to avoid unnatural productions. Thus, there were three conditions: unmarked-broad focus, unmarked-subject focus and cleft-subject focus.

Materials were constructed with a native speaker so that half of the target sentences contained only syllables carrying lexical tone 1 (high), except for the copula *shi4* in cleft conditions, while the other half of the target sentences contained only syllables carrying lexical tone 4 (falling), since f0 effects of information structure marking can be subtle compared to differences between lexical tones (B. Wang & Xu, 2011; T. Wang et al., 2020). Tones 1 and 4 were chosen to represent both level and contour tones. Further, since tones 2 and 3 were avoided, any observed effects on voice quality will be clearly attributable to focus (recall Section 1.2). To achieve this, target words with a length of one to three syllables were chosen; note that no three-syllable words appeared for tone 4. A pilot study checked the acceptability of the experimental materials (see Supplementary Materials, Section 1, for a full list of target sentences, example contexts and further details).

	Broad focus	**Subject focus**
Context	'It's Friday night, most of your coworkers are going home to enjoy the weekend. You are still waiting for your friend to pick you up because you guys decided to go to a club. When your friend comes to pick you up, she is surprised that you don't switch off the lights and lock the door when the two of you leave the office. Your friend asks':	'You are the secretary in a small law firm with two lawyers, Lijie and Fangwen. Your boss received a call from one of their clients. He will be stopping by later today to drop off his documents for his case. Your boss wanted someone to handle this because he cannot himself. He knows that someone is staying late tonight, so he asks you':
Question	'Friend: Why aren't you locking up?'	'Boss: Who is working overtime?'
Target (unmarked)	方温加班。 fang1wen1 jia1 ban1. Fangwen add shift. 'Fangwen is working overtime.'	方温加班。 fang1wen1 jia1 ban1. Fangwen add shift. 'Fangwen is working overtime.'
Target (cleft)		是方温加班。 shi4 fang1wen1 jia1 ban1. COP Fangwen add shift. 'It is Fangwen who is working overtime.'

Table 1: Example of target sentence, with preceding context and question, in all conditions.

The experimental materials (24 items * 3 conditions = 72 target sentences with contexts) were distributed onto four lists, so that each item appeared once on two of the lists and twice on the other two (in two different blocks), and the three experimental conditions each appeared twelve times per list. These 36 experimental trials per list were combined with 15 filler trials. Each list was preceded by two practice trials. Block order and trial order within blocks were randomized individually for each participant.

2.1.3 Procedure

Participants were recorded individually in a sound-attenuated booth. During each trial, contexts, questions and target sentences appeared in writing on a computer screen. Participants read the contexts and questions silently before pressing a button to see the answer, which they then spoke aloud. They were told to imagine themselves as being in the described context and to pronounce the sentence as if they were really in this situation.

Participants were recorded with a Countryman headset microphone (H6 Omni) placed about four centimeters from their mouths and a Fostex field recorder (Model FR-2LE) at 16-bit resolution and a sampling frequency of 44,100 Hz. Stimuli were presented and button presses recorded with E-Prime (Psychology Software Tools, 2016).

2.1.4 Data editing and measurements

All recorded utterances were manually segmented in Praat (Boersma & Weenink, 2020) based on spectrograms and waveforms, as well as auditory cues. Boundaries were set based on the following cues, where available: silent intervals, fricative noise, the third formant and the second formant. Automatically generated pitch objects were inspected to remove measurement errors such as octave jumps or measurements during obstruents, as well as removing measurements during the first few cycles at the beginning and end of voiced intervals to minimize microprosodic effects. A script then automatically identified and marked the highest (f0 maximum) and lowest (f0 minimum) points for each syllable; these were likewise checked manually. Stretches of speech with non-modal voice quality (mostly creaky or breathy) were also marked manually based on auditory and visual inspection of waveforms and spectrograms; see **Figure 1** for illustration. Manual annotation was chosen over acoustic measurements to obtain a single dependent variable that corresponds to human perception and to ensure comparability with the two existing production experiments that found focus effects on voice quality Mandarin (Zheng, 2006; Huang et al., 2018).

A Praat script (based on Arnhold, 2018) then measured the following acoustic variables for each syllable in the dataset: (1) f0 range, calculated as f0 maximum – f0 minimum; (2) f0 maximum and (3) f0 minimum, both measured in Hz and converted to st relative to a baseline

of 100 Hz; (4) duration, in ms; (5) mean intensity, in dB, during the center 50% duration of the vocalic nucleus. Before measuring, intensity of all utterances was scaled to 50dB with a Praat script (Vicenik, n.d.).
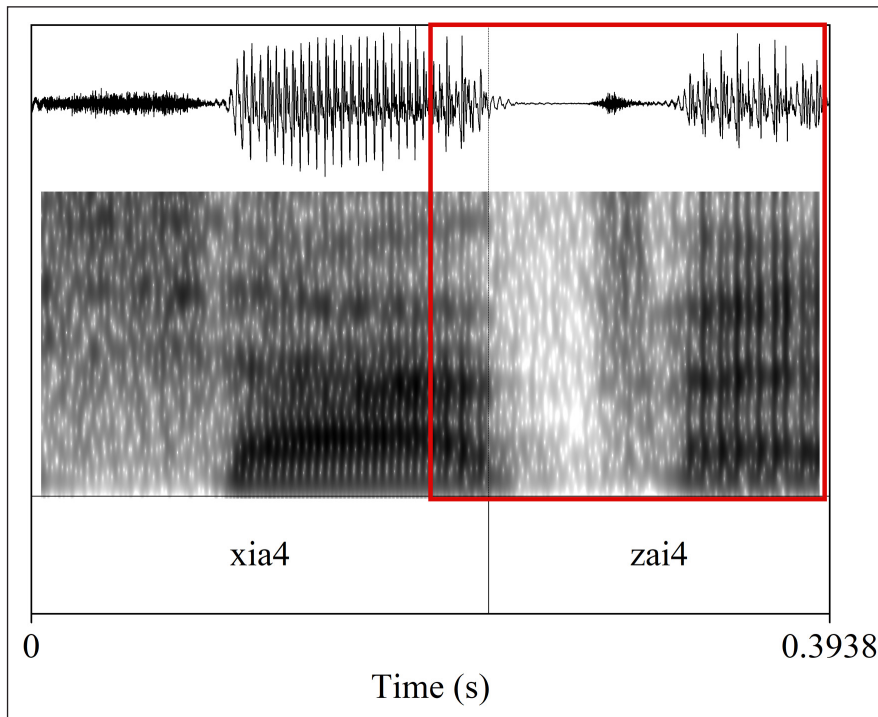
Figure 1: Illustration of manual annotation of non-modal voice quality, verb 下载 'download' (from item 20, realized by participant s03 in cleft condition). The interval annotated as non-modal, starting at the end of the first syllable and covering the complete second syllable of the verb, is highlighted with a red box. It is characterized by striations in the spectrogram and by irregularities in the cycles and period doubling visible in the waveform, as is typical of creaky voice.

2.1.5 Statistical analysis

All measurements were separately modelled as dependent variables by fitting linear mixed-effects models with the package lme4 (Bates et al., 2015, 2022) in R (R Core Team, 2022). Non-modal voice quality, the only dependent measure that was factorial instead of numeric (use of non-modal voice quality: yes, no), was modelled with binomial generalized linear mixed-effects models, with the yes = 1 and no = 0 responses logit (log-odds) transformed.

For all dependent variables, the final models were chosen through model comparison using the anova function, following the principle of only choosing a more complex model if it provided a significantly better fit than a simpler one (Matuschek et al., 2017). Model comparison always

started with a basic model containing an interaction between condition (levels: broad focus with unmarked syntax, subject focus with unmarked syntax, subject focus cleft) and constituent (subject, verb, object). The model additionally contained the predictors tone (1 = high, 4 = falling), syllable number within the word, counted from the left edge (1, 2, 3, as factor levels; below: 'syllable'), and, for f0 measures, participant gender (female, male), as well as random intercepts for participant and item. The optimal random effects structure was then determined by forward-fitting, and the fixed effects structure with a combination of backward-fitting and forward fitting, see supplementary materials, section 2, for details.

Based on the final models, pairwise comparisons were conducted with the emmeans function, adjusting p-values using the Sidak method (Lenth, 2022). All final models and full results of pairwise comparisons are reported in the supplementary materials, Section 3. Below, only estimates (β) and p-values are reported for pairwise comparisons between conditions, and they are also represented graphically with compact letter displays created with the multcomp package, showing confidence intervals with a confidence level of 0.95 (Hothorn et al., 2008, 2022). Additionally, p-values are given for effects and interactions as a whole, based on an anova-comparison between a model containing the respective effect or interaction and a model not containing it.

2.2 Results

Out of 1008 target trials (28 participants * 36 trials), 71 (7.0%) were discarded because of slips of the tongue or disfluencies (N = 29), because the participant deviated from the scripted text (N = 36), or both (N = 6). The remaining 937 target utterances contained 1916 subject syllables, 1166 verb syllables, 1384 object syllables and 304 instances of the cleft marker *shi4*. After removing the instances of *shi4*, 4466 syllables were retained for statistical analysis of duration, intensity and voice quality. Due to non-modal voice quality or extreme reduction, no reliable f0 values could be obtained for 170 syllables. Thus, the analyzed data sets for f0 measures only consisted of 4296 data points.

For almost all evaluated dependent measures, the best linear mixed-effects model contained a significant interaction between condition and constituent. As suggested by the editor, the subsections below only describe differences between conditions within constituents (i.e., paradigmatic focus effects). Readers who are interested in a fuller picture of the prosodic realization of Mandarin cleft sentences are referred to the supplementary materials, Section 3, for differences between constituents within conditions (i.e., syntagmatic focus effects), as well as all other effects revealed by the best models (also see Section 2.3 below for a brief summary).

As an overview before discussing results for individual measures, **Table 2** gives means and standard deviations for the five numeric acoustic measures and percentage of non-modal realizations by condition and constituent. Even though the best model for voice quality did not

show a significant interaction between condition and constituent, data is presented in parallel fashion for comparability. Additionally, **Figure 2a** gives an illustration of the three f0 measures. It displays average f0 contours for all conditions based on measurements at 10 equidistant points for each syllable (also see the example raw f0 contours in **Figure 3**). For the sake of comparability, *shi4* is omitted and points based on less than 10 individual measurements have been trimmed. Note also that while participants in the production experiment were not asked to produce cleft sentences in broad focus contexts, this combination was recorded for the stimuli used in the perception experiment, see **Figure 2b**.

Measure	Condition	Subject	Verb	Object
F0 range	Unmarked-Broad focus	2.1 (1.7)	1.9 (1.3)	1.7 (1.2)
F0 range	Cleft-Subject focus	2.4 (1.8)	2.1 (1.2)	1.6 (1)
F0 range	Unmarked-Subject focus	2.3 (1.8)	1.9 (1.1)	1.6 (0.9)
F0 maximum	Unmarked-Broad focus	15.6 (4.2)	14.7 (4.2)	14.3 (4.4)
F0 maximum	Cleft-Subject focus	15.9 (4.2)	13.1 (4.8)	12.4 (4.8)
F0 maximum	Unmarked-Subject focus	15.8 (4.1)	13.2 (4.6)	12.4 (4.9)
F0 minimum	Unmarked-Broad focus	13.5 (4.6)	12.8 (4.5)	12.6 (4.6)
F0 minimum	Cleft-Subject focus	13.5 (4.7)	11 (4.9)	10.7 (5)
F0 minimum	Unmarked-Subject focus	13.5 (4.5)	11.3 (4.7)	10.9 (4.9)
Duration	Unmarked-Broad focus	264.1 (75.8)	263.4 (67.4)	285.1 (73.4)
Duration	Cleft-Subject focus	269.5 (59.4)	251.8 (70)	267.4 (72)
Duration	Unmarked-Subject focus	270.1 (75.5)	251.5 (71.8)	274.2 (72.3)
Intensity	Unmarked-Broad focus	57.6 (2.8)	55.6 (3.3)	54.6 (3.3)
Intensity	Cleft-Subject focus	57.5 (2.9)	53.6 (4.1)	51.9 (3.6)
Intensity	Unmarked-Subject focus	58 (2.8)	54.6 (3.4)	52.5 (3.8)
Voice quality	Unmarked-Broad focus	4.5%	14.7%	32.7%
Voice quality	Cleft-Subject focus	5.3%	18.8%	36.2%
Voice quality	Unmarked-Subject focus	9%	23.9%	41.5%

Table 2: Mean and, in brackets, standard deviation for numeric dependent variables, and percent of non-modal realizations for all combinations of constituent and condition.
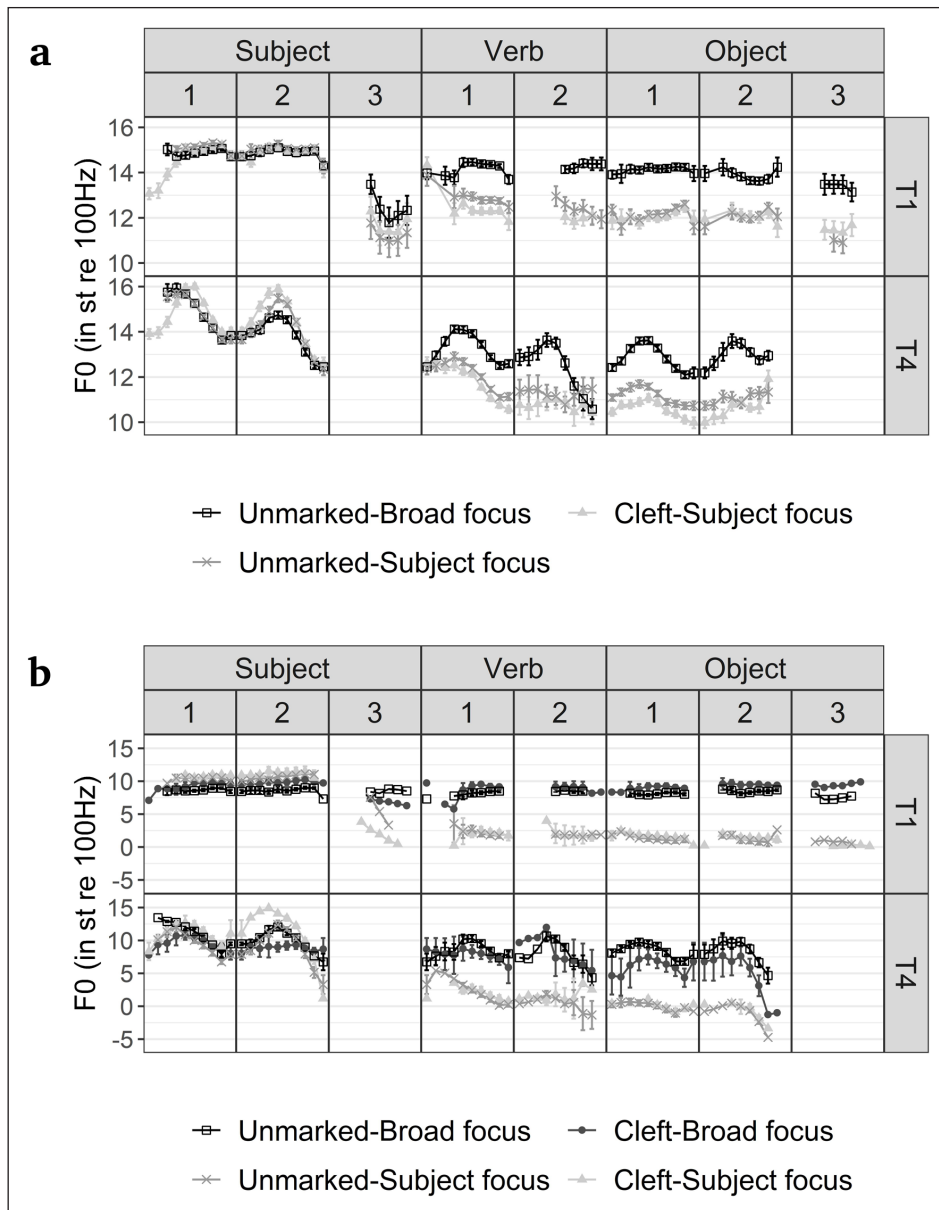
Figure 2: Time-normalized average f0 contours based on participants' realizations in the different conditions for the production experiment (panel a) and stimuli for the perception experiment (panel b) by constituent and syllable in the different syntactic and prosodic conditions.

2.2.1 F0 range

The best linear mixed-effects model of f0 range included an interaction between condition and constituent, an interaction between constituent and tone, as well as the predictors syllable and participant gender in the fixed effects (p = .02 for gender; p < .001 for all other effects and interactions).
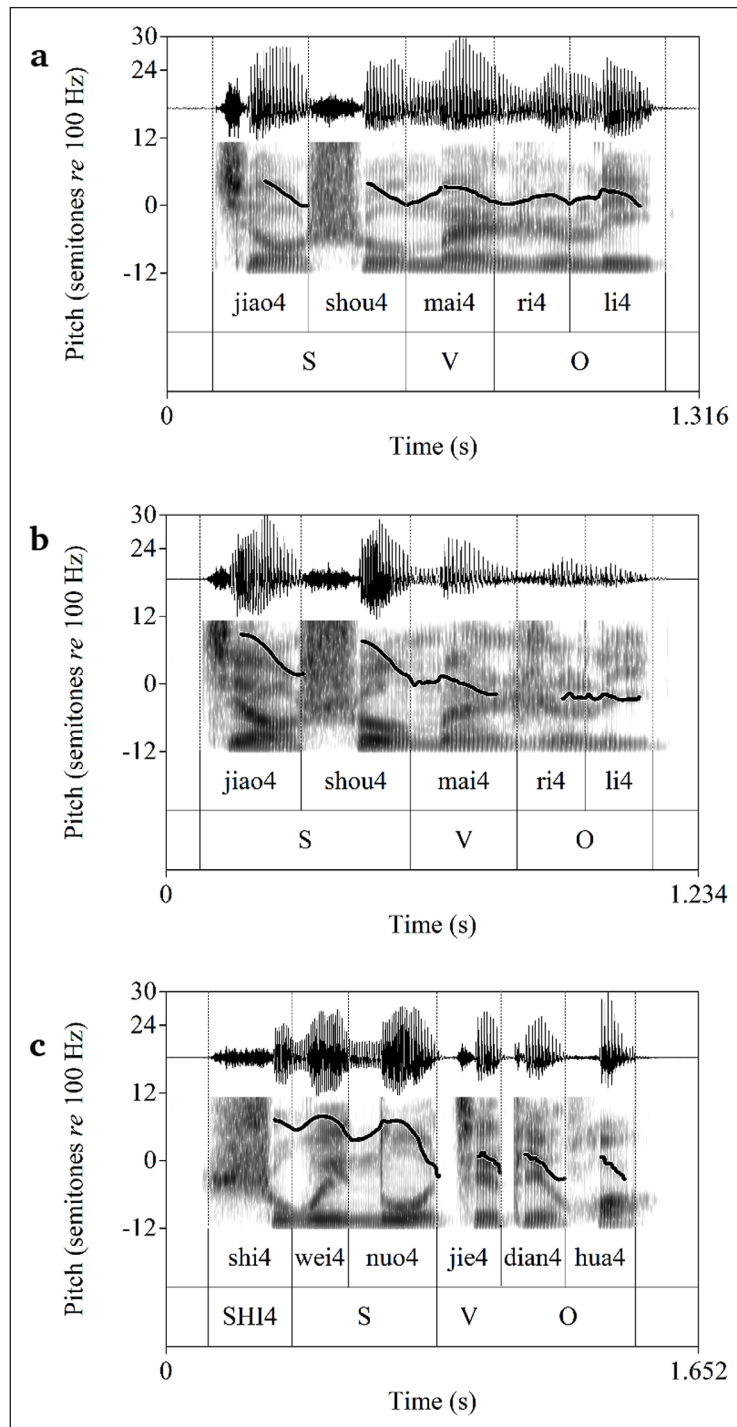
Figure 3: Example renditions for all three conditions in the production experiment. Item 17 教授卖日历。'A professor is selling calendars' with unmarked syntax in broad focus (panel a) and subject focus (panel b), and item 16 卫诺借电话。'Weinuo is borrowing a phone' in subject focus cleft (panel c), all spoken by participant s10.

Figure 4 illustrates the significant interaction between condition and constituent by showing estimated marginal means (plot symbols) and 95% confidence intervals (error bars) together with a compact letter display indicating the results of pairwise comparisons based on the model (letter combinations above the error bars). The letters in the compact letter display are not abbreviations; in fact, they have no meaning themselves. They are instead used as symbols: What is relevant is only whether each combination of letters differs from the others in the same figure. Factor combinations sharing one or more letters cannot be shown to differ significantly from each other, whereas factor combinations that do not share any letter differ significantly.

Figure 4 indicates that the conditions only differed significantly from each other on subject constituents: Subjects in both subject focus conditions (subject focus cleft, letter 'd', and subject focus with unmarked syntax, letters 'cd') had significantly larger ranges than those in broad focus (labelled 'b', a letter that is not shared by either of the other two conditions; for clefts: $\beta = -.36$, $p < .001$; for subject focus with unmarked syntax: $\beta = -.21$, $p = .01$). The two subject focus conditions did not differ significantly from each other (as indicated by the fact that their letters both include 'd'; $p = .27$). On verbs and objects, the three conditions showed no significant differences (within both constituents, all conditions share at least one letter; all $p > .1$).
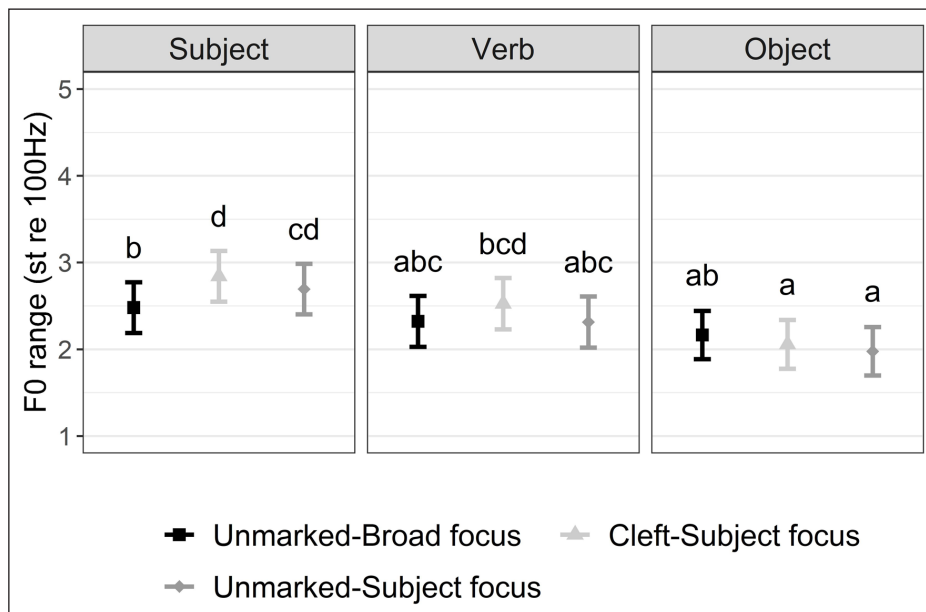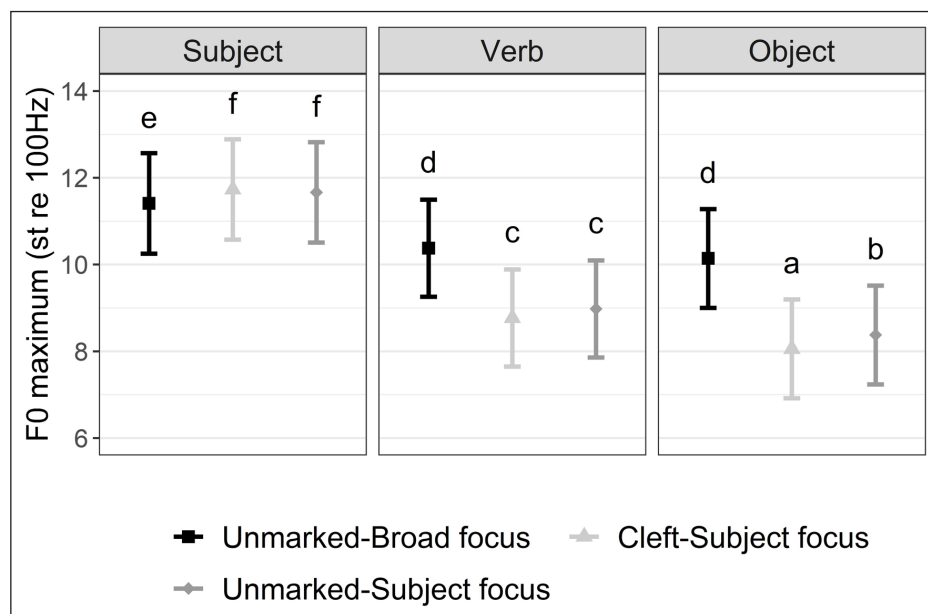

Figure 4: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of f0 range for different conditions by constituent.

2.2.2 F0 maximum

For f0 maxima, the best model contained interactions between condition and constituent, between constituent and tone, as well as syllable and gender as predictors (p < .001 for all).

As illustrated in **Figure 5**, the model indicated that subjects had significantly higher maxima in both subject focus conditions compared to broad focus (broad focus, subject – cleft, subject: β = −.32, p < .001; broad focus, subject – unmarked subject focus, subject: β = −.26, p = .01), whereas verbs and objects had lower maxima in the two subject focus conditions than in broad focus (broad focus, verb – cleft, verb: β = 1.61 p < .001; broad focus, verb – unmarked subject focus, verb: β = 1.40, p < .001; broad focus, object – cleft, object: β = 2.09, p < .001; broad focus, object – unmarked subject focus, object: β = 1.76, p < .001). There were no significant differences between the two subject focus conditions on subjects and verbs (all p > .1), but objects showed lower maxima in clefts than in subject focus with unmarked syntax (β = −.32, p = .01).
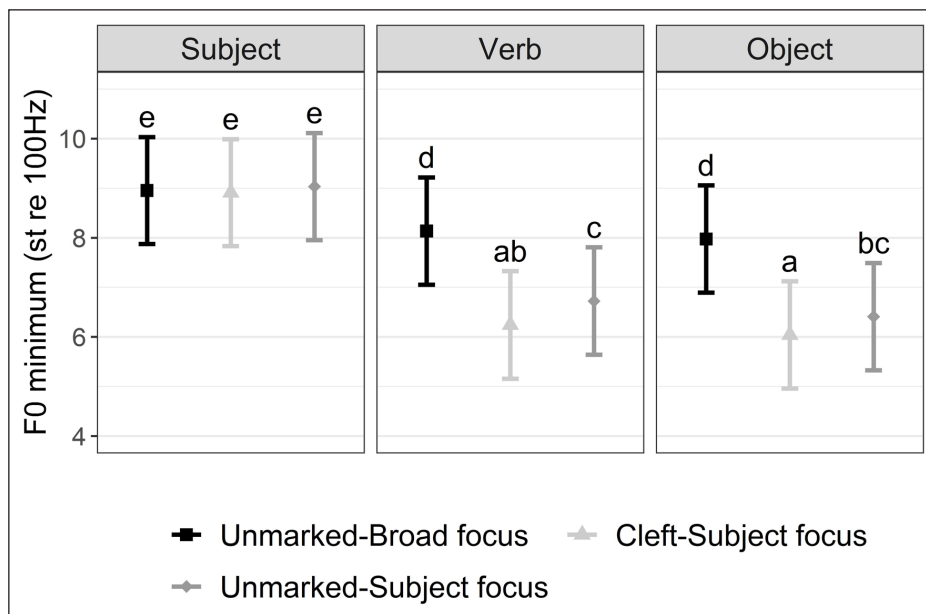

Figure 5: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of f0 maximum for different conditions by constituent.

2.2.3 F0 minimum

The best model of f0 minimum contained interactions between condition and constituent, between condition and tone, as well as syllable and gender as fixed effects (p = .02 for condition∗tone; p < .001 for all other effects and interaction condition∗constituent).

Condition did not show any significant effects on subject constituents themselves (all p > .1), but f0 minima of verbs and objects were significantly lower in both subject focus conditions compared to broad focus; see **Figure 6** (broad focus, verb – cleft, verb: β = 1.89, p < .001; broad focus, verb – unmarked subject focus, verb: β = 1.41, p < .001; broad focus, object – cleft, object: β = 1.94, p < .001; broad focus, object – unmarked subject focus, object: β = 1.57, p < .001). Additionally, verbs and objects had significantly lower minima in clefts than in subject focus with unmarked syntax (cleft, verb – unmarked subject focus, verb: β = –.48, p < .001; cleft, object – unmarked subject focus, object: β = –.37, p < .01).
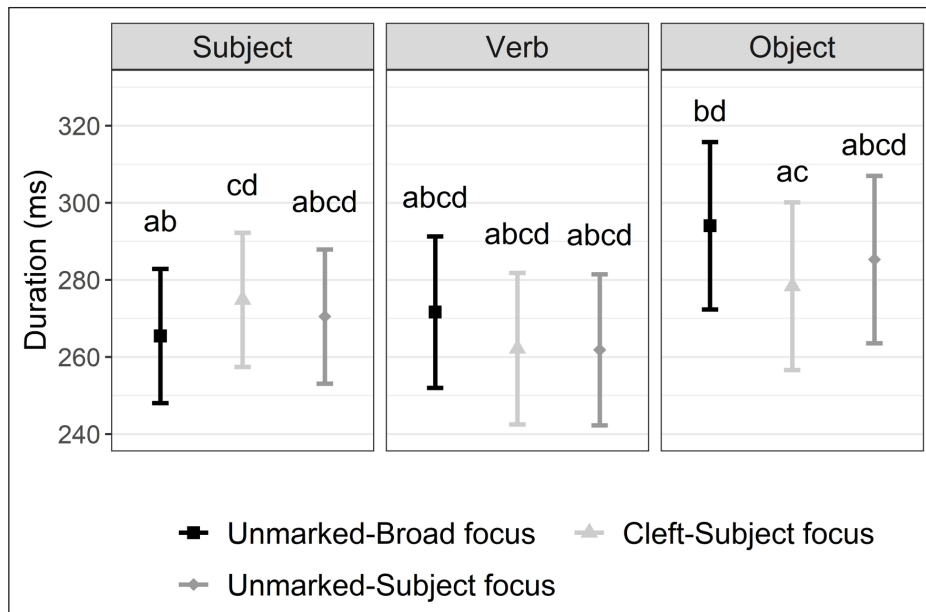

Figure 6: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of f0 minimum for different conditions by constituent.

2.2.4 Duration

For duration, the best model contained an interaction between condition and constituent (p < .001), with tone and syllable as additional predictors (p < .01 and p < .001; note that removing the interaction or tone from the best model resulted in convergence issues).

Figure 7 indicates a trend towards subjects having longer durations, and verbs and objects having shorter durations in the subject focus conditions compared to broad focus. However, only two pairwise comparisons between conditions were significant: Subjects had longer durations and objects shorter durations in clefts compared to broad focus (broad focus, subject – cleft, subject: β = –9.38, p = .01; broad focus, object – cleft, object: β = 15.67, p < .0001).

Figure 7: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of duration for different conditions by constituent.

2.2.5 Intensity

The best model of intensity included interactions between condition and constituent (p < .001), between constituent and tone (p = .02) and between condition and tone (p = .02), as well the predictor syllable (p < .001).

Condition did not significantly affect subjects (all p > .1), whereas verbs and objects had lower intensity in the two subject focus conditions than in broad focus; see **Figure 8** (broad focus, verb – cleft, verb: β = 1.86, p < .001; broad focus, verb – unmarked subject focus, verb: β = 1.11, p < .001; broad focus, object – cleft, object: β = 2.70, p < .001; broad focus, object – unmarked subject focus, object: β = 1.81, p < .001). Verbs and objects also showed significantly lower intensity in clefts than in subject focus with unmarked syntax (cleft, verb – unmarked subject focus, verb: β = –.74, p < .001; cleft, object – unmarked subject focus, object: β = –.88, p < .001).

2.2.6 Use of non-modal voice quality

Of the 4466 syllables in the dataset, 855 (19.1%) were realized with non-modal voice quality partially or completely. The best-fitting binomial model of whether or not a syllable was realized with non-modal voice quality contained an interaction between condition and tone (p = .03), as well as constituent and syllable as predictors (p < .001 for both). Unlike for the acoustic
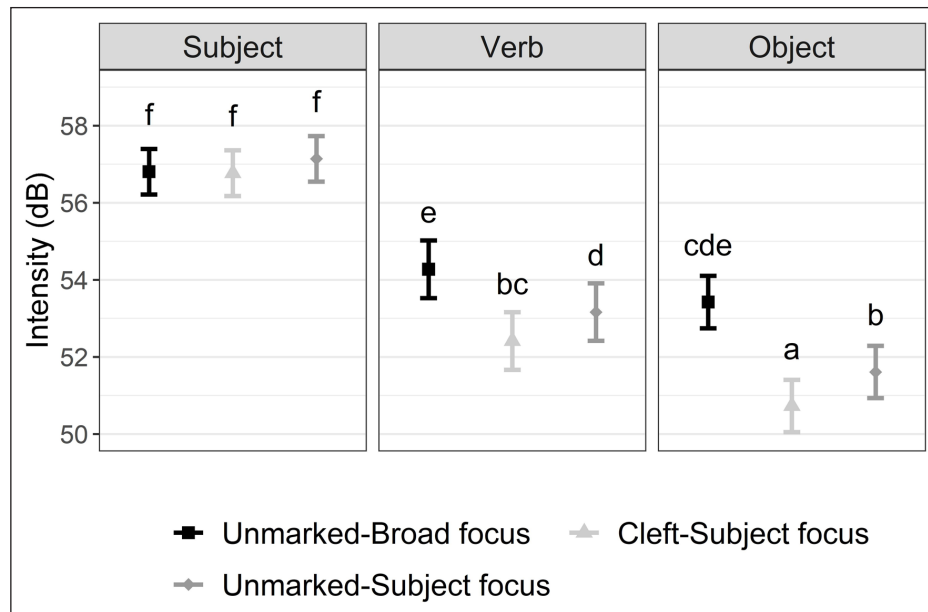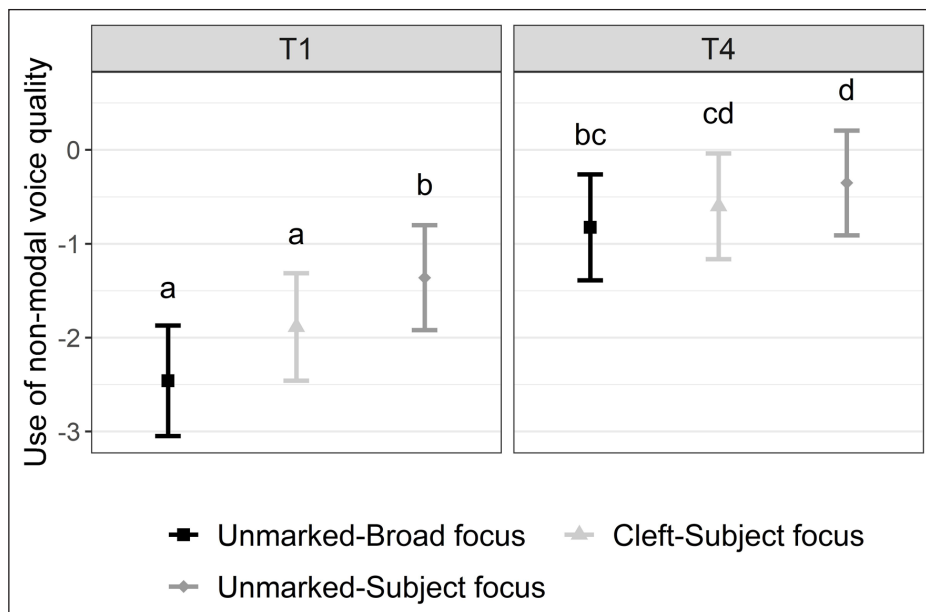
Figure 8: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of intensity for different conditions by constituent.

measures analyzed above, this model was not significantly improved by including an interaction between condition and constituent (p = .45, also note that adding this interaction to the best model resulted in convergence issues).

Regarding the effect of condition, **Figure 9** shows a similar pattern for both tones: Non-modal realizations were least frequent in broad focus and most frequent in subject focus with unmarked syntax (broad focus: 15.9%, clefts: 18.4%, subject focus with unmarked syntax: 23%). The differences between the conditions were, however, more pronounced for tone 1 items, where subject focus with unmarked syntax differed significantly from both other conditions (broad focus – unmarked subject focus: β = –1.10, p < .001; cleft – unmarked subject focus: β = –.52, p = .04; the difference between cleft and broad focus was marginal: β = –.57, p = .05). For tone 4, only the difference between broad focus and subject focus with unmarked syntax was significant (broad focus – unmarked subject focus: β = –.47, p = .02; p > .1 for both broad focus – cleft and cleft – unmarked subject focus).

2.3 Summary and discussion

The production study showed prosodic focus marking that can be described as focus expansion on subjects and post-focal reduction on verbs and objects, as summarized in **Table 3**. Focus expansion resulted in focused subjects having larger f0 ranges, higher f0 maxima and, in clefts

Figure 9: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) of use of non-modal voice quality for different conditions by tone (T1 = tone 1, T4 = tone 4). (More) negative values on the y-axis indicate fewer realizations with non-modal voice quality, positive or less negative values more realizations with non-modal voice quality.

only, longer durations compared to their realizations in broad focus. Post-focal reduction was even more pervasive, showing significant effects for all measures except f0 range. Thus, post-focal verbs and objects had lower f0 maxima and minima, as well as lower intensity than in broad focus, with objects in clefts additionally showing shorter durations compared to broad focus.

Results for non-modal voice quality do not clearly fit the classification into focus expansion and post-focal reduction, since there was no interaction between condition and constituent. Instead, sentences with unmarked syntax showed overall more frequent use of non-modal voice quality in subject focus than in broad focus. Clefts also showed marginally more frequent non-modal voice quality than broad focus, but significantly less than unmarked subject focus for tone 1. For tone 4, clefts did not differ significantly from the other two conditions.

With respect to the hypothesis of prosody-syntax trade-offs in (4), its strong version (4)a is falsified by each instance of prosodic focus marking on cleft sentences ('cleft' to the left of a > and 'broad focus' to its right in the cells of **Table 3**). As shown in the table, clefts showed prosodic focus marking, i.e., significant differences in the expected direction from the broad focus baseline, on at least one constituent with respect to all measures except the use of non-modal voice quality. Moreover, again with the exception of voice quality, in every instance

Measure	Focus expansion on subject	Post-focal reduction on verb	Post-focal reduction on object
F0 range	Clefts, unmarked subject focus > broad focus	–	–
F0 maximum	Clefts, unmarked subject focus > broad focus	Clefts, unmarked subject focus > broad focus	**Clefts > unmarked subject focus >** broad focus
F0 minimum	–	**Clefts > unmarked subject focus >** broad focus	**Clefts > unmarked subject focus >** broad focus
Duration	Clefts > broad focus	–	Clefts > broad focus
Intensity	–	**Clefts > unmarked subject focus >** broad focus	**Clefts > unmarked subject focus >** broad focus
Voice quality	Tone 1: unmarked subject focus > clefts (>) broad focus Tone 4: unmarked subject focus > broad focus		

Table 3: Summary of prosodic focus marking observed for subject focus in clefts and sentences with unmarked syntax ('unmarked subject focus'). > marks stronger prosodic focus marking for the condition on the left than for that on the right. For focus expansion, this means larger values for the subject for the condition on the left; for post-focal reduction, that means smaller values for the verb/object for the condition on the left. A comma between two conditions indicates that they did not differ significantly from each other. A dash marks that no significant differences appeared between any of the three conditions. Bolding marks significant differences between the two subject focus conditions, with clefts showing stronger focus marking than subject focus with unmarked syntax.

where clefts did not differ significantly from broad focus, neither did unmarked equivalents. Worse with respect to (4)a, significant focus marking via duration appeared only in clefts. In sum, the strong version of the hypothesis can clearly be rejected.

The weak version of the hypothesis, (4)b, would be upheld if both clefts and their unmarked equivalents showed prosodic focus marking, but less so in clefts than with unmarked syntax. This is not supported by the data, either. In all instances where the two conditions differed significantly, again with the exception of voice quality, clefts showed stronger post-focal reduction (bolding in **Table 3**). Together with the instances where clefts showed focus marking and unmarked equivalents did not, this suggests that not only can the hypothesis be rejected even in its weaker form–if anything, clefts show clearer focus marking than syntactically unmarked sentences.

The only prosodic correlate showing effects in the expected direction was voice quality: Non-modal voice quality appeared significantly more frequently in subject focus with unmarked syntax than in broad focus, whereas clefts only differed marginally from broad focus for tone 1 items. As clefts did thus not reliably mark subject focus via voice quality, these results are in line with hypothesis (4)a. Even taking into account the marginal effect for clefts, at the very least hypothesis (4)b can be upheld with respect to voice quality, as sentences with unmarked syntax had significantly more non-modal realizations than clefts for tone 1 items (and insignificantly more for tone 4 items, as well). However, in the face of all other prosodic cues contradicting the hypothesis in (4), it should not be upheld based on the findings for voice quality alone.

Two additional points speak for this conclusion: First, the hypothesis in (4) was more tentative for voice quality than for the other measures, as detailed in Section 1.3. Second, the findings for the five acoustic measures more clearly match previous research than those for voice quality. Prosodic focus marking observed in the present study generally echoes previous research for the five acoustic measures, although prosodic focus marking in terms of duration and intensity had previously only been found in terms of syntagmatic comparisons between constituents within a sentence (recall Section 1.2). In addition to differences between conditions, the present data also showed differences between the constituents, described in Section 3 of the supplementary materials. For the five acoustic measures, these observed differences between constituents, including findings of syntagmatic focus effects, are in line with previous research. They indicate a downtrend in f0 and intensity over the course of an utterance, reflected in subjects having significantly higher f0 maxima, f0 minima and mean intensity than verbs and objects (also see Xu, 1999; Yuan & Liberman, 2010, on declination in Mandarin). Verbs had significantly higher f0 maxima and mean intensity than objects only in subject focus, indicating that the declination was steeper post-focally than in broad focus, replicating previous research on post-focal compression as summarized in Section 1.2. Similarly, differences between the constituents in f0 range were only significant for tone 4, not for tone 1 (also see Y. Chen & Gussenhoven, 2008; B. Wang & Xu, 2011; and Xu for differences between lexical tones with respect to post-focal compression). In contrast to these findings echoing existing research, the present study indicated that the use of non-modal voice quality increases significantly later in the sentence, which has not been reported for Mandarin before. The present finding of prosodic focus marking via increased use of non-modal voice quality does agree with previous research, but was not localized to any particular constituent, unlike in previous studies. A follow-up study considering all lexical tones and varying focus location would be needed.

To summarize, with the possible exception of the use of non-modal voice quality, the results of the present production study clearly contradict the hypothesis of prosody-syntax trade-offs given in (4). As pointed out by an anonymous reviewer, this could be because the task of reading

sentences as answers in imagined dialogue situations is not as engaging as a more interactive or naturalistic scenario. However, two observations speak against this explanation. First, while significant differences often did not appear between the two subject focus conditions, significant differences between broad focus and subject focus appeared for all dependent variables. This suggests that the experimental manipulation was effective in engaging participants at least to a certain extent, since they did produce systematic prosodic focus marking (which, as pointed out, matches expectations based on previous research). Second, some significant differences did appear between the two subject focus conditions, and these almost exclusively indicated stronger prosodic focus marking for clefts than with unmarked syntax. Therefore, the present findings indicate that prosodic focus marking may even be stronger in clefts than in sentences with unmarked syntax.

3. Perception experiment

3.1 Methods

3.1.1 Participants

Of the 103 native speakers of Mandarin who participated in the study, data from one was sorted out because they gave incorrect answers to the majority of comprehension questions (see supplementary materials, Section 1). None of the remaining 102 participants (58 female, 43 male, one did not specify gender; age 17–28, mean 20.27, one did not specify their age) reported an uncorrected visual or hearing impairment. The supplementary materials provide more details on their language backgrounds in Section 7.

All participants were undergraduate students enrolled in an introduction to linguistics course at the University of Alberta and received partial course credit for their participation. None of them had participated in the production experiment. The study was approved by the Research Ethics Board 2 of the University of Alberta (study ID Pro00069978).

3.1.2 Materials and procedure

The same 24 target sentences as in the production experiment appeared in two syntactic conditions, as clefts and with unmarked syntax. Both syntactic conditions were recorded with two prosodic conditions, one with clear prosodic focus marking on the subject constituent and one with prosody appropriate for a broad focus context. All stimuli were spoken by a male native speaker of Mandarin from Northern China. At the time of recording, he was 23 years old and an undergraduate student of linguistics at the University of Alberta. As visible in **Figure 2**, his productions were similar to those of the participants in the production experiment; see supplementary materials, Section 4, for further details.

All four combinations of syntax (unmarked, cleft) and prosodic focus marking (broad focus, subject focus) appeared in two context conditions (broad focus, subject focus) for a $2 \times 2 \times 2$ design. The contexts were the same as in the production study. In each trial, participants first read the context, then heard the target sentence through headphones. They were asked to rate how appropriate this answer was given the preceding context on a scale from 1 (最不合适 'most inappropriate/unsuitable') to 7 (最合适 'most appropriate/suitable') by clicking the corresponding box on the screen. They were additionally instructed at the beginning of the session that a rating of 7 indicated that this was how they would answer this question in everyday life. Participants received no particular instruction regarding speed vs. accuracy of their response.

Target trials were distributed across eight lists following a Latin square design, such that each participant responded to all 24 target items, as well as 36 fillers. Of those, 24 contained target sentences of varying lengths, which were presented in two conditions, with prosodic focus marking either matching the preceding question or indicating narrow focus on a different constituent. The other fillers included answers that were completely incongruent with the preceding question (e.g., 'How many siblings do you have?' – 'The weather is nice today.').

All experimental sessions were conducted with small groups of participants in a university computer lab, using E-Prime to present stimuli and record responses and reaction times (Psychology Software Tools, 2016). Block and trial orders were randomized individually.

3.1.3 Statistical analysis

Ratings and reaction times were analyzed with generalized additive mixed-effects models (GAMMs, Wood, 2017), as implemented in the package mgcv in R (Wood, 2023), specified as ordinal GAMMS for the ratings. GAMMs have several advantages for the analysis of rating data in particular (Baayen & Divjak, 2017); see supplementary materials, Section 5, for further details.

To select the best-fitting model of each dependent variable, model comparisons were performed with the function compareML from the package itsadug (van Rij et al., 2022). The initial model of ratings contained a three-way interaction between context (broad focus, subject focus), syntax (cleft, unmarked) and prosody (broad focus, subject focus), as well as the predictors list (as a factor), a smooth for trial (centered by subtracting the mean) and random smooths for participant and item. The initial model of reaction times contained a four-way interaction between context, syntax, prosody and rating (as a factor), as well as list, a smooth for trial and random smooths for participants and item. As for the production study, the best

random-effects structure was first determined by forward-fitting, before the interactions and the effects of list and trial were assessed through backward-fitting. Again, pairwise comparisons were conducted with the emmeans function (Lenth, 2022), and compact letter displays created with the multcomp package (Hothorn et al., 2022).

3.2 Results

All final models and full results of pairwise comparisons are reported in the supplementary materials, Section 6. Below, only estimates and p-values are reported.

3.2.1 Ratings

Altogether 2448 rating responses were evaluated (102 participants * 24 items). The GAMM with the best fit to the data contained only an interaction between context and prosody, as well as a smooth for trial and by-participant and by-item random smooths for context ($p < .01$ for trial; $p < .001$ for smooths and interaction context*prosody). Other interactions, as well as including syntax or list, did not significantly improve model fit ($p = .08$ for context*prosody*syntax; $p = .16$ for context*syntax; $p = .86$ for list; adding syntax or prosody*syntax decreased model fit,
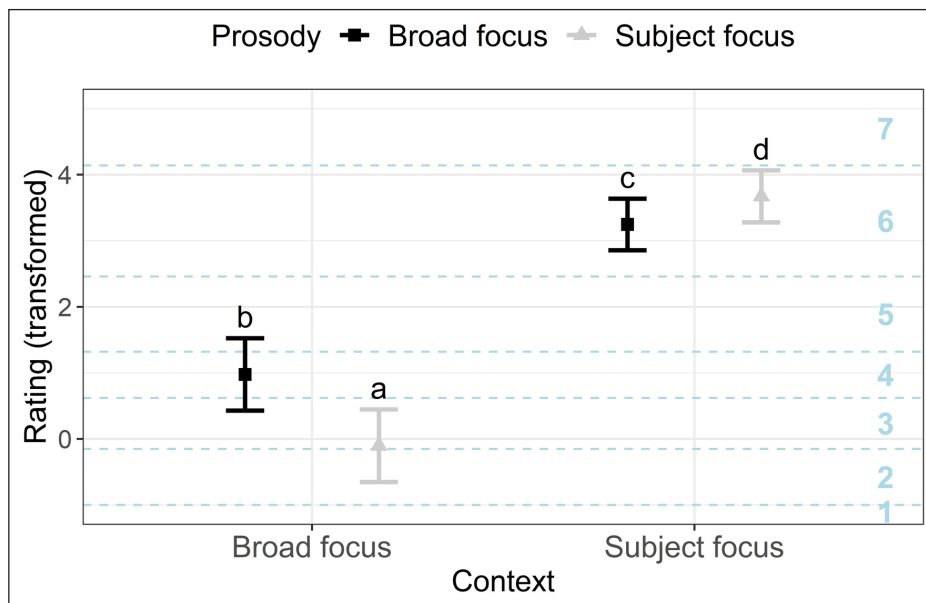
Figure 10: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) for rating by context and prosody. Since the y-axis reflects the transformed continuous dependent variable used in statistical modelling, dashed lines in light blue and bold numbers on the right indicate boundaries relating these to the original rating categories.

increasing AIC by 1.85 and 0.46 respectively). Thus, there was no evidence that participants rated clefts differently from unmarked syntax, nor that either syntactic structure was more appropriate in either context or prosodic realization.

Answers to subject focus questions were rated as more appropriate than those to broad focus questions; see **Figure 10**. This difference was significant with both broad focus and subject focus prosody on the answers themselves (broad focus context, broad focus prosody – subject focus context, broad focus prosody: $\beta = -2.27$, $p < .001$; broad focus context, subject focus prosody – subject focus context, subject focus prosody: $\beta = -3.77$, $p < .001$). Moreover, context affected the effect of prosody: In broad focus contexts, answers with broad focus prosody were rated significantly higher than those with subject focus prosody ($\beta = 1.08$, $p < .001$), whereas the opposite was the case in subject focus contexts ($\beta = -.43$, $p < .001$).

3.2.2 Reaction times

Of the 2448 rating responses, reaction times were discarded for 59 responses that occurred more than 500 ms before the end of the stimulus sound and for 67 responses that occurred more than 10,000 ms after its end (5.15% of the data total), and leaving 2322 trials for analysis. Reaction times from these trials (from the end of the stimulus sound, in ms) were log transformed after adding the largest negative reaction time to all reaction times, since negative numbers cannot be log transformed. This resulted in a distribution closely resembling a bell-shaped curve. Note that in **Figure 11** below, both the log transformation and the addition of the largest negative reaction time are reversed, i.e., model results are simply displayed on the original ms scale.

The best model of reaction times contained the predictors prosody, context and rating, a smooth for trial, a by-item smooth for context and a random smooth for participant ($p = .01$ for prosody; $p < .001$ for context, rating and smooths). Neither the effect of syntax, list, nor any interactions were significant (adding context*prosody*syntax, context*syntax, prosody*syntax, context*prosody, syntax or list decreased model fit, increasing AIC by 1.02, 1.07, 1.42, 1.32, 0.64 and 0.26 respectively).

Figure 11a shows that participants reacted slightly faster to stimuli with subject focus prosody than to stimuli with broad focus prosody ($\beta = -.06$, $p < .001$). Additionally, as shown in **Figure 11b**, reaction times were smaller for stimuli with subject focus contexts than for those with broad focus contexts ($\beta = -.155$, $p < .001$). Finally, reaction times were smaller with more extreme ratings, whereas the ratings in the middle of the scale were associated with longer reaction times (cf. **Figure 11c** and **Table 20** in the supplementary materials).
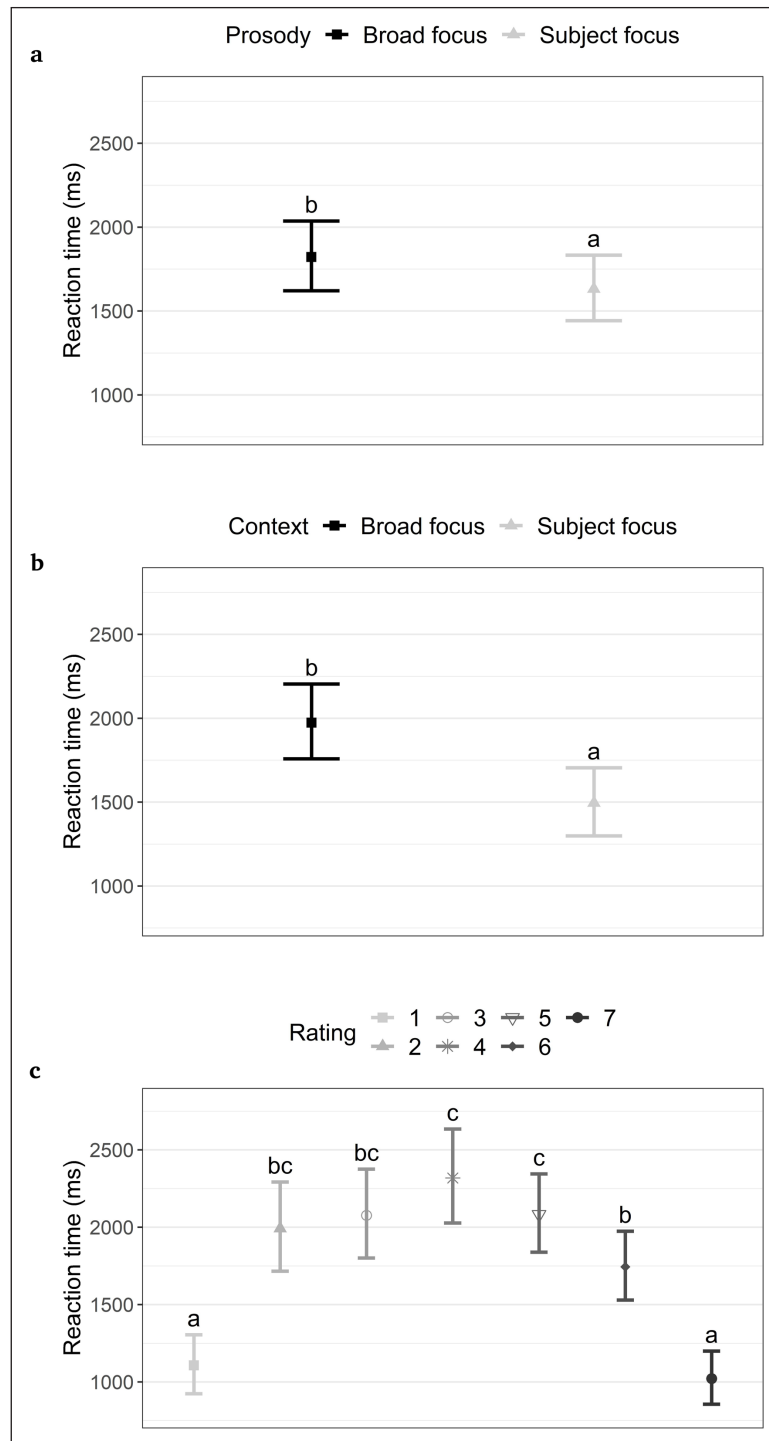
Figure 11: Estimated marginal means, 95% confidence intervals and results of pairwise comparisons (compact letter displays: different letters indicate significant differences, shared letters that no significant different difference could be shown) for reaction time (in ms) for stimuli with different prosody (panel a) and presented in different contexts (panel b) and receiving different ratings (panel c).

3.3 Summary and discussion

The results of the perception study showed a clear effect of prosodic focus marking on ratings, which interacted with context in the expected way: Participants rated target sentences higher when their prosody matched the information structure induced by the preceding context. In contrast to context and prosody, syntax did not significantly affect ratings or reaction times and did not interact significantly with either factor, contradicting both versions of the hypothesis of trade-offs in (5). The weak version of the hypothesis (5b) predicted that when syntactic and prosodic focus marking are combined, effects of both prosody and syntax would be detectable, but participants would be less affected by one of them compared to the effect this marking would have when occurring on its own. In contrast to this prediction, the results showed a total absence of syntax effects, whereas the effect of prosody persisted even when combined with concurrent syntactic marking. The strong version of the hypothesis (5a) predicted that both syntax and prosody would affect ratings and reaction times on their own, but that in conditions with concurrent prosodic and syntactic marking, either syntax or prosody would be redundant and its effect would not be detectable. However, a syntax effect was absent not only when syntactic and prosodic marking concurred, but also when they did not. Thus, the absence of a syntax effect in conditions with concurrent syntax and prosody cannot be explained as a trade-off. Put differently, in contrast to the hypothesis, there was no evidence for an effect of syntax that could have been reduced or neutralized by concurrent prosody.

The absence of any effect of syntax constitutes a difference from a parallel study on English, where an effect of syntax appeared with the same methods (Arnhold, 2021). This difference between the languages is in line with previous research showing that clefting has weaker effects than prosodic focus marking on perception and processing in Mandarin (Yan et al., 2020; Yan & Calhoun, 2019, 2020) and may also be relatively less important for perception and processing in Mandarin than in English (S. H. Chen et al., 2012; but see Yan & Calhoun, 2020). However, it is interesting that the present experiment showed no evidence that clefting affected the ratings at all. Previous studies on Mandarin also did not find clefting to induce lexical priming (Yan & Calhoun, 2019), speed up the rejection of false alternatives (Yan & Calhoun, 2020) or improve ratings compared to unmarked syntax, even when focus marking matched the context (Yan et al., 2020). Still, the syntactic manipulation affected all these tasks: It slowed reaction times overall in lexical decision and false alternative rejection tasks,[4] as well as lowering ratings when the clefted constituent was not the one that the context indicated to be focused (Yan et al., 2020; Yan & Calhoun, 2019, 2020). Importantly, all of these studies lacked a broad focus baseline. Therefore, either prosodic and syntactic focus marking indicated narrow focus on the same constituent (e.g., (6a) for subject focus and (6b) for object focus) or they provided conflicting clues, each pointing to a different constituent being in narrow focus (e.g., (6c) with subject cleft, but prosodic focus

---

[4] Analyses of reaction times are not reported for the rating experiment in Yan et al. (2020).

marking on the object, or (6d), an object cleft with prosodic focus marking on the subject; underlining indicates prosodic focus marking in these examples). As Yan et al. acknowledge, this mismatch in itself could have lowered ratings. Actually, if mismatch cases are discounted, as well as conditions differing only marginally from each other, the only significant effects of the syntactic manipulation in Yan et al.'s experiment that remain appeared in object focus contexts: Subject clefts were rated lower than unmarked syntax when prosody marked the subject as focused, i.e., clefting a background constituent additionally lowered ratings (e.g., following 'What did the captain put on?', (6a) was rated even lower than (6e)). However, when prosody marked the object as focused in accordance with the context, object clefts were also rated significantly lower than unmarked syntax, though still higher than (mismatch) subject clefts with the same prosody (e.g., (6f) was preferred over (6b), which was in turn preferred over (6c) as an answer to 'What did the captain put on?'). This could indicate that it is not only object clefts that are less frequent and, accordingly, less readily acceptable, than subject clefts in Mandarin (as stated by Yan et al., based on their own results and previous literature), but that clefts are generally less acceptable in object than in subject *focus*. This could be because the appropriate cleft in object focus would be an object cleft or, more likely, causality could run the other way around, in line with cross-linguistically common subject/non-subject asymmetries mentioned in Section 1.2. It is therefore possible that the present experiment did not detect any effects of syntax because it contained neither prosody-syntax mismatches of the type in (6c&d) nor object focus conditions.

(6)   Example stimuli from Yan et al.
   a.   是<u>船长</u>穿<u>上</u>的雨衣 'It was the <u>CAPTAIN</u> who put on the raincoat'
   b.   船长是穿上的<u>雨衣</u> 'It was the <u>RAINCOAT</u> that the captain put on'
   c.   是船长穿上的<u>雨衣</u> 'It was the captain who put on the <u>RAINCOAT</u>'
   d.   <u>船长</u>是穿上的雨衣 'It was the raincoat that the <u>CAPTAIN</u> put on'
   e.   <u>船长</u>穿上了雨衣 'The <u>CAPTAIN</u> put on the raincoat'
   f.   船长穿上了<u>雨衣</u> 'The captain put on the <u>RAINCOAT</u>'

The present experiment did not include items where prosody and clefting indicated narrow focus on different constituents, as unlike for Yan et al. (2020), the research question was not whether prosody or syntax would prevail in such a mismatch. Instead, the purpose of the present study was to test whether effects of syntactic and prosodic focus marking would diminish when combined. Therefore, conditions where prosody and clefting indicated narrow focus on the same constituent were compared to conditions where only one of them did, while the other was neutral and did not indicate narrow focus on any constituent. Of course, it can be questioned whether prosody can ever be truly neutral. While the unmarked syntactic structures used in the present study are equally compatible with broad focus and subject focus (and with other information structures), broad focus prosody might be considered infelicitous following a subject

focus question and therefore not truly be neutral–at the very least, the results of the present study indicate that Mandarin listeners find it significantly less appropriate than subject focus prosody in this context. Based on this, one might argue that clefts with broad focus prosody, as used in the present study, also constitute a mismatch between prosody and syntax. However, the present results suggest that this is not in itself enough to lower ratings significantly in the way that the mismatches in Yan et al. did. It appears that even if broad focus prosody is not neutral, it does not conflict with subject clefting as much as object focus prosody does or, alternatively, that the particular conflict tested by Yan et al. is just particularly unacceptable in Mandarin. This is in line with Greif and Skopeteas' (2021) conclusion that Mandarin subject clefts with object focus, even if supported by a suitable context, are "marginal in language use" (p. 17).[5]

It may also be relevant that bare *shi* clefts can be interpreted as verum focus (recall Section 1.2), although it is not agreed whether this possibility exists whenever prosody does not mark subject focus (Paul & Whitman, 2008) or, in line with cross-linguistic trends, only with prosodic focus marking on the copula *shi4* (Liu & Shi, 2022). Liu and Shi's description is compatible with the idea that the clefted constituent is focused per default, but this default can be overwritten with specific verum focus prosody. Paul and Whitman's account, on the other hand, suggests that Mandarin clefts are a means of marking focus on the clefted constituent only when prosody concurs. However, Greif and Skopeteas (2021) found that subject clefts were rated as more appropriate than unmarked syntax in subject focus contexts–even though their study used written stimuli. While participants likely supplied implicit prosody, as mentioned in footnote 1, it is interesting that their implicit prosody for clefts apparently marked the subject as focused. Alternatively, participants may have interpreted the subject clefts as verum focus but still judged this as a better fit with subject focus contexts than the default interpretation for unmarked syntax, which is that the subject is a topic (see Section 1.2).

---

[5] In Greif and Skopeteas' (2021) written rating study, Mandarin subject clefts with object focus received a mean rating of 3.1 out of 7 (compared to 5.7 for subject clefts with subject focus). Unlike for German and English, these ratings did not significantly improve with a context containing a cleft and prompting second occurrence focus (e.g., A: *It's John that sold the car*, B: *No, it's John that sold the bicycle*; Mandarin equivalents were rated 3.6 on average). The authors therefore suggest that Chinese clefts with focus following the clefted constituent are "marginal in language use" (p. 17). This is in line with the fact that Yan and co-authors refer to them as mismatches. Note also that bare *shi* clefts as used in the present experiments are compatible with broad focus according to Cheng (2008), as long as the broad (sentential) focus is contrastive focus, i.e., the sentence in (7a) could be interpreted as 'Aunt is steaming mushrooms (and not uncle frying onions)'. Cheng does not describe the prosodic realization of these sentences, and I am not aware of any studies comparing the prosody of contrastive and non-contrastive broad focus in Mandarin. If contrastive and non-contrastive broad focus prosody do not differ from each other, and the broad focus prosody used here corresponds to the prosody of Cheng's examples, then the clefts with broad focus prosody do not constitute a mismatch condition at all. Thus, this combination of syntax and prosody would be inherently grammatical, meaning syntax and prosody are compatible with *each other*. Of course, a contrastive broad focus interpretation should still be incompatible with the non-contrastive broad focus *contexts* used in the present perception study.

At any rate, the possibility of interpreting subject clefts as marking verum focus does not explain the absence of a syntax effect in the present experiment. Verum focus could perhaps be seen as more similar to broad focus than subject focus, but it is not the same (e.g., *My aunt IS steaming mushrooms* would be an odd answer to a broad focus question such as, *What's happening?*). Thus, in the broad focus contexts, it makes sense that clefts with broad focus prosody (indicating verum focus according to Paul & Whitman, 2008) were rated more highly than clefts with subject focus prosody due to prosody alone–for the same reason that sentences with unmarked syntax were rated more highly with broad focus prosody than with subject focus prosody in these contexts (e.g., (7a) was preferred over (7b), just like (7c) was preferred over (7d)). However, it is still unexpected that among sentences with broad focus prosody, clefts like (7a) were not rated lower than unmarked syntax equivalents like (7c) in broad focus contexts. Even with a possible verum focus interpretation, a cleft should not be an appropriate answer to a broad focus question. It is possible that participants accommodated an unprompted insistence on the truth of the utterance, but if so, it is surprising that this affected neither ratings nor reaction times. Even if participants ultimately judged the fit between question and answer as high, accommodation of an initially less expected interpretation could be expected to lead to processing delays.

(7)   Example stimuli from the present study (item 3)
    a.   是姑妈蒸冬菇 'It is aunt who is steaming mushrooms'
    b.   是<u>姑妈</u>蒸冬菇 'It is <u>AUNT</u> who is steaming mushrooms'
    c.   姑妈蒸冬菇 'Aunt is steaming mushrooms'
    d.   <u>姑妈</u>蒸冬菇 '<u>AUNT</u> is steaming mushrooms'

Regarding subject focus contexts, when prosody marks the subject as focused, higher ratings would be expected for clefts like (7b) than for unmarked syntax as in (7d) if prosodic and syntactic focus marking were additive. The fact that no significant differences appeared could be taken as support for the strong version of the hypothesis in (5), i.e., the effect of syntax being undetectable when combined with the effect of prosody. However, as stated above, given that effects of syntax appeared nowhere else in the present experiment, it seems more likely that clefting simply did not affect ratings at all. A potential explanation for this is explored in the general discussion (Section 4).

A final finding that should be discussed, though it is not directly relevant to the question of trade-offs, is that context itself had a significant effect in the present study, with answers to subject focus questions receiving higher ratings than answers to broad focus questions. Arnhold (2021) observed the same context effect in a parallel experiment on English and speculated that the more specific narrow focus questions may have induced the perception of a better fit between question and answer in these cases. In line with this interpretation, in the present experiment participants responded significantly faster in subject focus contexts than in broad focus contexts.

Additionally, reaction times were also significantly smaller for stimuli with subject focus prosody compared to broad focus prosody. Interestingly, however, the English results showed a ceiling effect where all answers to subject focus questions were rated so highly that no difference could be observed between the two prosodic conditions (or the syntactic ones, which likewise differed significantly in broad focus contexts). This contrasts with the present findings for Mandarin, which showed a significant difference between the ratings of the two prosodic realizations in both context conditions. It is possible that the context effect was weaker in Mandarin than in English, or that the effect of prosodic focus marking was stronger in Mandarin. The latter explanation seems more likely, given that previous studies on the perception of clefts also show a stronger effect of prosodic focus marking in Mandarin than in English (S. H. Chen et al., 2012; Yan & Calhoun, 2020).

To sum up, this experiment did not provide evidence for the hypothesis of prosody-syntax trade-offs.

4. General discussion

The two experiments reported in this paper set out to test the hypothesis of prosody-syntax trade-offs in information structure marking in Mandarin. With regard to the production experiment, the strong version of the hypothesis (4a) was that prosodic marking of subject focus would not occur in subject clefts, which already mark subject focus syntactically, so that only sentences with unmarked syntax should show significant differences from the broad focus baseline (subject focus with unmarked syntax > subject focus with clefts, broad focus). The weak version hypothesis (4b) was that clefts would show prosodic focus marking, but that it would be significantly weaker than in sentences with unmarked syntax (subject focus with unmarked syntax > subject focus with clefts > broad focus). These predictions were tested by analyzing six measures of prosody, with the stipulation that the hypothesis could be considered supported if at least three of the measures showed the predicted patterns. However, only one measure showed results in accordance with the hypothesis: Focus marking in terms of voice quality–a higher frequency of non-modal realizations throughout the sentence reported here for the first time–was significant only for subject focus with unmarked syntax and not for clefts when compared to broad focus. This is in line with the strong version of the hypothesis in (4a). By contrast, the results of the five acoustic measures did not support either version of the hypothesis. F0 minimum, f0 maximum and intensity showed stronger prosodic focus marking for clefts than for unmarked equivalents (bolding in **Table 3** above), while f0 range showed no significant differences between subject focus with unmarked syntax and clefts: Both conditions exhibited the prosodic differences from broad focus that are expected based on the previous literature on Mandarin focus marking. Thus, the findings for the overwhelming majority of evaluated indicators of prosodic focus marking clearly contradicted both versions of the hypothesis of trade-offs in (4).

The second experiment tested the corresponding hypothesis for perception by asking participants to rate subject clefts and sentences with unmarked syntax, each of which had prosody indicating either subject focus or broad focus. These four combinations were presented in either subject focus or broad focus contexts. The strong version of the trade-off hypothesis (5a) predicted that when prosodic and syntactic focus marking concurred (clefts with subject focus prosody), the effect of one of them should become undetectable, either the effect of syntax (clefts with subject focus prosody = unmarked syntax with subject focus prosody), or the effect of prosody (clefts with subject focus prosody = clefts with broad focus prosody). The weak version (5b) hypothesized that both effects would still be present when combined, but that one of them would be reduced (i.e., effect of prosody x + effect of syntax $y < x + y$). The results of the perception experiment showed that syntax did not affect ratings when combined with concurrent prosodic focus marking, as hypothesized in (5a). However, effects of syntax did not appear in any of the other conditions, either. This casts doubt on the idea that concurrent prosody neutralized the effect of syntax, as it appears that there was no effect of syntax in the first place that could have been neutralized.

In sum, neither the production nor the perception study provided clear evidence for prosody-syntax trade-offs in focus marking for Mandarin clefts. These findings indicate that prosodic focus marking is not redundant in clefts. On the contrary, they suggest that prosodic focus marking in clefts is necessary. In fact, the results of the perception study suggest that prosody alone is sufficient for information structure marking in Mandarin. Since prosodic focus marking in syntactically unmarked sentences is a perfectly ordinary way to signal information structure, this claim should not be controversial in itself. The interesting question is what follows from that with respect to the meaning and function of clefts.

I would like to argue that syntax and prosody make separate but interrelated contributions to the pragmatics of clefts, as suggested by Delin (1995) on the basis of English data. In particular, clefts have complex meaning and multiple functions and, specifically, the present results suggest that prosody performs the function of information structure marking, as it does in non-cleft sentences. It is clear that clefts are semantically and pragmatically complex constructions, as mentioned in footnote 2 (see e.g., Hole, 2011; Hole & Zimmermann, 2013; Liu & Yang, 2016, on exhaustivity of Mandarin clefts). It then seems likely that the use of a cleft construction contributes precisely those nuances, such as exhaustivity, which were not tested in the present experiments. This would explain why clefting did not affect ratings or reaction times in the perception experiment–participants were asked about information structure only, not about exhaustivity or other pragmatic aspects. Effects of clefting did, however, appear in the production study. This supports the suggestion that prosody and syntax of clefts are interrelated.

The precise origin and nature of the interplay of prosody and syntax in deriving the semantics and pragmatics of clefts is beyond the scope of the present article (for more discussion, see the literature cited in the introduction and footnote 2, and Delin, 1995; Westera, 2017; and Bourgoin,

2022). But it seems reasonable that the meaning contributed by the cleft construction interacts with prosody by exerting an influence on plausible information structures. For example, narrow focus on the clefted constituent goes particularly well with the implicature that this constituent is to be interpreted exhaustively: In (8), prosody marks 姑妈 'aunt' as narrow focus, meaning that speaker and listener already know that someone from a set of people (e.g., mother, father, aunt, uncle) steamed mushrooms, and the information that should be added to the common ground is that aunt did so (cf. semantic analysis of focus in Rooth, 1992; Krifka, 2008; Zimmermann & Onea, 2011). This fits naturally with the implicature of exhaustivity arising from clefting, i.e., it was only aunt and no one else, explaining why clefts are often used with prosody marking the clefted constituent as narrow focus (e.g., Collins, 2006, p. 1708, finds that "the great majority of *it*-clefts" in a spoken corpus have nuclear accents on the clefted constituent). Other types of information structure are possible, but may be limited by their fit with the pragmatics implied by the syntactic structure.

(8)   是   <u>姑妈</u>   蒸   (的)   冬菇   (的)。
       shi4   <u>gu1ma1</u> zheng1 (de0)   dong1gu1 (de0)
       COPULA   <u>aunt</u>   steam   (DE)   mushroom   (DE)
       'It was <u>AUNT</u> who steamed mushrooms.'

Thus, prosodic focus marking is not simply independent of and orthogonal to the meaning conveyed by the syntax of clefts. Instead, the two seem to be truly interrelated, with one influencing the other. In the production experiment, prosodic focus marking was significantly stronger in clefts than with unmarked syntax in terms of three of the six evaluated measures. This is hard to square with the assumption that prosody performs information structure marking independently of the syntactic construction. If that were the case, why would subject focus prosody not be the same in clefts and unmarked SVO sentences? Instead, speakers produced clefts with exaggerated prosodic focus marking, in line with the idea that prosody and syntax of clefts make related contributions, as advocated here. In particular, it seems that the presence of clefting enhanced the suitability of prosody marking the subject as focused beyond what was required by the preceding subject focus question (as detailed with respect to (8) above, participants not only stated that the backgrounded information was true of the subject, but that it was uniquely true of the subject).

Further evidence comes from the results of the rating experiment in Greif and Skopeteas (2021). They showed that Mandarin subject clefts are rated higher than unmarked syntax in subject focus contexts, even though their experiment used written stimuli. Thus, it seems that clefts themselves may convey focus on the clefted constituent, without prosodic focus marking. Even if this is explained via implicit prosody, as suggested in Section 3.3 above, it needs to be explained why the prosodic realization that comes to mind when reading a cleft is the one marking the clefted constituent as focused. On the assumption that syntax and prosody of clefts

are completely independent, Greif and Skopeteas' participants should have been equally likely to imagine the subject focus prosody fitting the context for clefts as they were to imagine it for sentences with unmarked syntax. Also, the prosody fitting default interpretation for unmarked SVO sentences—that the subject is a topic, should have been just as likely to come to mind for participants with clefts as with unmarked syntax. Thus, clefts and unmarked syntax should have received the same ratings. One might say that subject focus prosody might come to mind more easily for a subject cleft because it is the most frequent or the most prototypical prosody, but why? If prosodic focus marking were orthogonal to the syntax of clefting, such an association should not emerge, since clefts should be possible with the full range of information structures that prosody can also signal in syntactically unmarked sentences. Instead, Greif and Skopeteas' finding fits the idea of related contributions advocated here, as well as the results of the present production study, where prosodic focus marking was stronger in clefts than in unmarked equivalents with subject focus. If the prosody that participants in the present production study used corresponds to the implicit prosody that the participants in Greif and Skopeteas' rating study imagined (either because this is the most frequent or prototypical prosody for clefts, or because of the fit between the meaning of this prosodic form and the meaning of clefting, as discussed above), this would explain why they rated written clefts as more suitable than unmarked syntax in subject focus contexts. Thus, implicit prosody could indeed explain why clefting alone can apparently indicate information structure and, in particular, focus on the clefted constituent in writing. Actually, this would mean that written clefts do not necessarily cue focus on the clefted constituent *in the absence of prosodic cues*, since the invocation of the relevant prosodic cues via implicit prosody would be the crucial mechanism. More generally, the use of clefts may be a way for writers to cue the desired prosody precisely because spoken clefts usually come with this prosody. If substantiated, this could also explain why clefts are more prone to appearing in writing than in spoken language (see summary of the literature in Section 1.1). Future experiments testing the role of implicit prosody in the perception of written clefts are desirable to test this hypothesis.

An alternative–though not necessarily conflicting–hypothesis is that the interrelatedness of syntax and prosody of clefts that I advocate here goes as far as transferring some of the ability to mark information structure from prosody to the syntactic structure of clefts. In other words, one could argue that focus marking, originally and primarily a function of prosody, has to some degree become associated with the syntactic form of cleft constructions. This, of course, is in line with the literature assuming that clefts are commonly a means of focus marking, with some authors explicitly stating that clefts are equivalent to employing prosodic focus marking with unmarked syntax (Atlas & Levinson, 1981; Lambrecht, 2001). For Mandarin specifically, for example Hole's (2011) analysis of complete *shi…de* clefts assumes a syntactic partition into focus and background. Moreover, several authors analyze *shi4* as a focus marker by itself (see summary

and discussion in Lee, 2005, p. 22–24), and even among scholars who disagree widely about the structural analysis of the Chinese cleft construction, "[t]he general agreement is that it is a focus construction" (Cheng, 2008, p. 235). Thus, the idea is that focus marking does not just happen to happen *in* clefts, via prosody, just as in sentences with unmarked syntax, but that focus marking can also happen *by means of* using of clefts.

For Mandarin, the evidence reviewed above is compatible with this extended hypothesis, but it does not lend any positive support to it. Thus, the discussion so far has ascribed all information structure marking to explicit or implicit prosody without running into any problems. However, it could also be argued that clefting itself can be interpreted as focusing the clefted constituent in the absence of evidence to the contrary. Conflicting prosody would constitute evidence to the contrary, explaining the absence of a syntax effect in the present perception study: Subject clefts were not able to cue subject focus in the presence of contradictory broad focus prosody, but when prosody concurred in indicating subject focus, participants treated subject clefts as indicating subject focus, as expected. Importantly, when prosody marked subject focus, there was no indication that the interpretation as subject focus was somehow clearer or stronger for clefts than with unmarked syntax. Thus, the experiment failed to provide positive support for clefts themselves indicating focus on the clefted constituent. This differs from the corresponding study for English reported in Arnhold (2021), which showed additive effects of prosody and syntax, suggesting that clefting itself constituted a cue to information structure. As discussed in Section 3.3, this difference between the two studies using the same design is in line with other research indicating that Chinese listeners are more likely to favor the interpretation suggested by prosody, whereas for English, clefts themselves are more likely to be interpreted as marking focus on the clefted constituent, even in the absence of concurring prosodic cues (Yan et al., 2020; Yan & Calhoun, 2019, 2020; S. H. Chen et al., 2012; Kember et al., 2019). It is possible that English has moved further on a trajectory of entangling information structure with the syntax of clefts. It is also possible that clefting signals information structure equally in both languages, but that it is more specifically associated with narrow focus on the clefted constituent in English, whereas it simply marks that information structure differs from the default in Mandarin. For both languages, it has been shown that clefts can come in more than one type and with more than one information structure (e.g., for English, Prince, 1978; Hedberg, 1990; Declerck, 1988; Huber, 2006; Van Praet & O'Grady, 2018; Karssenberg et al., 2019; for Mandarin, Lee, 2005; Cheng, 2008; Paul & Whitman, 2008). Therefore, more research would be needed to clarify these cross-linguistic differences.

What is clear from the preceding discussion is that prosodic focus marking is far from being redundant in clefts. Therefore, even though clefts have played a prominent role in the literature on syntax-prosody interactions in marking information structure, and their involvement in trade-offs has been suggested frequently (see Section 1.1), clefts may not actually be the best place to

look for trade-offs. While the present results did not show any evidence for prosody-syntax trade-offs in information structure marking, this does not necessarily disprove the idea that language is efficient and avoids redundancy. Instead, these findings may suggest that we should look for evidence supporting this idea elsewhere.

5. Conclusion

Based on a production and a perception study on Mandarin clefts, the present article did not find evidence for trade-offs between prosodic and syntactic focus marking. In production, consistent prosodic focus marking in terms of f0 range, f0 maximum, f0 minimum and intensity appeared for both clefts and in unmarked syntax, while effects of focus on duration only appeared in clefts. These results are consistent with the previous literature on prosodic focus marking in Mandarin sentences with unmarked syntax, and notably expand our knowledge on prosodic focus marking in clefts, as previously published production studies only reported results on f0. The present production study additionally suggested a role for voice quality in focus marking, which should be followed up in future research. In perception, the difference between clefts and unmarked syntax neither significantly affected ratings of the fit between the target sentence and the preceding context, nor did the syntactic manipulation interact with context or prosody of the target sentence. Prosodic realization, in contrast, affected ratings in the expected manner, with higher ratings for prosodic focus marking matching the preceding question, as well as affecting reaction times. These findings suggest that prosody is necessary to mark focus on the clefted constituent, instead of being a potentially redundant cue.

Data availability

Funding information

Acknowledgements

Competing interests

The author has no competing interests to declare.

References

Akmajian, A. (1970). On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry*, *1*(2), 149–168. http://www.jstor.org/stable/10.2307/4177550

Arnhold, A. (2018). *MeasureIntensityDurationF0minF0maxF0contourpoints.praat*. Anja Arnhold's Praat Scripts. https://sites.ualberta.ca/~arnhold/praatScripts.html

Arnhold, A. (2021). Prosodic focus marking in clefts and syntactically unmarked equivalents: Prosody–syntax trade-off or additive effects? *Journal of the Acoustical Society of America*, *149*(3), 1390–1399. https://doi.org/10.1121/10.0003594

Atlas, J. D., & Levinson, S. C. (1981). It-clefts, informativeness and logical form: Radical pragmatics (revised standard version). In P. Cole (Ed.), *Radical Pragmatics* (pp. 1–62). Academic Press.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309 040470010201

Baayen, R. H., & Divjak, D. (2017). Ordinal GAMMs: A New Window on Human Ratings. In A. Makarova, S. M. Dickey & D. Divjak (Eds.), *Each Venture a New Beginning: Studies in Honor of Laura A. Janda* (pp. 39–56). Slavica. http://www.sfs.uni-tuebingen.de/~hbaayen/publications/BaayenDivjak2017.pdf

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *lme4: Linear Mixed-Effects Models Using "Eigen" and S4* (R package version 1.1-31).

Beekhuizen, B., Bod, R., & Zuidema, W. (2013). Three design principles of language: The search for parsimony in redundancy. *Language and Speech, 56*(3), 265–290. https://doi.org/10.1177/0023830913484897

Blything, L. P., Järvikivi, J., Toth, A. G., & Arnhold, A. (2021). The influence of focus marking on pronoun resolution in dialogue context. *Frontiers in Psychology, 12*, 2876. https://doi.org/10.3389/fpsyg.2021.684639

Boersma, P., & Weenink, D. (2020). *Praat. Doing phonetics by computer.* Version 6.1.14.

Bourgoin, C. (2022). *A corpus-based study of the prosody and information structure of English it-clefts and French c'est-cleft.* [Doctoral dissertation, Cardiff University and University of Leuven]. ProQuest Dissertations and Theses.

Büring, D. (2010). Towards a typology of focus realization. In C. Féry & M. Zimmermann (Eds.), *Information Structure: Theoretical, Typological, and Experimental Perspectives* (pp. 177–205). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570959.003.0008

Büring, D., & Gutiérrez-Bravo, R. (2002). Focus-related word order variation without the NSR: A prosody-based crosslinguistic analysis. In S. Mac Bhloscaidh (Ed.), *Syntax & Semantics at Santa Cruz* (Vol. 3, pp. 41–58). Linguistics Research Center University of California, Santa Cruz.

Calhoun, S., Wollum, E., & Kruse Va'ai, E. (2021). Prosodic prominence and focus: Expectation affects interpretation in Samoan and English. *Language and Speech, 64*(2), 346–380. https://doi.org/10.1177/0023830919890362

Cao, W., & Zhang, J. (2008). Tone-3 accent realization in short Chinese sentences. *Tsinghua Science and Technology, 13*(4), 533–539. https://doi.org/10.1016/S1007-0214(08)70085-3

Cassarà, A., Adli, A., & Karssenberg, L. (2022). Clefts in context: A QUD-perspective on c'est/il y a utterances in spoken French. *Isogloss, 8*(1), 1–29. https://doi.org/10.5565/rev/isogloss.197

Chen, S. H., Chen, S. C., & He, T. H. (2012). Surface cues and pragmatic interpretation of given/new in Mandarin Chinese and English: A comparative study. *Journal of Pragmatics, 44*(4), 490–507. https://doi.org/10.1016/j.pragma.2011.12.006

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724–746. https://doi.org/10.1016/j.wocn.2008.06.003

Cheng, L. L. S. (2008). Deconstructing the *shì ... de* construction. *Linguistic Review, 25*(3–4), 235–266. https://doi.org/10.1515/TLIR.2008.007

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology* (pp. 183–216). Cambridge University Press.

Collins, P. C. (1991). *Cleft and pseudo-cleft constructions in English*. Routledge. https://doi.org/10.4324/9780203202463

Collins, P. C. (2006). It-clefts and wh-clefts: Prosody and pragmatics. *Journal of Pragmatics*, *38*(10), 1706–1720. https://doi.org/10.1016/j.pragma.2005.03.015

Coupé, C., Oh, Y., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, *5*(9). https://doi.org/10.1126/sciadv.aaw2594

de Cesare, A.-M. (Ed.) (2014). *Frequency, forms and functions of cleft constructions in Romance and Germanic: Contrastive, corpus-based studies*. De Gruyter Mouton. https://doi.org/10.1515/9783110361872

De Clercq, K. (2017). Prosody as an argument for a layered left periphery. *Nederlandse Taalkunde*, *22*(1), 31–39. https://doi.org/10.5117/nedtaa2017.1.decl

Declerck, R. (1988). *Studies in copular sentences, clefts, and pseudo-clefts*. Leuven University Press.

Delin, J. (1995). Presupposition and shared knowledge in it-clefts. *Language and Cognitive Processes*, *10*(2), 97–120. https://doi.org/10.1080/01690969508407089

Di Tullio, Á. (2006). Clefting in spoken discourse. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., Vol. 2, pp. 483–491). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00564-2

Dufter, A. (2009). Clefting and discourse organization: Comparing Germanic and Romance. In A. Dufter & D. Jacob (Eds.), *Focus and Background in Romance Languages* (pp. 83–121). John Benjamins. https://doi.org/10.1075/slcs.112.05duf

Fedzechkina, M. (2014). *Communicative efficiency, language learning, and language universals* [Doctoral dissertation, University of Rochester]. ProQuest Dissertations and Theses.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, *41*(2), 416–446. https://doi.org/10.1111/cogs.12346

Fenk-Oczlon, G., & Fenk, A. (2008). Complexity trade-offs between the subsystems of language. In M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language Complexity. Typology, Contact, Change* (pp. 43–65). John Benjamins. https://doi.org/10.1075/slcs.94.05fen

Fenk-Oczlon, G., & Pilz, J. (2021). Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.626032

Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory*, *31*(3), 683–734. https://doi.org/10.1007/s11049-013-9195-7

Féry, C., & Arnhold, A. (2019). Verum focus and negation. In J. M. M. Brown, A. Schmidt & M. Wierzba (Eds.), *Of Trees and Birds: A Festschrift for Gisbert Fanselow* (pp. 213–229). University of Potsdam. https://doi.org/10.25932/publishup-43235

Féry, C., Paslawska, A., & Fanselow, G. (2007). Nominal split constructions in Ukrainian. *Journal of Slavic Linguistics*, *15*(1), 3–48.

Frascarelli, M., & Ramaglia, F. (2013). (Pseudo)clefts at the syntax-prosody-discourse interface. In K. Hartmann & T. Veenstra (Eds.), *Cleft Structures* (Issue 1, pp. 97–138). John Benjamins. https://doi.org/10.1075/la.208.04fra

Geluykens, R. (1984). *Focus phenomena in English. An empirical investigation into cleft and pseudo-cleft sentences*. (Vol. 36) *Antwerp papers in linguistics*. University of Antwerp.

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Greif, M., & Skopeteas, S. (2021). Correction by focus: Cleft constructions and the cross-linguistic variation in phonological form. *Frontiers in Psychology*, *12*, 648478. https://doi.org/10.3389/fpsyg.2021.648478

Gundel, J. K. (2006). Clefts in English and Norwegian: Implications for the grammar-pragmatics interface. In V. Molnár & S. Winkler (Eds.), *The Architecture of Focus* (Issue 2, pp. 517–548). De Gruyter Mouton. https://doi.org/10.1515/9783110922011.517

Gundel, J. K. (2008). Contrastive perspectives on cleft sentences. In M. de los Á. Gómez González, J. L. Mackenzie & E. M. González Alvarez (Eds.), *Languages and Cultures In Contrast and Comparison* (pp. 69–87). John Benjamins. https://doi.org/10.1075/pbns.175.06gun

Gutzmann, D., Hartmann, K., & Matthewson, L. (2020). Verum focus is verum, not focus: Cross-linguistic evidence. *Glossa: A Journal of General Linguistics*, *5*(1), 51. https://doi.org/10.5334/gjgl.347

Haan, J. (2001). *Speaking of questions. An exploration of Dutch question intonation* [Doctoral dissertation, LOT].

Hamlaoui, F. (2007). French cleft sentences and the syntax-phonology interface. In M. Radišić (Ed.), *Proceedings of the 2007 Annual Conference of the Canadian Linguistic Association* (pp. 1–11). Canadian Linguistic Association. https://cla-acl.ca/pdfs/actes-2007/Hamlaoui.pdf

Han, C., & Romero, M. (2014). Disjunction, focus and scope. *Linguistic Inquiry*, *35*(2), 179–217.

Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics*, *57*(3), 605–633. https://doi.org/10.1017/S0022226720000535

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford University Press.

Hedberg, N. (1990). *Discourse pragmatics and cleft sentences in English* [Doctoral dissertation, University of Minnesota].

Hedberg, N. (2000). The referential status of clefts. *Language*, *76*(4), 891–920. https://doi.org/10.2307/417203

Höhle, T. N. (1992). Über Verum-Fokus im Deutschen [About verum focus in German]. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (pp. 112–141). Westdeutscher Verlag. https://doi.org/10.1007/978-3-663-12176-3_5

Hole, D. (2011). The deconstruction of Chinese shì...de clefts revisited. *Lingua*, *121*(11), 1707–1733. https://doi.org/10.1016/j.lingua.2011.07.004

Hole, D., & Zimmermann, M. (2013). Cleft partitionings in Japanese, Burmese and Chinese. In K. Hartmann & T. Veenstra (Eds.), *Cleft Structures* (pp. 285–318). John Benjamins. https://doi.org/10.1075/la.208.11hol

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., & Scheibe, S. (2022). *multcomp: Simultaneous Inference in General Parametric Models* (R package version 1.4-20).

Huang, Y. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *The Journal of the Acoustical Society of America, 147*(3), 1441–1458. https://doi.org/10.1121/10.0000721

Huang, Y., Athanasopoulou, A., & Vogel, I. (2018). The effect of focus on creaky phonation in Mandarin Chinese tones. *University of Pennsylvania Working Papers in Linguistics*, *24*(1), 1–9.

Huber, S. (2006). The complex functions of it-clefts. In V. Molnár & S. Winkler (Eds.), *The Architecture of Focus* (pp. 549–578). De Gruyter Mouton. https://doi.org/10.1515/9783110922011.549

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press.

Jespersen, O. (1937). *Analytic syntax*. Allen and Unwin.

Jun, S.-A., & Bishop, J. (2015). Priming Implicit Prosody: Prosodic Boundaries and Individual Differences. *Language and Speech, 58*(4), 459–473. https://doi.org/10.1177/0023830914563368

Karssenberg, L., Lahousse, K., Lamiroy, B., Marzo, S., & Drobnjakovic, A. (2019). Non-prototypical clefts. Formal, semantic and information-structural properties. *Belgian Journal of Linguistics*, *32*(1), 1–20. https://doi.org/10.1075/bjl.00014.kar

Kember, H., Choi, J., Yu, J., & Cutler, A. (2019). The processing of linguistic prominence. *Language and Speech*. https://doi.org/10.1177/0023830919880217

Kiss, K. É. (1999). The English cleft construction as a focus phrase. In L. Mereu (Ed.), *Boundaries of Morphology and Syntax* (pp. 217–231). John Benjamins. https://doi.org/10.1075/cilt.180.14kis

Koplenig, A., Meyer, P., Wolfer, S., & Müller-Spitzer, C. (2017). The statistical trade-off between word order and word structure–Large-scale evidence for the principle of least effort. *PLoS ONE*, *12*(3), e0173614. https://doi.org/10.1371/journal.pone.0173614

Kratzer, A., & Selkirk, E. (2020). Deconstructing information structure. *Glossa: A Journal of General Linguistics*, *5*(1). https://doi.org/10.5334/gjgl.968

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*, 243–276. http://doi.org/10.1556/aling.55.2008.3-4.2

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America, 142*(3), 1693–1706. https://doi.org/10.1121/1.5003649

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics, 39*(373), 463–516. https://doi.org/10.1515/ling.2001.021

Lee, H. (2005). *On Chinese focus and cleft constructions* [Doctoral dissertation, National Tsing Hua University].

Lenth, R. (2022). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.3). https://cran.r-project.org/web/packages/emmeans/index.html

Levshina, N. (2021). Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology, 12*, 648200. https://doi.org/10.3389/fpsyg.2021.648200

Liu, Y., & Shi, W. (2022). Verum shi, sentence-final de and the emphatic effects in Mandarin. *Lingua, 267*, 103186. https://doi.org/10.1016/j.lingua.2021.103186

Liu, Y., & Yang, Y. (2016). Exhaustivity in Mandarin shi … (de) sentences: experimental evidence. In M. Köllner & R. Ziai (Eds.), *Proceedings of the ESSLLI 2016 Student Session. 28th European Summer School in Logic, Language & Information August 15–26, 2016, Bozen-Bolzano, Italy* (pp. 167–178). Free University of Bozen-Bolzano. https://esslli2016.unibz.it/wp-content/uploads/2016/09/esslli-stus-2016-proceedings.pdf

Lohnstein, H. (2016). Verum focus. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 290–313). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642670.013.33

Maddieson, I. (2005). Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts. *UC Berkeley Phonology Lab Annual Reports, 1*, 259–268. https://doi.org/10.5070/p73cm3w6ck

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McWhorter, J. H. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology, 5*(2), 125–166. https://doi.org/10.1515/lity.2001.001

Miestamo, M. (2008). Grammatical complexity in cross-linguistic perspective. In M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language Complexity. Typology, Contact, Change* (pp. 23–41). John Benjamins. https://doi.org/10.1075/slcs.94.04mie

Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 118*(49), e2025993118. https://doi.org/10.1073/pnas.2025993118

Nichols, J. (2009). Linguistic complexity: A comprehensive definition and survey. In G. Sampson, D. Gil & P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 110–125). Oxford University Press.

Onea, E. (2019). Exhaustivity in it-clefts. In C. Cummins & N. Katsos (Eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics* (pp. 401–417). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.17

Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(1–2), 57–72. https://doi.org/10.1080/01690965.2013.805795

Paul, W., & Whitman, J. (2008). Shi ... de focus clefts in Mandarin Chinese. *Linguistic Review*, *25*(3–4), 413–451. https://doi.org/10.1515/TLIR.2008.012

Pimentel, T., Roark, B., & Cotterell, R. (2020). Phonotactic complexity and its trade-offs. *Transactions of the Association for Computational Linguistics*, *8*, 1–18. https://doi.org/10.1162/tacl_a_00296

Pinelli, M. C., Poletto, C., & Avesani, C. (2020). Does prosody meet syntax? A case study on standard Italian cleft sentences and left peripheral focus. *Linguistic Review*, *37*(2), 309–330. https://doi.org/10.1515/tlr-2019-2045

Prince, E. F. (1978). A comparison of wh-clefts and it-clefts in discourse. *Language*, *54*(4), 883. https://doi.org/10.2307/413238

Psychology Software Tools. (2016). *E-Prime* (Version 2.0.10.252). [Computer software].

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2). R Foundation for Statistical Computing. http://www.r-project.org/

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 237–263). Wiley Blackwell. https://doi.org/10.1002/9781118346136.ch11

Reinhart, T. (2006). Focus: The PF interface. In T. Reinhart (Ed.), *Interface Strategies: Optimal and Costly Computations* (pp. 125–163). MIT Press.

Roberts, G., & Fedzechkina, M. (2018). Social biases modulate the loss of redundant forms in the cultural evolution of language. *Cognition*, *171*, 194–201. https://doi.org/10.1016/j.cognition.2017.11.005

Rochemont, M. S. (1986). *Focus in generative grammar*. John Benjamins.

Rooth, Mats. (1992). A Theory of Focus Interpretation. *Natural Language Semantics*, *1*(1), 75–116. http://doi.org/10.1007/BF02342617

Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory*, *23*(3), 687–755. https://doi.org/10.1007/s11049-004-2874-7

Sánchez-Alvarado, C. (2020). Syntactic and prosodic marking of subject focus in American English and Peninsular Spanish. In A. Morales-Front, M. J. Ferreira, R. P. Leow & C. Sanz (Eds.), *Hispanic Linguistics: Current Issues and New Directions* (pp. 184–203). John Benjamins. https://doi.org/10.1075/ihll.26.09san

Shosted, R. K. (2006). Correlating complexity: A typological approach. *Linguistic Typology*, *10*(1), 1–40. https://doi.org/10.1515/LINGTY.2006.001

Siewierska, A. (1998). Variation in major constituent order: A global and a European perspective. In A. Siewierska (Ed.), *Constituent Order in the Languages of Europe* (pp. 475–552). De Gruyter Mouton. https://doi.org/10.1515/9783110812206.475

Simpson, A., & Wu, Z. (2002). From D to T: Determiner incorporation and the creation of tense. *Journal of East Asian Linguistics*, *11*(2), 169–209. http://www.jstor.org/stable/20100822

Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. In M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language Complexity. Typology, Contact, Change* (pp. 67–88). John Benjamins. https://doi.org/10.1075/slcs.94.06sin

Sinnemäki, K. (2010). Word order in zero-marking languages. *Studies in Language, 34*(4), 869–912. https://doi.org/10.1075/sl.34.4.04sin

Sinnemäki, K. (2014). Complexity trade-offs: A case study. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring Grammatical Complexity* (pp. 179–201). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685301.003.0009

Skopeteas, S., & Fanselow, G. (2010). Focus types and argument asymmetries: a cross-linguistic study in language production. In C. Breul & E. Göbbel (Eds.), *Comparative and Contrastive Studies of Information Structure* (pp. 169–197). John Benjamins.

Szendrői, K. (2017). The syntax of information structure and the PF interface. *Glossa: A Journal of General Linguistics, 2*(1), 32. https://doi.org/10.5334/gjgl.140

Tal, S., & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition, 224*, 105055. https://doi.org/10.1016/j.cognition.2022.105055

Tönnis, S., Fricke, L. M., & Schreiber, A. (2016). Argument asymmetries in German cleft sentences. In M. Köllner & R. Ziai (Eds.), *Proceedings of the ESSLLI 2016 Student Session. 28th European Summer School in Logic, Language & Information August 15–26, 2016, Bozen-Bolzano, Italy* (pp. 208–218). Free University of Bozen-Bolzano. https://esslli2016.unibz.it/wp-content/uploads/2016/09/esslli-stus-2016-proceedings.pdf

Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. In P. W. Culicover & L. McNally (Eds.), *The Limits of Syntax* (pp. 79–108). Academic Press. https://doi.org/10.1163/9789004373167_005

van Heuven, V. J. (2017a). Prosody and sentence type in Dutch. *Nederlandse Taalkunde, 22*(1), 3–29. https://doi.org/10.5117/nedtaa2017.1.heuv

van Heuven, V. J. (2017b). Functional trade-off of prosody and syntax in question marking? *Nederlandse Taalkunde, 22*(1), 41–42. https://doi.org/10.5117/nedtaa2017.1.heue

Van Praet, W., & O'Grady, G. (2018). The prosody of specification: Discourse intonational cues to setting up a variable. *Journal of Pragmatics, 135*, 87–100. https://doi.org/10.1016/j.pragma.2018.07.013

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2022). *itsadug: Interpreting time series and autocorrelated data using GAMMs* (R package version 2.4.1).

van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning and function.* Cambridge University Press.

Vander Klok, J., Goad, H., & Wagner, M. (2018). Prosodic focus in English vs. French: A scope account. *Glossa, 3*(1), 71. https://doi.org/10.5334/gjgl.172

Vicenik, C. (n.d.). *intensity-scaler.txt.* Retrieved May 17, 2011, from http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/IntensityScaler.txt%0A

Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics, 39*(4), 595–611. https://doi.org/10.1016/j.wocn.2011.03.006

Wang, B., Xu, Y., & Ding, Q. (2017). Interactive prosodic marking of focus, boundary and newness in Mandarin. *Phonetica*, *75*(1), 24–56. https://doi.org/10.1159/000453082

Wang, T., Liu, J., Lee, Y. H., & Lee, Y. C. (2020). The interaction between tone and prosodic focus in Mandarin Chinese. *Language and Linguistics*, *21*(2), 331–350. https://doi.org/10.1075/lali.00063.wan

Wehr, B. (2005). Focusing strategies in Old French and Old Irish. In J. Skaffari, R. Hiltunen, R. Carroll & M. Peikola (Eds.), *Opening Windows on Texts and Discourses of the Past* (pp. 353–379). John Benjamins. https://doi.org/10.1075/pbns.134.28weh

Westera, M. (2017). *Exhaustivity and intonation. A unified theory*. Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC.

Wood, S. N. (2023). *Package 'mgcv'* (R package version 1.8-42).

Xie, Z. (2012). The modal uses of de and temporal shifting in Mandarin Chinese. *Journal of East Asian Linguistics*, *21*(4), 387–420. https://doi.org/10.1007/s10831-012-9093-8

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 55–105. https://doi.org/doi:10.1006/jpho.1999.0086

Yadav, H., Vaidya, A., Shukla, V., & Husain, S. (2020). Word order typology interacts with linguistic complexity: A cross-linguistic corpus study. *Cognitive Science*, *44*(4). https://doi.org/10.1111/cogs.12822

Yan, M. (2020). *Prosodic and syntactic focus in speech processing in Mandarin Chinese* [Doctoral dissertation, Victoria University of Wellington].

Yan, M., & Calhoun, S. (2019). Priming effects of focus in Mandarin Chinese. *Frontiers in Psychology*, *10*, 1985. https://doi.org/10.3389/fpsyg.2019.01985

Yan, M., & Calhoun, S. (2020). Rejecting false alternatives in Chinese and English: The interaction of prosody, clefting, and default focus position. *Laboratory Phonology*, *11*(1), 17. https://doi.org/10.5334/LABPHON.255

Yan, M., Calhoun, S., & Warren, P. (2020). Prosody or syntax? The perception of focus by Mandarin speakers. *Proceedings of the International Conference on Speech Prosody 2020* 347–351. https://doi.org/10.21437/SpeechProsody.2020-71

Yan, M., Warren, P., & Calhoun, S. (2022). Focus Interpretation in L1 and L2: The Role of Prosodic Prominence and Clefting. *Applied Psycholinguistics*, *43*(6), 1275–1303. http://doi.org/10.1017/S0142716422000376

Yuan, J., & Liberman, M. (2010). F0 declination in English and Mandarin broadcast news speech. *Proceedings of Interspeech 2010 (Eleventh Annual Conference of the International Speech Communication Association)*, 134–137.

Zheng, X. (2006). Voice quality variation with tone and focus in Mandarin. *Second International Symposium on Tonal Aspects of Languages (TAL 2006)*, 132–136. https://www.isca-speech.org/archive_v0/tal_2006/papers/tal6_132.pdf

Zimmermann, M., & Onea, E. (2011). Focus Marking and Focus Interpretation. *Lingua*, *121*(11), 1651–1670. http://doi.org/10.1016/j.lingua.2011.06.002