Quaderni di Linguistica e Studi Orientali QULSO
Working papers in Linguistics and Oriental Studies

Firenze University Press
https://oajournals.fupress.net/index.php/bsfm-qulso

OPEN ACCESS

Citation: M. Casentini (2025) More than a Topic Marker: The Functions of a in Mandarin Chinese. *Qulso* 11: pp. 57-76. doi: http://dx.doi.org/10.13128/QULSO-2421-7220-18662

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

More than a Topic Marker: The Functions of *a* in Mandarin Chinese

Marco Casentini
Università degli Studi di Verona (<marco.casentini@univr.it>)

Abstract:

This study investigates the role of the Chinese particle *a* as a Topic Marker (TM). Our findings show that TM-*a* is closely tied to Speaker/Writer Attitude and is context-dependent, typically marking a Topic selected from a broader set of entities. This selection often carries a contrastive-like reading, reinforcing the Producer's personal involvement. Statistical analysis reveals that Involvement is the strongest predictor, followed by Type and Attitude. Additionally, TM-*a* appears more frequently in negative statements, suggesting a mitigating function. Unlike Sentence-Final Particle (SFP) *a*, which softens the entire utterance post-factum, TM-*a* operates as a preparatory mitigation strategy, setting the interpretive stance before the predication unfolds. Following Pan (2017), who posits that TMs are located in Att° rather than Top°, we argue that TM-*a* and SFP-*a* share semantic properties, as they both function as Attitude markers, but differ in mitigation strategy. This analysis aligns with Chu (2009) and Li & Thompson (1981), supporting the hypothesis that TM-*a* is primarily used to signal personal involvement and soften negative evaluations.

Keywords: Topic Markers, Attitude Markers, Involvement Markers

1. Introduction

In recent years, an increasing number of generative studies have embraced quantitative methodologies to evaluate theoretical proposals (cf. Van Craenenbroeck and Van Koppen 2025; Samo and Si 2022). These works highlight the relevance of corpus-based and experimental data in informing formal syntax, including frameworks such as Cartography. The present study aligns with this line of research, adopting corpus data to reassess the categorial status of the particle a in Mandarin Chinese. In doing so, it contributes to ongoing efforts to bridge qualitative theoretical insights and quantitative empirical validation (cf. Samo and Si 2022 on the optionality of de in Mandarin).

The present study investigates the role of the Mandarin particle 啊*a* when it functions as a Topic Marker (TM). Although the literature has devoted considerable attention to *a* as a Sentence-Final Particle (SFP) (e.g., Chao 1968; Li and Thompson

1981; Chu 2006), its use as a TM has often been overlooked or treated as pragmatically opaque and semantically unrestricted (cf. Deng 2015).

Recent proposals within the cartographic approach to syntax (e.g., Pan 2017; Paul and Pan 2016) challenge the traditional view that TMs are located in Top°, a dedicated projection within the CP layer. In earlier accounts (e.g., Badan 2007; Paul 2005), TMs were assumed to occupy the head of TopP, while SFPs were analyzed as final complementizers in Force° or associated with other high CP projections. However, more recent work suggests that both TMs and SFPs may in fact occupy Att°, a functional projection responsible for encoding speaker or writer attitude (Pan 2017; Paul 2015).

If this unified analysis is correct, then particles like *a* – which appear both as TMs and as SFPs – should exhibit similar semantic and pragmatic properties across these two uses. This raises the central question of the present study: whether *a*, when used as a TM, shares with SFP-*a* its function as an Attitude Marker and its involvement in the mitigation of speaker stance.

To answer this question, the study adopts a corpus-based approach, drawing on authentic language data to analyze the distribution and discourse functions of TM-*a* in Mandarin. By doing so, the analysis contributes to broader discussions on the relationship between syntax, Information Structure, and speaker stance in Chinese.

The paper is structured as follows. Section 2 reviews definitions of Topic and the main features of TMs in Mandarin. Section 3 discusses previous accounts of *a* as an SFP. Section 4 presents the methodology of the corpus analysis. Section 5 outlines the findings, which are discussed in Section 6. Section 7 concludes the study and outlines directions for further research.

2. Topic markers

The notion of Topic plays a central role in Information Structure, typically alongside Focus and Givenness (cf. Féry e Ishihara 2016). However, there is considerable variation in how the term is defined.

For instance, Lambrecht (1994: 118) defines Topic as "the thing which the proposition expressed by the sentence is about". However, the author himself acknowledges that this definition may be too narrow, as it primarily aligns with the grammatical Subject. Yet, as Lambrecht (1994) argues, the Subject is not necessarily the Topic, nor is the Topic necessarily the Subject, as illustrated by the following examples:

(1) Topic = Subject
[The children]$_{\text{Topic/Subject}}$ went to school.
(Lambrecht 1994: 121)

(2) Topic ≠ Subject
[That kind of thing]$_{\text{Topic}}$, [I]$_{\text{Subject}}$ don't think I'll ever do.
(Akaruese 2015: 153)

As shown in (1) and (2), the Topic can either coincide with the relevant Subject (the former) or with the Object (the latter). Furthermore, from a semantic perspective, the Topics in (1) and (2) are closely linked to the argument structure of the verb with which they are associated. In some cases, they must be co-indexed with a gap (or a deleted copy) in the sentence, at least in languages like English. For instance, the sentence in (2) can be represented with the following underlying structure in (2'):

(2') [That kind of thing], I don't think I'll ever do ~~[that kind of thing]~~

However, not all Topics behave in the same way. According to (Chafe 1976), certain types of Topics are not fully integrated into the predicate-argument structure of the clause they are associated with. These Topics serve as "scene-setting" expressions, providing a spatial, temporal, or individual framework within which the main predication takes place. This phenomenon, referred to by Chafe (1976) as Chinese-style Topic, has been described under various terms in the literature, including Hanging Non-Clausal Topic (Deulofeu 2008), Dangling Topic (Shi 2000), and Limiting Topic (Carella 2015), among others. An example is given in (3), in which the main predication holds true only within the "setting" established by the Framing-Topic, that is, within the context of 'the typical family today':

(3) [The typical family today]$_{\text{Topic}}$, the husband and the wife both work.
(Deulofeu 2008:218)

In the present analysis, we will focus exclusively on referential Topics – that is, elements that indicate what the sentence is about, in Lambrecht's (1994) terms.

Since the Topic-like constituent in (3) serves to establish a setting for the main clause to hold (cf. Chafe 1976), it functions as a "framing" constituent. As such, it will be disregarded, along with other similar structures, in the present study.

While Topics can be identified through their position and function within a sentence or discourse, they are not always marked overtly. However, in Mandarin Chinese, clause-initial Topics can be followed by partciles that are traditionally described as TMs, as they appear to signal the topical status of the preceding constituent.

TMs are morphemes that signal expressions referring to Topics, which can be characterized by three core features: (i) optionality, (ii) interchangeability, and (iii) lack of inherent semantic meaning (cf. Li and Thompson 1989; Gundel 1988, among others).

In this regard, consider examples (4) and (5) in Chinese in terms of optionality and interchangeability. Specifically, in example (4), Dong (2019) points out that the TM 呢*ne* is optional. Moreover, Qiang (2011) states that the sentence in (5) would be feasible with different TMs, such as *ne*, 吧*ba*, 么*me*, or *a*:

(4) Optionality
这个问题（呢），我们还要研究一下。
Zhe-ge wenti (*ne*), *women hai yao yanjiu yixia.*
this-CL issue NE 1PL still need discuss a little
'(As for) this issue, we still need to discuss it a little.'
(Dong 2019: 474)

(5) Interchangeability
小王，呢/吧/么/啊，是上海人。
Xiao Wang ne/ba/me/a, shi Shanghai-ren.
Xiao Wang NE/BA/ME/A be Shanghai-person
'(As for) Xiao Wang, he is from Shanghai.'
(Qiang 2011: 191)

Furthermore, TMs are typically not restricted to marking a specific syntactic relation or semantic role (Gundel 1988). For instance, examining examples (4) and (5) above, we can observe that in (5), the relevant Topic corresponds to the Subject, whereas in (4), it corresponds to the Direct Object of the sentence.

In Mandarin Chinese, several particles – such as 啊/呀*a/ya*[1], *ba*, *ma*, *me*, and *ne* – have been traditionally described as TMs. Among these, *ne* is one of the most extensively studied, particularly for its use in contrastive constructions (cf. Li 2006; Huang 2007; Chu 2009; Qiang 2010, 2011; Deng 2015).

However, it is important to note that in some studies (i.e., Qiang 2010, 2011; Deng 2015), the authors frequently group together cases where these "marking morphemes" function not only as TMs but also as markers of Vocative constituents. As a matter of fact, it has been shown that both Vocatives and Topics can be marked by the same particles (e.g., in Japanese and Chinese). However, as demonstrated by Maynard (2002), a distinction must be made, as Vocatives serve to call someone's attention, while Topics identify what the sentence is about. Furthermore, the presence of one does not exclude the presence of the other (cf. Hill 2014).

Considering the above, the following paragraphs will focus exclusively on the use of *a*, *ba*, *ma*, and *ne* as TMs, outlining the state of the art on TMs in Chinese. This review will highlight the existing gaps in research on TM-*a*, which is the central focus of the present study.

Let us first introduce the results of previous studies on different TMs, starting with *ne*.

According to Chu (2009) and Deng (2015), the TM *ne* is used to mark contrastive Topics in Chinese. For instance, Chu (2009) analyzes the sentence in (6) as expressing contrast through the use of *ne*. However, this interpretation is not necessarily shared by all native speakers or scholars, and alternative analyses may be possible depending on the context.[2] More specifically, Chu (2009) argues that *ne* appears in sentences where the relevant Topic is contrasted or compared with another Topic in a different sentence or with an entity in the context. For instance, in example (6), even if the sentence is presented out of the blue, the presence of the TM *ne* makes it clear that 'that book' is being contrasted with something else in the context:

(6) 那本书呢，你看完了没有?

Na-ben shu	*ne,*	*ni*	*kan-wan-le*	*mei*	*you?*	
that-CL book	NE	2SG	read-finish-LE	NEG	have	

'That book, have you finished reading?'

(Chu 2009: 13)

Obviously, the comparison is acceptable only if it occurs in a context where the contrast is meaningful, both in relation to the Topic itself and to the predication concerning the Topic (Chu 2009). For instance, a sentence like (7b) would be considered acceptable because, within the discourse, the speaker first states in (7a) that 'this book' has just been borrowed, implying that the listener has not yet finished reading it. Consequently, in (7b), the speaker asks whether

[1] I am grateful to an anonymous reviewer for pointing out the relevance of *ya*. This particle is generally regarded as a phonological variant of *a*, often surfacing after high vowels due to phonotactic constraints. Although some variation exists in terms of speaker gender or register (cf. Chang 2019), most studies treat *a* and *ya* as functionally equivalent, both as TMs and SFPs (see e.g., Qiang 2011; Deng 2015). For this reason, *ya* is typically subsumed under *a* in analyses of discourse particles. The present study follows this convention and focuses on *a*, with the understanding that many observations likely extend to *ya* as well.

[2] I thank the reviewer for raising this important point. The examples cited from Chu (2009) and Qiang (2011) reflect the analyses proposed by the original authors. While these accounts have been influential, their interpretations may not be universally accepted, and further investigation is warranted to determine the extent to which these uses of *ne* and other particles are understood consistently by native speakers.

the listener has finished reading 'that book', making the contrast between the two books contextually relevant:

(7) a. 这本书，刚借来（还没看）。
Zhe-ben shu gang jie- lai (hai mei kan).
that-CL book just borrow-come still NEG read
'This book has just been borrowed (hasn't been read yet).' (Chu 2009: 14)

b. 那本书呢，你看完了没有?
Na-ben shu ne, ni kan-wan-le mei you?
that-CL book NE 2SG read-finish-LE NEG have
'That book, have you finished reading?'
(Chu 2009: 14)

On the contrary, (8b) would be judged as unacceptable, as the contrast – which is obligatorily triggered by the presence of the TM *ne* – does not create a meaningful contrast with 'this beef' in (8a):

(8) a. 这种牛肉（你）刚买来（还没吃）。
Zhe-zhong niu-rou (ni) gang mai-lai (hai mei chi).
that-CL beef-meat 2SG just buy-come still NEG eat
'This beef has just been bought (hasn't been eaten yet).' (Chu 2009: 14)

b. #那本书呢，你看完了没有?
#Na-ben shu ne, ni kan-wan-le mei you?
#that-CL book NE 2SG read-finish-LE NEG have
#'That book, have you finished reading?'
(Chu 2009: 14)

Chu's (2009) claims are further supported by Deng (2015), who states that the Topic marked by *ne* must be semantically linked to the co-text or context. Furthermore, unlike other TMs discussed in the following sections, *ne* is not associated with the speaker's attitude (cf. Deng 2015). Thus, as already mentioned, it serves purely to mark contrast, either with something previously stated in the discourse or with an element present in the physical context.

Another TM that has been extensively studied is 嘛*ma*.[3] According to Qiang (2010), this TM is used to mark a Topic for which, in a given context, the relevant predication, or the Topic itself, is obvious. For instance, in example (9) below, within the socio-cultural context in which the sentence is produced, it is obvious that if you are a man, you can only rely on yourself. Therefore, *nanren* ('men') is marked by *ma*:

(9) 男人嘛，就是得靠自己。
Nan-ren ma, jiu shi dei kao ziji.
masculine-person MA only be need relay on self
'Men, they can only relay on them-self.'
(Qiang 2010: 59)

[3] Notice that we refer to the TM 嘛 (*ma*), which, despite its phonological similarity, is semantically distinct from the sentence-final particle 吗 (*ma*), used in yes-no questions.

Furthermore, Qiang (2011) and Deng (2015) observe that *ma* can mark Topics followed by a negative evaluation (10) and can also appear in imperative clauses. In the latter context, Qiang (2011) argues that *a*, *ne*, *ma*, and *ba* are all acceptable alternatives, as shown in (10) and (11). This classification remains tentative, since speakers' acceptability judgments often diverge and further empirical validation is needed:[4]

(10) 这本书吧/啊/嘛/呢，写得一点儿也不深刻。
Zhe-ben shu ba/a/ma/ne, xie-de yi-dianr ye bu shenke.
this-CL book BA/A/MA/NE write-DE a little also NEG deep
'This book, is not written in a profound way at all.'
(Qiang 2011: 199)

(11) 这瓶酒吧/啊/嘛/呢，（你）先全喝！
Zhe-ping jiu ba/a/ma/ne, (ni) xian quan he!
this-CL$_{(BOTTLE)}$ alcohol BA/A/MA/NE 2SG first complete drink
'This bottle of alcohol, you go ahead and drink it all!'
(Qiang 2011: 198)

Similar to TM *ma*, there is TM *ba*, which, as already shown in examples (10) and (11) above, can mark the Topic of an imperative clause or a Topic for which the speaker or writer expresses a negative judgment (Qiang 2011; Deng 2015). Furthermore, topics marked by both *ma* and *ba* must be contextually connected to the broader discourse in which they appear (cf. Deng 2015).

However, while the topic marked by *ma* must be anaphorically linked to another constituent in the text, *ba* can also refer to a new entity, as long as it remains interpretable within a given context (whether social, physical, or otherwise) (cf. Deng 2015).

The last TM we will discuss here is *a*, which, unlike the previously presented TMs, does not appear to have specific restrictions or properties. In fact, Deng (2015) argues that the Topic marked by *a* is not necessarily anaphoric, contextually connected, or indicative of the speaker's or writer's attitude toward the Topic, as summarized in Table 1 below:

	A	*Ne*	*Ba*	*Ma*
Anaphoric reading	-	+	-	+
Context-dependent	-	+	+	+
Sp./Wr.'s Attitude	-	-	+	+

Table 1. Functions of Chinese TMs

Furthermore, Qiang (2011) shows that *a* can mark Topics in various contexts, including imperative sentences and sentences expressing a negative judgment about the Topic, as already exemplified in (10) and (11) above, where *a* is a viable option alongside *ne*, *ma*, and *ba*.

[4] I thank an anonymous reviewer for urging a more critical stance toward Qiang's (2011) classification. The examples in (10) and (11) are cited verbatim from Qiang's study; their acceptability has not been systematically tested in controlled experiments, and fieldwork indicates that judgments are heterogeneous. A comprehensive survey of speaker intuitions lies beyond the scope of the present paper but is an obvious avenue for future research.

However, Qiang (2011) suggests that there is one case in which only *a* can function as a feasible TM, while the other TMs are not acceptable: the Topic of exclamative clauses. According to the author, an exclamative sentence with a Topic marked by *a* is acceptable (12a), whereas marking it with other TMs results in unacceptability (12b):

(12) a. 祖国啊，多么伟大啊！
Zuguo *a,* *duome* *weida* *a!*
Motherland A how mighty A
'Motherland (China), how mighty it is!' (Qiang2011:199)

b. *小王呢/吧/嘛，多么调皮啊！
**Xiao Wang* *ne/ba/ma,* *duome* *tiaopi* *a!*
*Xiao Wang NE/BA/MA how naughty A
(Qiang 2011: 199)

Thus, considering Deng (2015) and Qiang (2011), on the one hand, the TM *a* does not seem to have specific restrictions, as it does not need to be linked to the context and does not necessarily reflect the attitude of the speaker/writer. On the other hand, in exclamative sentences, the Topic can only be marked by the TM *a* (Qiang 2011).

It should be noted that all these TMs also function as Sentence-Final Particles (SFPs) in Chinese, where they serve different semantic and pragmatic functions in discourse. In fact, the particle *a* can also be used as an SFP to mark exclamative sentences, and it is indeed present in the examples given in (12a-b).

This specific point will be addressed in the next two sections, where we will examine the syntactic structure of TMs and the use of the particle *a* as an SFP.

2.1 Syntactic analysis of Topic Markers

According to generative theory (Chomsky 1981 and subsequent works), the clause universally consists of three major phrases: the Verb Phrase (VP), the Inflectional Phrase (IP) – often referred to as Tense Phrase (TP) in the literature – and the Complementizer Phrase (CP), hierarchically structured as shown in (13):

(13) $[_{CP}\ [_{TP}\ [_{VP}\ ]]]$

Following standard assumptions, the VP is the layer where *theta*-role assignment takes place; the IP/TP layer is responsible for the licensing of formal features such as case and agreement; and the CP is the layer where illocutionary force is encoded and discourse-related categories are hosted.

In Rizzi's (1997) seminal work, it is argued that Topics are hosted in the Spec position of specTopP (Topic Phrase) projections within the left periphery of the sentence, meaning they are located within the CP layer.

As for TMs, studies such as Badan (2007), Pan (2014; 2015), and Paul (2002; 2005; 2014; 2015), analyze them as heads of a TopP in Chinese.

Thus, a sentence like (11) above, reproduced as (14a) below, should have the structure shown in (14b), where the TMs occupy the head of TopP (Top°), the DP *zhe ping jiu* ('this bottle of alcohol') is in Spec,TopP, and the rest of the sentence is represented as a TP in the Complement position of TopP (Compl,TopP):

(14) a. 这瓶酒吧/啊/嘛/呢，（你）先全喝！
Zhe-ping jiu ba/a/ma/ne (ni) xian quan he!
this-CL$_{(BOTTLE)}$ alcohol BA/A/MA/NE 2SG first complete drink
'This bottle of alcohol, you go ahead and drink it all!'
(Qiang 2011: 198)

b. [$_{TopP}$ *Zhe ping jiu* [$_{Top°}$ *ba/a/ma/ne* [$_{TP}$ (*ni*) *xian quan he*!]]] (Qiang2011:198)

However, according to Pan (2017), the fact that these particles can function not only as TMs but also as SFPs or even as Focus Markers challenges their analysis as heads of TopP.

To address this, Pan (2017) proposes that these particles are actually heads of a higher functional projection, namely Attitude Phrase (AttP) (cf. Paul 2015; Pan 2015; Pan and Paul 2016). In this view, the relevant particle occupies Att°, the head of AttP, while the associated DP – typically a Topic – occupies Spec,AttP. Crucially, AttP itself is embedded within the Left Periphery, and specifically occupies Spec,TopP, as shown in (15). This configuration allows the analysis to capture the flexible distribution of these particles across different discourse configurations (e.g., Topic, Focus, or sentence-final position), while maintaining their core attitudinal function.

Thus, the structure in (14b) can be revised as in (15), where the so called "TMs" occupy the head of AttP (Att°), the DP *zhe ping jiu* ('this bottle of alcohol') is in Spec,AttP, and AttP itself is located in Spec,TopP. The rest of the sentence is represented as a TP in the complement position of TopP:

(15) [$_{TopP}$ [$_{AttP}$ *Zhe ping jiu* [$_{Att°}$ *ba/a/ma/ne*]] [$_{Top°}$ [$_{TP}$ (*ni*) *xian quan he*!]]]

Thus, this proposal analyzes these particles as attitude markers, accounting for their ability to appear in different positions within the sentence (following a Topic or a Focus constituent, or functioning as SFPs) while maintaining similar semantic properties.

This perspective also challenges Deng's (2015) classification of *a* as a TM unrelated to the speaker's or writer's attitude, even though Qiang (2011) had already noted that TM *a* can mark the Topic of a sentence expressing a negative judgment (see example (10) above, where *a* is a viable option).

Furthermore, if Pan's proposal is correct, we should expect *a* to mark Topics – or, more precisely, attitude-mark DPs that happen to be Topics – in the same contexts where *a* functions as an SFP, since in both cases, *a* should occupy the head of AttP.

Given that *a*, like the other particles, functions as an attitude marker according to Pan (2017), in the present paper, I will continue to use the terms TM and SFP solely for the sake of clarity: TM will refer to cases where *a* acts as an attitude marker for a DP that serves as a Topic, while SFP will indicate instances where *a* functions as an attitude marker for an entire sentence.

Thus, our research question is:

- If *a* is in Att°, whether marking the relevant Topic (TM) or the entire sentence (SFP), should its properties as a TM be the same (or at least similar) to those of SFP-*a*?

Before outlining the methodology and results of this study, it is necessary to first examine the function of SFP *a*.

3. Sentence-Final Particle A

According to Chao (1968), the particle *a* has at least ten different functions. However, some of these are not strictly related to *a* as an SFP. In this section, we will focus only on those functions where *a* operates as an SFP.

Following Chao's (1968) classification, SFP-*a* can be used to: (i) ask questions (16a), (ii) ask for confirmation (16b), (iii) give an order (16c), (iv) mark impatience (16d), (v) give a warning (16e), (vi) mark exclamative sentences (17a), and (vii) express reminders (17b):

(16) a. 你明天出去不出去啊?
Ni *mingtian* *chu-qu* *bu* *chu-qu* *a?*
2SG tomorrow exit-go NEG exit-go A
'Are you going out or not tomorrow?'

b. 你不去啊?
Ni *bu* *qu* *a?*
2SG neg exit A
'Aren't you going out?'

c. 走啊！咱们都走啊！
Zou *a!* *Zanmen* *dou* *zou* *a!*
walk A 1PL.INCL DOU walk A
'Walk! Let's walk!'

d. 我并没做错啊！
Wo *bing mei* *zuo-cuo* *a!*
1SG NEG do-wrong A
'I didn't do it wrong!"

e. 这个人的话是靠不住的啊！
Zhe-ge *ren* *de* *hua* *shi* *kao-bu-zhu* *de* *a!*
this-CL person DE word be rely on-NEG-RES DE A
'What this person says is not reliable!'

(Chao 1968)

(17) a. 我就跑啊，跑啊，跑啊！
Wo *jiu* *pao* *a,* *pao* *a,* *pao* *a!*
1SG just ran A ran A ran A
'I ran and run and run!'

b. 本来你也知道啊，也用不着再说啊。
Benlai *ni* *ye* *zhidao* *a,* *ye* *yong-bu-zhao* *zai* *shuo* *a.*
originally 2SG also know A also use-NEG-RES again speak A
'As you already know, and I don't have to say it again…'

(Chao 1968)

However, as pointed out by Li and Thompson (1981), what SFP-*a* does in sentences like (16a-f) is not to mark the sentence type *per se* (e.g., questions, requests for confirmation, etc.), as these can already function independently. Instead, SFP-*a* primarily "reduces the forcefulness of the message" (Li and Thompson 1981: 313).

Similarly, Chu (2006) agrees with Li and Thompson (1981) and further argues that SFP-*a* indicates the speaker's personal involvement or concern in what is being stated.

Finally, Chao (1968) also claims that the particle *a* is used to mark list elements (18a) or signal a pause (18b). However, in both cases, *a* does not appear to function as an SFP but rather as a TM – especially considering that *yaoshi ni bu ken* ('if you are not willing to') in (18b) acts as a frame-setting Topic:

(18) a. 什么天啊，地啊，日啊，这些字都会写。
Shenme Tian a, Di a, Ri a, zhi-xie zi dou hui xie.
such us Tian A Di A Ri A this-CL character DOU can write
'Tian, Di, Ri, (he) is able to write all these characters.'

b. 要是你不肯啊，那我就不管了。
Yaoshi ni bu ken a, na wo jiu bu guan le.
if 2SG NEG will A then 1SG JIU NEG bother LE
'If you are not willing to, than I won't bother about it.'

(Chao 1968)

Thus, considering our research question, we should expect TM-*a* to mark Topics in the same types of sentences where SFP-*a* occurs, functioning as an attitude marker that serves one of the functions listed above, while also carrying a mitigating or involvement-related effect.

4. Method

To answer our research question, a corpus-based analysis was conducted using the Chinese Web Corpus (Jakubíček et al. 2013), accessed through the SketchEngine platform (Kilgarriff et al. 2014). The decision to rely on corpus data for identifying attested uses of *a* reflects a broader trend in generative syntax toward the empirical grounding of theoretical claims (cf. Van Craenenbroeck and Van Koppen 2025; Samo and Si 2022). This approach allows for the systematic testing of syntactic hypotheses, following the logic of falsifiability and pattern generalization that has characterized recent quantitative work in the field.

To identify Topics marked by *a*, the Corpus Query Language was used to extract all occurrences of the particle *a* that were preceded by a noun and followed by a comma, using the following query:

(19) [pos="N.*"] [word="啊"] [word="，"]

Thus, only tokens matching this exact sequence – "noun + 啊 + ," – were considered for analysis.

From a total of 10,059 occurrences, cases where *a* was used to mark vocatives, framing-Topics, or Focus, as well as instances where *a* functioned as a proper SFP or an interjection, were excluded. After filtering, only 60 occurrences of *a* as a TM remained for analysis.

For each instance, we tagged the function of *a* as a TM (Type), the Producer's attitude (Attitude)[5], and the presence or absence of personal involvement markers (Involvement).

[5] Since the corpus data may include cases of direct speech transcription, the term "Producer" will be used from this point onward to refer to both Speakers (in the case of written direct speech) and Writers (in the case of originally written texts).

Since annotation was based on a mixed method, both deductive and data-driven (Stefanowitsch 2020), the specific types of functions served by *a* as a TM will be presented in the next section. Meanwhile, the speaker's Attitude was categorized as follows: Positive (20a), when a favorable judgment was expressed; Neutral (20b), when the utterance merely reported facts without personal evaluation; and Negative (20c), when a negative judgment was conveyed. The classification was determined based on the full sentence:

(20) *a.* 政治啊，历史，都考得不错。
Zhengzhi a, lishi, dou kao-de-bu-cuo
politics A history DOU test-DE-NEG-bad
'Politics (exam), history (exam), I did well.'

b. 这些东西啊，都是我男人从山地下跳上来的。
Zhe-xie dongxi a, dou shi wo nanren cong shan dixia tiao-shang-lai de.
this-CL thing A DOU be 1SG man from mountain foot pick-up-come DE
'These thing, they have been all pick up by my man from the foot of the mountain.'

c. 我这张口啊，让人扫兴。
Wo zhe zhang-zui a, rang ren saoxing.
1SG this open-mouth A let person disappointed
'My words, they let people disappointed.'

Furthermore, personal involvement markers were identified based on Timmi (2014). Specifically, the presence of intensifiers, superlatives, or strongly evaluative adjectives was classified as "presence of involvement" ("Yes" in the figures below). Conversely, their absence was classified as "absence of involvement" ("No" in the figures below).

Finally, to examine category distributions and assess which variable exerts the strongest influence, we performed chi-square tests (cf. Field, Miles, and Field 2012; Agresti 2018; Sharpe 2015).

5. Data analysis

One of the first findings from the data analysis is that TM-*a* serves at least three distinct functions: (i) it can mark a Topic that represents an entity selected from a larger group of entities ("Selection"); (ii) it can mark a Topic for which the Producer expresses a purely personal judgment ("Standpoint"); (iii) it can mark a Topic that refers to a group consisting of multiple entities ("Grouping"). Additionally, some instances were found where *a* was used to mark other individual functions that did not form a uniform category (4 cases), which were categorized under "Other".

Among these functions, TM-*a* most frequently serves as a Selection marker (30 out 60: 50% of occurrences) or as a Standpoing marker (21 out of 60: 35%), while it is less commonly used for Grouping (5 out of 60: 8.3%) or for "Other" functions (4 out of 60: 6.7%). These data are also represented in Figure 1 below:

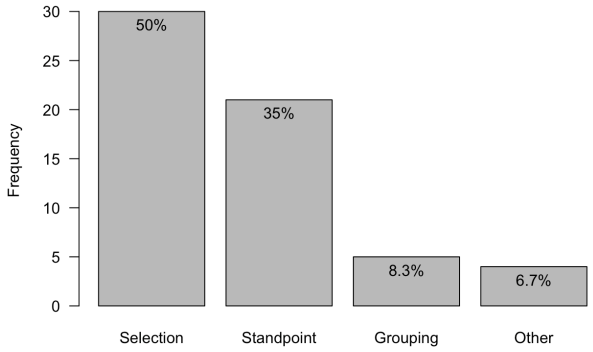

Figure 1. Distribution of TM-*a* by Discourse Function Type.

Let us now examine an example for each of these types, starting with a case of "Selection", shown in (21), followed by its relevant description:

(21) 原来他在小院里面，生活成本很低，院里种的菜啊，水啊，什么的都挺方便。
Yuanlai ta zai xiao yuan limian, shenghuo chengben hen di,
Actually 3SG in small compound inside life cost very low
yuan li zhong de cai a, shui a,
compound in CL DE vegetable A water A
shenme de dou ting fangbian.
what DE DOU very convenient
'Actually, he lives in a small compound, life costs (there) are low, **the type of vegetables in the compound**, **the water**, it is all very convenient.'

The Producer states that someone (i.e., 'he' in the example) lives in a countryside compound and claims that the cost of living there is low. They then provide an exemplary list of expenses, such as those for 'vegetables' and 'water', all marked by the TM-*a*, emphasizing that all of them are very convenient.

Furthermore, it should be noted that the relevant predication carries a positive connotation, and the presence of the adverbial modifier *ting* ('very, quite') further indicates the Producer's involvement. The use of intensifiers, as observed in *ting*, can signal an evaluative stance (cf. Timmi 2014).

Let us now examine a case of Standpoint. In (22), the Producer – an elementary school student – writes about the Great Wall of China in their essay and states that 'the Great Wall is so fucking long', expressing a strong evaluative stance. This stance is marked by the presence of both an imprecation and the use of an intensifier, *zenme name* ('how so'), before the adjectival predicate *chang* ('long').

Furthermore, from the text extracted it can be assumed that the Producer was already writing about their trip to the Great Wall, making it a referential Topic that anaphorically refers to something previously mentioned in the context:

(22) 长城啊，长城tmd 怎么那么长……
Changcheng a, *Changcheng tmd zenme name chang…*
Great Wall A Great Wall fucking how so long…
'**The Great Wall**, Great Wall is so fucking long…'

Finally, let us examine a case where TM-*a* is used as a "Grouping" marker. As shown in the context provided for (23), the Producer reflects on different aspects of their past life – such as when 'love and hate were intertwined' or when 'class was as lively as a party'. All these aspects are encapsulated by the phrase *na-ge shihou* ('that time'), which serves as a general reference grouping together the various experiences mentioned. Finally, once again, we can observe the Producer's attitude and involvement in this case, as indicated by the presence of intensifiers such as *zhende* ('really'):

(23) 这是发生在那一段让我爱恨交织、哭不出笑不来、浑浑噩噩的日子。班里每天都想开PARTY那么热闹。从早上7:20一直持续到晚自习下课。那个时候啊，真的是快乐无忧。
Zhe shi fasheng zai na yi duan rang wo ai hen jiao zhi, ku bu chu xiao bu lai, hun hun e e de rizi. Ban li mei tian dou xiang kai PARTY name renao. Cong zaoshang 7:20 yizhi chixu dao wanzixi xiake.
Na-ge shihou a, *zhende shi kuaile wu-you.*
That-CL time A really be happy NEG-care
'This happened during that period when love and hate were intertwined for me, when I couldn't cry or laugh, and lived in a daze. Every day in the class was as lively as a party. It went on from 7:20 in the morning until the end of the evening self-study session. **That time,** it was truly happy and carefree.

Excluding "Other" from the analysis[6], we performed a chi-square test to determine whether one function is predominant over the others.

The results indicate a significant difference between the three identified functions ($\chi^2(2) = 17.2$, $p < .001$). However, the primary distinction lies in the fact that Grouping occurs significantly less than Selection ($p < .001$) and Standpoint ($p < .01$), while no significant difference is observed between Selection and Standpoint ($p = .62$).[7] This suggests that TM-*a* is primarily used to mark a Topic that represents an entity selected from a group (Selection) or to mark a Topic for which the Producer expresses a personal judgment (Standpoint).

Another key finding of this analysis is that, in most cases, the predication following the Topic marked by *a* expresses a negative evaluation (30 out 60: 50%). In contrast, positive and neutral evaluations are less common (16 out of 60: 26.7%, and 14 out 60: 23.3% respectively), as illustrated in Figure 2:

[6] "Other" is a small category encompassing various Types that could not be meaningfully grouped together and was excluded for the sake of clarity in results and analysis.

[7] For the comparisons between groups, the p-values were adjusted using the Bonferroni method to account for multiple comparisons.

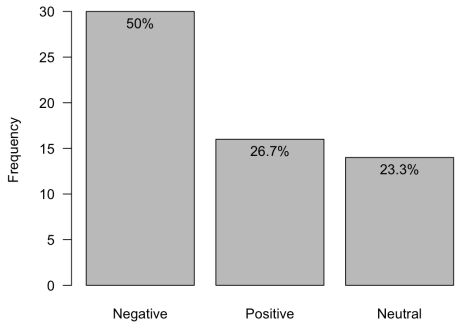

Figure 2. Distribution of TM-*a* by Speaker Attitude.

Importantly, a chi-square test revealed a significant difference among the categories of Attitude, $\chi^2(2) = 10.4$, $p = .005$. Specifically, Negative occurs significantly more often than Positive ($p = .04$) and Neutral ($p = .01$), while no significant difference is observed between Positive and Neutral ($p = 1$).[8]

If we examine the distribution of Attitude with respect to Type (Figure 3), we observe that most occurrences fall under "Selection" and "Standpoint" with negative predications. This is not unexpected, given that Selection and Attitude are the most frequently occurring Types, and negative predications are the most common:

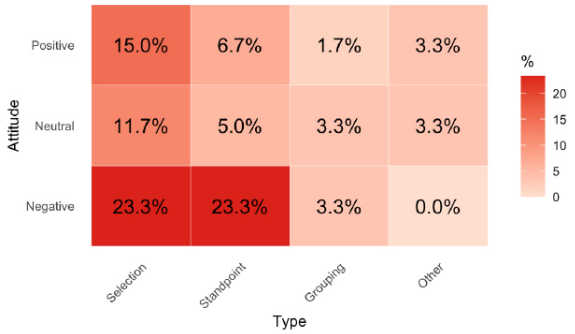

Figure 3. Heatmap of TM-*a* Functions Cross-Tabulated by Speaker Attitude.[9]

However, when testing for significant correlations, the chi-square test revealed no significant difference between the combinations for Type and Attitude ($\chi^2(6) = 7.4$, $p = .288$).

As for Producer's involvement, the analysis reveals that most predications following Topics marked by *a* contain explicit constituents that indicate the Producer's involvement (46 out of 60: 76.7%), as illustrated in Figure 4:

[8] For the comparisons between groups, the p-values were adjusted using the Bonferroni method to account for multiple comparisons.

[9] Raw scores for the heatmap (Figure 3): Positive: Selection = 9, Standpoint = 4, Damping = 1, Other = 2; Neutral: Selection = 7, Standpoint = 3, Damping = 2, Other = 2; Negative: Selection = 14, Standpoint = 14, Damping = 2, Other = 0.

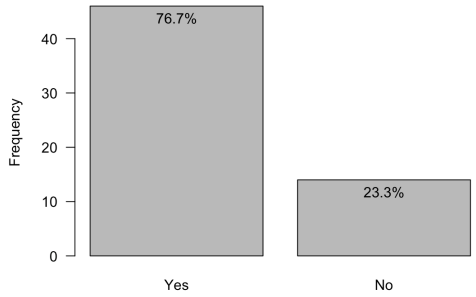

Figure 4. Distribution of TM-*a* by Speaker Involvement.

Furthermore, a chi-square test revealed a significant difference among the categories of Involvement ($\chi^2(1) = 17.1$, $p < .0001$), indicating that when the Topic is marked by TM-*a*, the following predication is more likely to reflect Producer's involvement.

Regarding the distribution of Involvement markers across Attitude, Producer's involvement appears most frequently in Negative (23 out of 46: 52.3%) or Positive (15 out of 46: 31.8%) predications. However, at the same time, most cases of non-explicit Involvement occur in Negative Attitude statements (30 out of 60). Furthermore, Positive statements always include explicit involvement markers, as illustrated in Figure 5:

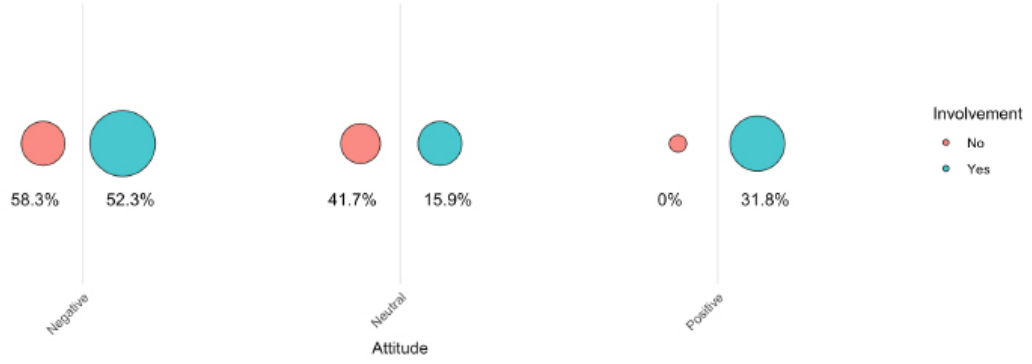

Figure 5. Co-occurrence of Speaker Attitude and Involvement (Bubble Chart.)

Importantly, a chi-square test confirmed that these differences are statistically significant ($\chi^2(2) = 6.8$, $p = .033$).

Regarding the distribution of Involvement markers across Type[10], as shown in Figure 6, most instances of predications without an Involvement marker occur when TM-*a* marks a Selection (i.e., selecting an entity from a larger group; 20 out 56: 45.5%) or a Topic for which the Producer expresses a judgment (Type: Standpoint; 20 out 56: 45.5%):

[10] Since "Other" is a small category encompassing various Types that could not be meaningfully grouped together, it was excluded from the present argumentation for the sake of clarity in results and analysis.

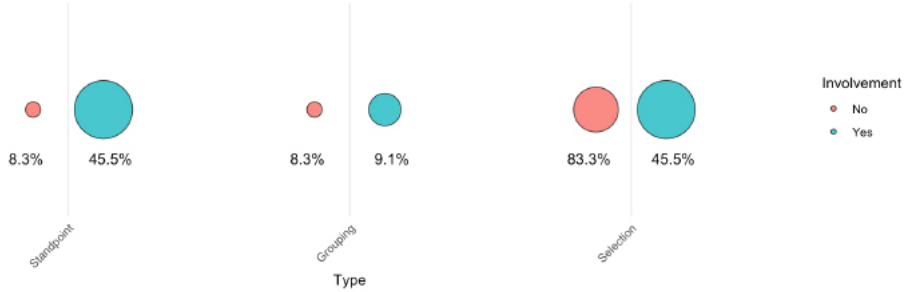

Figure 6. Co-occurrence of TM-*a* Functions and Involvement (Bubble Chart).

The chi-square test revealed that these differences are marginally statistically significant ($\chi^2(2) = 5.9$, $p = .049$).

6. Discussion

Comparing the p-values for each variable, we observe that all the variables considered are statistically significant. However, when comparing the magnitude of the effects (Table 2), Involvement emerges as the most influential variable, followed by Type and Attitude:

Variable	**P-value**
Involvement	< .0001
Type	< .001
Attitude	= .005

Table 2. Magnitude of the effects.

These data suggest that TM-*a* is primarily used to mark a Topic for which the Producer makes a personal predication. Furthermore, this personal predication typically refers to an item selected from a broader set of entities, which could be interpreted as having a contrastive-like reading.

For instance, in example (21) above – repeated as (24) below for clarity – the Producer not only selects specific "costs" from the broader category of "living costs", but also expresses a personal judgment, implicitly contrasting these costs with others in different contexts:

(24) 院里种的菜啊，水啊，什么的都挺方便。
Yuan li zhong de cai a, shui a,
compound in CL DE vegetable A water A
shenme de dou ting fangbian.
what DE DOI very convenient
'The type of vegetables in the compound, the water, it is everything very convenient.'

Furthermore, the data in Figure 2 indicate a tendency for TM-*a* to be used in contexts where the Producer expresses a negative evaluation of the relevant Topic.

These findings are also visually confirmed by the plot in Figure 7, which shows that Involvement is present across all subgroups (blue balloons), while its absence (red balloons) is limited to only a few cases. Specifically, non-Involvement appears only in Selection with Neutral and Negative Attitude, as well as in Grouping and Standpoint with Negative Attitude – accounting for 4 out of 9 subgroups:

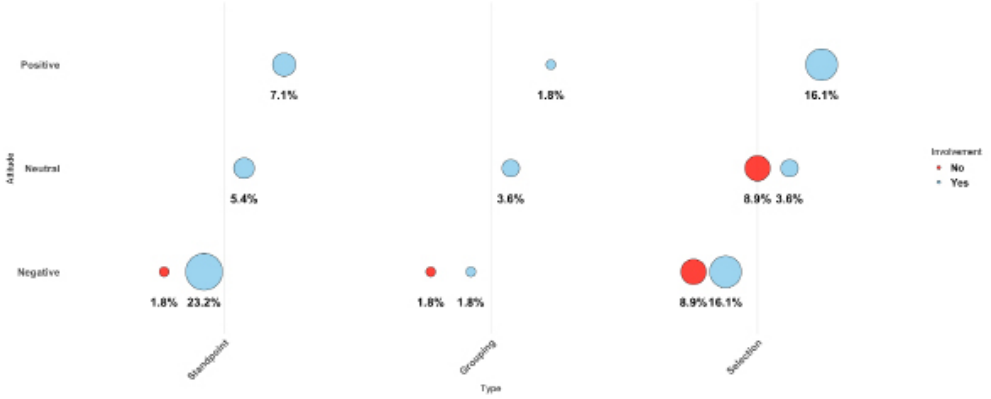

Figure 7. Distribution of Involvement.[11]

Furthermore, in line with the previously presented data, the highest number of instances occurs in Selection with personal involvement markers and either Negative or Positive Attitude (both 16.1%), as well as in Standpoint with personal involvement markers and Negative Attitude (23.2%).

These findings further support the analysis that TM-*a* is primarily used to mark a Topic for which the Producer expresses a personal judgment, making neutral Attitude and the absence of personal involvement markers relatively infrequent.

Comparing these findings with previous claims on SFP-*a* (see §3), notable similarities emerge. Specifically, our data align with Chu (2009), demonstrating that both TM-*a* and SFP-*a* are associated with the Producer's personal involvement.

Additionally, following Li and Thompson's (1981) analysis, the presence of TM-*a* in instances where the Producer expresses a negative statement about the Topic may serve a mitigating function, softening the negative evaluation.

In this regard, it can be hypothesized that the difference between SFP-*a* and TM-*a* lies in how the Producer sets the mitigation strategy. While SFP-*a* mitigates the entire sentence by appearing at the end of the clause, TM-*a* functions as a preparatory strategy, signaling the Producer's stance before the predication unfolds.

This is not unexpected, as other elements in Att° serve similar preparatory functions. For instance, as Casentini (2022) shows, the Chinese Discourse Marker *na* at the beginning of a sentence acts as an instruction-giver, signaling to the hearer that what follows should be connected to previous discourse. Similarly, TM-*a* appears to function in parallel, providing a cue that prepares the hearer/reader by indicating that the upcoming statement is a personal judgment, reflecting the Producer's Attitude and Involvement, rather than a neutral or general assertion.

[11] Raw scores for the bubble plot (Figure 7), (Involvement: Yes) Negative: Grouping = 13, Selection = 1, Standpoint = 9, Neutral: Grouping = 3, Selection = 2, Standpoint = 2, Positive: Grouping = 4, Selection = 1, Standpoint = 9; (Involvement: No) Negative: Grouping = 1, Selection = 1, Standpoint = 5; Neutral: Grouping = 0, Selection = 0, Standpoint = 5; Positive: Grouping = 0, Selection = 0, Standpoint = 0.

However, unlike SFP-*a*, TM-*a* is positioned immediately after the Topic, ensuring that this interpretive cue is provided at the very beginning of the utterance.

This proposal also accounts for Qiang's (2011) observation that in exclamative sentences, TM-*a* is the only viable option. Considering that exclamative clauses indicate a violation of the Speaker's expectations (Rett 2011), personal involvement is inherently present in these sentences. Therefore, in such cases, TM-*a* serves to instruct the hearer that what follows is a personal judgment of the Speaker.

Thus, our findings support a positive answer to our research question, showing that TM-*a* and SFP-*a* are not semantically distinct, but differ in their mitigation strategy: preparatory for the former and (post-)reparatory for the latter.

7. Conclusions

The present study aimed to clarify the role of *a* as a TM, which previous literature has described as a non-restricted TM (unlike other TMs such as *ba*, *ma*, and *ne*). However, our data challenge this view, showing that Speaker/Writer Attitude is indeed connected with TM-*a*—contra Deng (2015). Moreover, TM-*a* is also context-dependent, as it typically marks a Topic selected from a broader set of entities, meaning its interpretation is expected to be linked to the co-text or context in which the sentence is uttered.

Additionally, the present analysis builds on Pan (2017), who argues that TMs are not located in Top°, but rather function as Attitude markers positioned in Att°, placing TMs and SFPs on the same syntactic and semantic level. Following this framework, our research question was:

- If *a* is in Att°, whether marking the relevant Topic (TM) or the entire sentence (SFP), should its properties as a TM be the same (or at least similar) to those of SFP-*a*?

Based on our analysis, we argue that the answer is affirmative, as notable similarities exist between TM-*a* and SFP-*a*, in line with the analyses of Chu (2009) and Li and Thompson (1981). Specifically, our data show that TM-*a* is used to mark a Topic associated with Producer involvement and attitude, and can therefore function as a preparatory mitigation strategy for the following comment.

Nonetheless, it should be noted that the present analysis is based on a limited number of occurrences of TM-*a*. Although the corpus investigation yielded 10,059 instances of *a* marking a DP, only 60 were identified as proper Topics. This suggests that further research is needed to determine whether *a*, in different syntactic contexts, serves the same function(s).

Additionally, experimental studies could be conducted to further validate the present findings and assess whether the mitigating and preparatory functions of TM-*a* hold across different communicative settings.

References

Agresti, Alan. 2018. *Statistical Methods for the Social Sciences*. Harlow: Pearson.

Akaruese, Christopher U. 2015. "Topicalization and Passivisation in the English Language." *UNIZIK Journal of Religion and Human Relations* 7 (2): 152-160.

Badan, Linda. 2007. *High and Low Periphery: A Comparison Between Italian and Chinese.* PhD diss. Padova: Università degli Studi di Padova.

Carella, Giorgio. 2015. "The Limiting Topic." *Annali di Ca' Foscari. Serie occidentale* 49: 363-391.

Casentini, Marco. 2022. "The Sentence Initial Discourse Marker Nà 那 and Its Correlation with Sentence Final Particles." *L'Analisi Linguistica e Letteraria* 30 (3): 177-198.
Chafe, Wallace L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View." In *Subject and Topic: A New Typology of Language*, ed. by Charles N. Li, 25-55. New York, NY: Academic Press.
Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley-Los Angeles, CA: University of California Press.
Chu, Chauncey C. 2006. 汉语篇章语法 [*A Discourse Grammar of Mandarin Chinese*]. Beijing: Beijing yu yan da xue chu ban she.
Chu, Chauncey C. 2009. "A Contrastive Approach to Discourse Particles: A Case Study of the Mandarin Ufp Ne?." In *Contrasting Meanings in Languages of the East and West*, ed. by Dingfang Shu, and Ken Turner, 237-279. Oxford: Peter Lang.
Deulofeu, Henri-José. 2008. "Peripheral Constituents as Generalized 'Hanging' Topics." *Phantom Sentences*, 227-257.
Dong, Hongyuan. 2019. "A Semantic Analysis of -ne as a Topic Marker: A Grammaticalization Perspective". In *Proceedings of the 30th North American Conference on Chinese Linguistics (NACCL-30)*, ed. by Yuhong Zhu, Skylor E. Gomes, Junyu Ruan, Qian Wang, Seo-Jin Yang, Jinwei Ye, and Wei Zhou, 472-489. Columbus, OH: The Ohio State University.
Féry, Caroline, and Shinichiro Ishihara, 2016 [2014]. *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.
Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. Los Angeles-London-New Delhi-Singapore-Washington, DC: SAGE Publications Ltd.
Gundel, Jeanette K. 1988. "Universals of Topic-Comment Structure." *Studies in syntactic typology* 17 (1): 209-242.
Hill, Virginia. 2014. *Vocatives: How Syntax Meets with Pragmatics*. Leiden: Brill.
Huang, Xiao-You K. 2007. "Initialness of sentence-final particles in Mandarin Chinese." *Proceedings of the 21st Pacific Asia conference on language, information and computation,* 182-191.
Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlỳ, and Vít Suchomel. 2013. "The TenTen Corpus Family." In *7th International Corpus Linguistics Conference CL*: 125-127.
Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1 (1): 7-36.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents.* Cambridge: Cambridge University Press.
Li, Boya. 2006. *Chinese final particles and the syntax of the periphery.* PhD diss. Leiden: Universiteit Leiden.
Li, Charles N., and Sandra A. Thompson. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley-Los Angeles, CA: University of California Press.
Maynard, Senko K. 2002. *Linguistic Emotivity: Centrality of Place, the Topic-Comment Dynamic, and an Ideology of Pathos in Japanese Discourse*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
Pan, Victor J. 2014. "Wh-ex-situ in Mandarin Chinese: Mapping Between Information Structure and Split CP." *Linguistic Analysis* 39: 371-413.
Pan, Victor J. 2015. "Mandarin Peripheral Construals at the Syntax-Discourse Interface." *The Linguistic Review 32* (4): 819-868.
Pan, Victor J. 2017. "Optional projections in the left-periphery in Mandarin Chinese." In *Studies on Syntactic Cartography* 句法制图研究, ed. by Si Fuzhen, 216-248. Beijing: China Social Science Press.
Pan, Victor J., and Waltraud Paul. 2016. "Why Chinese SFPs Are Neither Optional nor Disjunctors." *Lingua* 170: 23-34.
Paul, Waltraud. 2002. "Sentence-Internal Topics in Mandarin Chinese: The Case of Object Preposing." *Language and Linguistics* 3 (4): 695-714.
Paul, Waltraud. 2005. "Low IP Area and Left Periphery in Mandarin Chinese." *Recherches linguistiques de Vincennes* 33: 111-134.

Paul, Waltraud. 2014. “Why Particles Are Not Particular: Sentence-final Particles in Chinese as Heads of a Split CP”. *Studia Linguistica* 68 (1): 77-115.
Paul, Waltraud. 2015. *New Perspectives on Chinese Syntax*. Berlin-München-Boston: De Gruyter Mouton.
Paul, Waltraud, and Victor J. Pan. 2016. “What You See is What You Get: Chinese Sentence-Final Particles as Head-Final Complementizers.” In *Discourse Particles: Formal Approach to their Syntax and Semantics*, ed. by Josef Bayer, and Volker Struckmeier, 49-77. Berlin-Boston: De Gruyter.
Rett, Jessica. 2011. “Exclamatives, Degrees and Speech Acts.” *Linguistics and Philosophy* 34: 411-442.
Rizzi, Luigi. 1997. “The Fine Structure of the Left Periphery.” In *Elements of Grammar*, ed. by Liliane Haegeman, 281-337. Dordrecht: Springer.
Sharpe, Donald. 2015. “Your Chi-Square Test is Statistically Significant: Now What?” *Practical Assessment, Research & Evaluation* 20 (8): 1-10.
Shi, Dingxu. 2000. “Topic and Topic-Comment Constructions in Mandarin Chinese.” *Language* 76 (2): 383-408.
Stefanowitsch, Anatol. 2020. *Corpus Linguistics: A Guide to the Methodology*. Berlin: Language Science Press.
Samo, Giuseppe, and Fuzhen Si. 2022. “Optionality of 的 De in Chinese Possessive Structures: A Quantitative Study”. *Quaderni Di Linguistica E Studi Orientali* 8: 37-53.
Timmis, Ivor. 2015. “Evaluation: Tails.” In *Corpus Pragmatics: a Handbook*, ed. by Karin Aijmer, & Christoph Rühlemann, 304-327. Cambridge: Cambridge University Press.
Qiang, Xingna. 2010. “话题标记 ‘嘛’ 与语气词 “嘛” [On “Ma” from Indicative Mood Particle to Topic Marker”. *Chinese Language Learning* 4: 59-64.
Qiang, Xingna. 2011. “话题标记与句类限制 [Topic Markers and Their Occurrence Constraints in Differing Sentence Types]”. *Linguistic Science* 10 (2): 190-204.
Van Craenenbroeck, Jeroen, and Marjo Van Koppen. 2022. “Quantitative Approaches to Syntactic Variation.” In Ms. KU Leuven/Meertens Institute.
Van Craenenbroeck, Jeroen, and Marjo Van Koppen. 2025. “Quantitative Approaches to Syntactic Variation.” In *The Cambridge Handbook of Comparative Syntax*, eds. by Sjef Barbiers, Norbert Corver, and Maria Polinsky, 1-23. Cambridge: Cambridge University Press.