Form and Function of Connectives in Chinese Conversational Speech

Nien-Heng Wu
Institute of Linguistics
Academia Sinica

Shu-Chuan Tseng
Institute of Linguistics
Academia Sinica

Abstract
Connectives convey discourse functions that provide textual and pragmatic information in speech communication on top of canonical, sentential use. This paper proposes an applicable scheme with illustrative examples for distinguishing Sentential, Conclusion, Disfluency, Elaboration, and Resumption uses of Mandarin connectives, including conjunctions and adverbs. Quantitative results of our annotation works are presented to gain an overview of connectives in a Mandarin conversational speech corpus. A fine-grained taxonomy is also discussed, but it requires more empirical data to approve the applicability. By conducting a multinomial logistic regression model, we illustrate that connectives exhibit consistent patterns in positional, phonetic, and contextual features oriented to the associated discourse functions. Our results confirm that the position of Conclusion and Resumption connectives orient more to positions in semantically, rather than prosodically, determined units. We also found that connectives used for all four discourse functions tend to have a higher initial F0 value than those of sentential use. Resumption and Disfluency uses are expected to have the largest increase in initial F0 value, followed by Conclusion and Elaboration uses. Durational cues of the preceding context enable distinguishing Sentential use from discourse uses of Conclusion, Elaboration, and Resumption of connectives.

Keywords: discourse function, Mandarin connectives, conversational corpus, phonetic representation

1. Introduction
Connectives are a class of lexical items that signal the relationship between units of text or discourse, connecting two different abstract objects, such as events, states or propositions, in discourse (Asher, 1993). It is posited that connectives have little lexical impact at the local segment level but serve significant pragmatic functions (Hjalmarrson, 2011). In conversational discourse, the position in which connectives occur and the phonetic form of connectives provide cues that help listeners process speakers’ intentions and the structure of the ongoing discourse (Didirková et al., 2018; Hirschberg and Litman, 1987, 1993; Horne et al., 2001; Rennie et al., 2016; Rhee, 2020). The use of connectives in conversation may be aimed more at marking discourse structure than at referring to the canonical meaning of the connectives themselves. It
is our goal to investigate whether there are differences between the canonical use (sentential) and the functional use (discourse) of connectives and whether it is possible to effectively disambiguate between the two uses by looking into the associated phonetic properties in conversation.

Previous research on Mandarin Chinese has identified various discourse-pragmatic functions in an array of connectives (Biq, 1994, 2001; Wang and Huang, 2006; Wang, 2018; Wang, 1998, 2005; Wang et al., 2013; Wang and Tsai, 2007; Wang et al., 2010; Yang, 2006), but only limited results have described the relationship between the functions and the form-related properties of connectives. For instance, the topic-shift ránhòu, ‘then’, has a significantly longer duration and pitch range larger than the canonical use, while the trail-off ránhòu (i.e., marking the closure of the current turn and inviting the hearer’s response) shows decreased loudness and durational lengthening. While the connection between the discourse function and the phonetic form of connectives may be strong (Biq, 2001; Wang, 2018; Yang, 2006), it lacks a systematic schema to describe discourse functions of Mandarin connectives and their phonetic forms. Previous studies offer mostly qualitative descriptions of individual connectives. In this study, we pursue a corpus-based discourse function annotation scheme and a quantitative analysis of Mandarin connectives and their phonetic features for discourse function disambiguation. The correlation examined through a statistically grounded method could be critical to the success of spoken language generation and understanding tasks.

This study considers conjunctions and adverbs to be the main lexical categories of target connective (Prasad et al., 2008; Zufferey and Degand, 2017). We conduct an annotation project on target connectives in a Chinese conversational speech corpus and analyze the positional, phonetic, and contextual properties of the associated discourse functions in a multinomial logistic regression model. With this task, we investigate whether a connective’s phonetic forms orient to sentential/discourse uses. If they do, are we able to find consistent patterns in terms of specific discourse functions? If speakers show sensitivity to the distinction of sentential/discourse uses of Mandarin connectives in their speech production, statistically significant coefficients that support consistent phonetic patterns in sentential/discourse uses are expected.

This article is organized as follows. Section 2 reviews the previous literature on Mandarin connectives’ discourse functions. Section 3 presents the literature on connectives’ form-related properties. In Section 4, we present the data, the annotation scheme, and the descriptive results of the positional, phonetic, and contextual features. In Section 5, we examine closely whether there is any coupling between the phonetic form and the discourse function of connectives by a multinomial logistic regression model. Finally, we discuss our main findings in Section 6.
2. Discourse functions of Mandarin connectives

Taking topic transitions in spoken discourse\(^1\) as the main focus in the consideration of discourse function grouping of Mandarin connectives, previous works basically distinguish discourse functions that signal initiation (resumption) and conclusion of topics, elaboration of various types, and disfluency. First of all, Wang (2005) and Wang and Tsai (2007) noted that the adverbs búguò, kěshì and dànsì, ‘but/yet/however’, may initiate a new topic in the discourse. Since the adverbs conventionally imply a contrast in propositions (Miracle, 1991; Ross, 1978), they may extend to imply a contrast between the new topic and the old topic in discourse. Similarly, Wang et al. (2010) observed that the adverb qíshí, ‘actually’, functioned to introduce a new topic or a new aspect of the current topic that may not be in accordance with the current claims in discourse (Biq, 1994). Working with the conjunction suǒyì, ‘so’, Wang and Huang (2006) proposed that suǒyì may initiate a new topic in the discourse, as the consequences introduced by suǒyì can be treated as new information on the discourse level. However, they also noted that suǒyì may signal a resumption of a previous topic and prevent further departure into irrelevant topics. The topic-resuming function was also found in the conjunction ránhòu, ‘then’. Yang (2006) and Wang (2018) maintained that ránhòu, canonically indicating either a temporal or consequential relationship between two adjacent clauses, can extend to organize utterances in discourse. As such, ránhòu may signal not only a change in topic but also a return to a previous topic after an intervening subtopic.

Connectives can also be used to signal conclusion of a current topic. Wang and Huang (2006) noticed that the speaker employed suǒyì to paraphrase or summarize the previous talk so that no misunderstanding was ensured. Relatedly, Wang (2018) described a trail-off use (Local and Kelly, 1986) in the turn-final ránhòu tokens. She maintained that the trail-off use of ránhòu may express “the speaker’s intent to close the turn and to invite the hearer’s responses of various types, such as acknowledgement, comment, or elaboration on the current topic” (p. 18).

More often in conversation, connectives allow the speaker to provide elaboration or clarification on a current topic. Biq (2001) claimed that the conjunction jiùshì, ‘that is’, when being slightly semantically reduced, may signal elaboration or clarification of the previous utterance. The elaboration function was also noted in zhìshí, ‘only’, typically a restrictive marker (Guo, 1999; Lü, 1980) or a focus marker (Wang, 2005). Wang (2005) and Wang and Tsai (2007) observed that the speaker may employ zhìshí to introduce an afterthought that comments or elaborates on the previous utterance. However, due to its implicature of contrast, the piece of discourse that zhìshí introduces may be incongruent with or divergent from the preceding discourse, such as a counterexpectation or surprising fact. Wang et al. (2013) similarly claimed that zhìshí may indicate “a more detailed or more correct formulation of

\(^1\) Studies reviewed in this section mainly focused on spontaneous conversations and/or TV/radio interviews and call-ins.
something stated previously” (p. 203). Wang and Huang (2006) noted that *ránhòu* featured an additive use that introduces new information to the current topic and connects successive ideas in discourse (Wang, 1998). Propositionally, *qíshí* is a commentary adverb that delivers the speaker’s attitude toward the propositional content. Wang et al. (2010) noticed that the canonical meaning of *qíshí* has developed several discourse functions that comment on the form or content of an utterance, such as elaborating on the previous utterance with more accurate or specific information.

All utterances encode the speaker’s attitude about the proposition to some extent, which may be indicated by connectives. For instance, Hsieh and Huang (2005) concluded that a *qíshí*-embedded clause may disclose a fact that the speaker believes that the hearer does not know. Propositional attitudes have been discussed in the form of emphasizing or supporting a proposition and limiting the validity of a proposition. Yang (2006) and Wang et al. (2010) suggested that *ránhòu* and *qíshí* may place emphasis on the distinctiveness of the following content. Wang et al. (2010) further contended that the fact-introducing *qíshí* may support and strengthen the speaker’s assertion. In contrast, Wang et al. (2013) claimed that the elaboration function of *zhǐshí* may imply incompatibility or insufficiency of the previous utterance and consequently limit the validity of the proposition in the utterance. The downtoning effect may further imply the speaker’s mild negative stance toward the propositions and his or her intent to instruct the hearer to reject previous claims.

Connectives may convey some interactional meanings in the speaker-hearer interaction, such as grabbing the hearer’s attention and establishing (dis)alignment between the interlocutors’ stances. Yang (2006) and Wang et al. (2010) suggested in their respective projects that the emphasis of *ránhòu* and *qíshí* may also function to attract the hearer’s attention to the discourse. Regarding the interlocutor stance, Hsieh and Huang (2005) and Wang et al. (2010) found that the speaker may use *qíshí* to disalign, and sometimes align, him/herself with the previous speaker’s stance. Wang (2005) and Wang and Tsai (2007) claimed that the contrastive *búguò*, *kěshì* and *dànshì* may preface the speaker’s dispreferred response, such as a rejection or disagreement, to the previous speaker’s utterance since disagreement can be one kind of contrastiveness (Ford, 2000). While Wang (2005) and Wang and Tsai (2007) did not explicitly note the function of expressing disagreement in *zhǐshí*, Wang et al. (2013) lent support to such a description by claiming that the incompatibility introduced by *zhǐshí* may extend to establish a contrast in the interlocutors’ stances, allowing the speaker to express minor or indirect disalignment with the previous speaker’s claims. The authors also noticed that in their data, *zhǐshí* was sometimes prefaced by a brief pause, which signals the speaker’s hesitation to agree with the previous speaker.

Last, connectives may sometimes contribute no propositional meaning to the discourse. They may simply signal disfluencies such as filled pauses. In her study of *jiùshí* and its variants, Biq (2001) noted that a more reduced *jiùshí* may serve as a filled pause or floor holder that does
not contribute to the proposition. This is evidenced by the fact that the utterance would not be understood differently if *jiùshì* were omitted. Yang (2006) further added that in the case of *ránhòu*, the connective may act to perform floor negotiations such as floor holding and turn-taking. These functions strategically enable the speaker to have more time to plan what to say next.

3. Positional, phonetic, and contextual encoding of connectives

In this section, related works on the positional, phonetic, and contextual encodings of connectives with various discourse functions are reviewed. Section 3.1 presents the positional encoding. Section 3.2 introduces various types of phonetic encoding, including pitch, intensity, and word duration correlates. Section 3.3 covers contextual cues that often co-occur with connectives.

3.1 Word position

Word position is the location of the lexical item being discussed in relation to a given unit in spoken discourse, be it a meaning-, interaction- or prosody-oriented unit, for instance, intonational unit (Hirschberg and Litman, 1987, 1993), interpausal unit (IPU) (Gravano et al., 2007; Gravano et al., 2011), speaker turn (Gravano et al., 2007; Gravano et al., 2011; Wang, 2018) or utterance (Rennie et al., 2016; Rhee, 2020). Recognizing the importance of word position, Hirschberg and Litman (Hirschberg and Litman, 1987, 1993) differentiated between the discourse and sentential use of the English *now* based on their positions in an intonational phrase and intermediate phrase (Pierrehumbert, 1980). They found that almost all tokens of discourse *now* (98.41%) were absolutely first or followed only another cue phrase in an intermediate phrase (e.g., *well now, ok now*), while only 13.5% of the tokens of sentential *now* were so placed. Sentential *now*, on the other hand, tended to occur intermediate phrase-finally (59.45%), whereas only 1.58% of the tokens of discourse *now* did. Similar tendencies were supported in another study of conjunctions (e.g., *and, but, or*, etc.), adverbs (e.g., *actually, also, indeed*, etc.), and other cue phrases (e.g., *okay, say, like*, etc.) (Hirschberg and Litman, 1993).

The utterance or speaker turn position of connectives is relevant to the associated discourse function. Investigating the English *so*, Rennie et al. (2016) posited that the utterance-initial and -second *so* may introduce either a topic shift or a conclusion of a previous utterance. In addition, the utterance-initial *so* may initiate a speaker change or a new utterance by the same speaker. On the other hand, the utterance-internal *so* may perform a resultative function that connects a new piece of information to an old utterance. The utterance-final *so* may release the speaker’s turn. Last, the standalone *so* may function as a turn-yielding device that urges the hearer to continue with the dialog or as a filled pause that holds the floor for the speaker. Rennie and colleagues, however, examined only the interaction between utterance positions and the two turn-organizing functions for the utterance-initial *so*. The interaction between utterance positions and other discourse functions remains unknown. Investigating *mak*, ‘coarsely’, in
Korean spontaneous conversations and scenarios of dramas and movies, Rhee (2020) observed that the filled-pause use of *mak* had a tendency to occur utterance-finally, which may reflect the tendency of the speaker to perform lexical search and floor holding at the end of the utterance. Despite stating that in principle, filled-pause use can occur anywhere in an utterance, Rhee did not provide any statistics. Concerning the relationship between discourse functions and speaker turn positions, Wang (2018) found that some of the utterance-final tokens of the Mandarin *ránhòu* had a turn-holding use in her conversational data, while the turn-final tokens all delivered the trail-off use, bringing the current talk to a closure and yielding the floor to another speaker. The trail-off use also occupied an independent intonation unit instead of being embedded in the previous intonation unit. Her findings, however, suffered from the problem of data sparseness, as she found only six tokens of the trail-off *ránhòu* and did not report any quantitative data for the turn-holding *ránhòu*. It is unclear whether the observed positional property is significant. Aside from the previously mentioned studies, Gravano and colleagues (Gravano et al., 2007; Gravano et al., 2011) found that the English cue words *alright* and *okay* tended to occur initially or independently in an IPU, signaling the beginning of a discourse segment. In summary, all these findings point to a strong tendency that positional encoding reflects discourse functions of connectives in speech production. It is unclear, however, to what extent different positional encodings (e.g., the initial, medial, final, or standalone position) correlate with each discourse function. It also remains an empirical question of which kind of production unit can reliably and effectively disambiguate between discourse functions.

### 3.2 Pitch, intensity, and duration

Previous studies have also reported phonetic evidence for various discourse functions. One phonetic encoding consistently discussed in the literature is pitch, which has been operationalized on different bases, such as pitch accent (Hirschberg and Litman, 1987, 1993), pitch reset (Didirková et al., 2018; Horne et al., 2001), and pitch range (Wang, 2018; Yang, 2006). Hirschberg and Litman (Hirschberg and Litman, 1987, 1993) found that the discourse use of the English *now* was more often deaccented than the sentential use of *now*. When forming part of a larger intermediate phrase, the majority of sentential uses of *now* received an H* or complex pitch accent, while all discourse uses of *now* bore an L* accent. Investigating pitch reset, Horne et al. (2001) and Didirková et al. (2018) each described its relation to topic shift in narrative speech and in spontaneous speech. Horne et al. (2001) observed a mean F0 reset for tokens of the Swedish *men*, ‘but’, that performed a topic-shift function, similar to the size of the reset one would observe at a topic-shift boundary. The effect of the F0 reset also led to 70% accuracy in distinguishing between discourse and sentential *men* in a linear classifier model. Analyzing the French *alors*, ‘then’, Didirková et al. (2018) reported that the connective tended to be marked by a reset on the word when introducing a new topic or specification. A reset in pitch was also observed in the French *et*, ‘and’, when introducing specification. Focusing on pitch range, both Yang (2006) and Wang (2018) described the way the phonetic parameter
manifested various functions of the Mandarin ránhòu. Yang (2006) showed that ránhòu, when signaling a topic shift or returning to a previous topic after an intervening subtopic, tended to have a larger pitch range and more perceptual prominence. In contrast, a use of ránhòu to signal a continuation of the current topic had a narrow pitch range and a more gradual and smoother contour. Yang also added that as an emphasis marker or a floor-negotiating device, ránhòu was marked with an expanded pitch range, suggesting the hearer pay attention to the speaker. However, no quantitative data were reported. Adopting a quantitative approach, Wang (2018) showed supporting evidence for Yang’s findings that the topic-shifting ránhòu had a much larger average pitch range than its other sentential uses. Moreover, the turn-initial ránhòu showed a larger pitch range than the noninitial ránhòu.

Meaningful phonetic variation has also been observed in the study of intensity, commonly referred to as loudness or volume. Rennie et al. (2016) compared the intensity of the English so and that of its adjacent segments. They showed that both the so tokens that introduced either a topic shift or a conclusion of a previous utterance and the so tokens that connected a new piece of information to an old utterance were significantly quieter than their following segment in paired t-tests. In contrast, the so tokens that introduced a conclusion of the current utterance appeared to be louder than its preceding segment, even though the difference was not significant. The increase in intensity after so, as suggested by Rennie and colleagues, may signal a more important status of the segment following so in the utterance. The decrease in intensity before so, on the other hand, seems to be in line with the tendency of which intensity increases at the start of a new topic and decreases at the end (Brown et al., 1980). Analyzing the Mandarin ránhòu, Wang (2018) also reported a similar observation: that the tokens marking the closure of the current speaker’s turn were marked with gradually decreased loudness as well as lengthening. Her observation, however, was qualitative and based on only six tokens in her data.

In connection to the relationship between phonetic encodings and the structure of discourse, attention has been given to temporal features such as word duration. For instance, Horne et al. (2001) observed a significant difference in mean duration between the discourse and sentential use of the Swedish men, where discourse men was longer than sentential men. Studies on the topic-shift function of the French alors (Didirková et al., 2018) and of the Mandarin ránhòu (Wang, 2018) have similarly reported a longer word duration than that of their respective sentential functions. Yang (2006) also offered insights into the durational correlates of other discourse functions of ránhòu. Yang found that the ránhòu tokens signaling a continuation of the current topic were short in word duration. In contrast, the tokens signaling an emphasis on the distinctiveness of the following content had a much longer word duration. She claimed that a longer duration attracts the hearer’s attention to the current topic, while the lack of prominent duration reflects less of a need to call for the hearer’s attention, which typically occurs when the following utterance develops step-by-step within the same topic. Rennie et al. (2016), on the other hand, presented a case where the English so was shorter in
duration mean when functioning to grab the hearer’s attention in the utterance-initial position. This is also evidenced by the lack of a perceptible pause between so and subsequent speech.

In addition to the aforementioned phonetic evidence, it has also been suggested that high-frequency disyllabic connectives used in Mandarin conversational speech are often produced in a phonetically extremely reduced form (Liu et al., 2016). The duration and position of disyllabic connectives tend to correlate with the degree of word reduction. Therefore, in our later analysis, we will include four types of phonetic correlates, including pitch, intensity, duration, and reduction degree.

3.3 Context
There has been some research into how contextual cues such as silent pauses and paralinguistic events may aid speakers in structuring discourse. A silent pause typically signals a major prosodic boundary. Horne et al. (2001) observed that 34% of the topic-shift use of the Swedish men was both preceded and followed by a pause, while none of the sentential uses were. Similarly, Wang (2018) found that during trail-off use, the Mandarin ránhòu was marked with a pitch contour independent of the previous intonation unit. In addition, it was immediately followed by laughter. On silent pause duration, Didirková et al. (2018) revealed that the silent pause preceding the French alors tended to be longer when the connective opened a new topic or introduced specification. Moreover, they noted that discourse uses of alors was almost never followed by a silent pause. Rhee (2020) reported that the filled-pause use of the Korean mak was often realized after a short pause, signaling that floor holding was needed. He further stated that a pause may distinguish the filled-pause use from other discourse functions of mak, such as the speaker expressing a negative stance and intensifying an utterance, as these functions tended to be marked with no pauses before or after mak. Swerts (1998) also discussed silent pauses preceding and following the Dutch filled pause uh and um. It showed that almost all tokens of phrase-initial uh and um had a neighboring silent pause.

4. Data and annotation of Mandarin connectives
In this section, we present the scheme with which we labeled our target connectives and the annotation results with illustrative examples. Section 4.1 gives a brief overview of the corpus used in this study. Sections 4.2 and 4.3 describe the annotation scheme with which we labeled all tokens of connectives in our corpus in terms of their sentential/discourse use and individual discourse functions. Representative examples of the annotated discourse functions are illustrated in Section 4.4. Lastly, Section 4.5 presents descriptive statistics of our data in terms of positional, phonetic, and contextual features.

4.1 Target connectives in Sinica MCDC8
Sinica MCDC8 contains eight free conversations produced by seven male and nine female Mandarin Chinese speakers aged between 16 and 46. The speakers were randomly sampled
from the citizens of Taipei City in 2001. Each pair of speakers who were invited to participate in the recording project met each other for the first time. The corpus has approximately eight hours of speech recording with 90K transcribed words/122K syllables. Acoustic properties that will be used for our later analysis were measured based on the signaled-aligned syllable boundary information (Tseng, 2019). In the present study, we performed an exploratory analysis to identify our target connectives in the corpus, including böguǎn, ‘no matter’; jiārú, ‘if’; jíshí, ‘even if’; jùshí, ‘is precisely’; háishí, ‘still’; huōshí, ‘or’; huōzhé, ‘or’; rúguǒ, ‘if’; suīrán, ‘even though’; suóyí, ‘so’; yaōshí, ‘if’; zhīshí, ‘only’; zhīyaō, ‘as long as’; ránhòu, ‘then’; and qīshí, ‘actually’. As part of the phonetic features, we adopted the labels of disyllabic reduction degree from Liu and colleagues (Liu et al., 2016) and selected only connectives that had such annotation in the corpus. Eventually, we obtained a total of 1370 connective tokens, as shown in Table 1.

<table>
<thead>
<tr>
<th>Connective</th>
<th>Tokens</th>
<th>Connective</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>böguǎn</td>
<td>11</td>
<td>suīrán</td>
<td>9</td>
</tr>
<tr>
<td>jiārú</td>
<td>7</td>
<td>suóyí</td>
<td>215</td>
</tr>
<tr>
<td>jíshí</td>
<td>2</td>
<td>yaōshí</td>
<td>3</td>
</tr>
<tr>
<td>jùshí</td>
<td>387</td>
<td>zhīshí</td>
<td>32</td>
</tr>
<tr>
<td>háishí</td>
<td>67</td>
<td>zhīyaō</td>
<td>23</td>
</tr>
<tr>
<td>huōshí</td>
<td>20</td>
<td>ránhòu</td>
<td>298</td>
</tr>
<tr>
<td>huōzhé</td>
<td>11</td>
<td>qīshí</td>
<td>216</td>
</tr>
<tr>
<td>rúguǒ</td>
<td>69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Frequency of Connectives in Sinica MCDC8

To present the semantic content and the prosodic organization of spontaneous conversation, Prévot et al. (2015) proposed two types of production units, discourse units (DUs) and prosodic units (PUs). A DU consists of a main predicate and the related complements and adjuncts. A PU is a stretch of speech content separated by perceptible pitch reset, changes in speech rate, and pauses. We adopted the definition of DUs and PUs proposed by Prévot et al. (2015) and used the word position in which a connective occurs relative to the respective DU/PU as our positional features of connectives in our later analysis.

4.2 Sentential/discourse labeling

We posited that a connective delivers a sentential use if the meaning of the DU is inevitably changed or becomes incomplete when the connective is removed from the DU in which it occurs. This is illustrated by (1), where the two speakers talked about what they do for work. After asking Speaker A where her office is and failing to get a satisfying answer, Speaker B asked Speaker A whether she could describe the direction to her office from Nangang District in Taipei, prefaced by rúguǒ in bold. Canonically, rúguǒ indicates hypotheticality, which suggests that the proposition of the DU (nà rúguǒ cóng Nángǎng guòqū ‘if going there from

---

2 Sinica MCDC8 is publicly available via the Association for Computational Linguistics and Chinese Language Processing.
Nangang’) is a purely hypothetical statement. When *rüguō* is removed from the DU, the semantic meaning of the DU is also affected. This shows that *rüguō* is used sententially in this case.

(1)

B: NA 如果/從南港過去要怎麼去A.
NA rüguō丛nángāng guòqù yào zěnme qù A
‘If I go there from Nangang, how do I get there?’

If a connective adds a designated discourse interpretation to the DU in which it occurs in relation to the local context, it is considered discourse use. An instance of such connective is shown in (2), which presents a conversation in which Speaker A told Speaker B about her trip to a hot spring resort in Japan. Speaker A described the indoor baths as separated from the outdoor baths and stated that the experience was quite interesting. She was going to share her opinion about Japanese people using the expression *wǒ jüede*, ‘I feel/think’. She then abandoned the thought and shifted to talking about liking Japan, which is less directly related to what is being discussed. The transition to a new topic (i.e., her liking Japan) on Line 25 was introduced by *qíshí*, which is discourse use.

(2)

1 A: NHNN, /
    NHNN
2 他們原則上就是說他一般 BREAK /
    tàmén yuánzéshàng jiùshìhuo tā yī BREAK
3 有 隔, /
    yǒu gé
4 就是, /
    jiùshì
5 [把 它隔開] LA. / $
    [bǎ tā gékāi] LA
6 B: [還是有隔開]. / $
    [háishì yǒu gékāi]
7 A: 就是]隔室內跟室外而已 LA. $ INHALE/ $
    jiùshì] gé shìnèi gēn shìwài éryǐ LA INHALE
8 B: ON [HON]. / $ ON [HON]
9 A: [就是]讓 NE GE, /
    jiùshì ràng NE GE
10 可能是讓 NE GE, INHALE /
    kěnéng shì ràng NE GE INHALE
11 NE GE 氣 O, /
    NE GE qì O
12 溫度 O, /
    wēndù O
13 不要, PAUSE /
    búyào PAUSE
14 那樣子交流這樣子, / $
    náyàngzi jiāoliú zhèyàngzǐ

---

3 The symbol / indicates a prosodic unit boundary, and the symbol $ indicates a discourse unit boundary in Sinica MCDC8.
Horne et al. (2001) mentioned that both a sentential use and a discourse use interpretation of connectives can seem possible. We also noticed that in some cases, the distinction of sentential or discourse use can be ambiguous. For instance, in (3), Speaker A told Speaker B about the harmful effect of formaldehyde and her effort to educate people about it. She mentioned that all she could do is to share the information with people, and it is up to people to do something with the information. She then said じゅうし じゅつよう, literally ‘that is it’ in English, on Line 11. Here, じゅうし may sententially indicate the preciseness of the equation between じゅつよう ‘like this’ and the previous content. However, we identified a discourse meaning in which the speaker introduced a conclusion for her previous topic and signaled to the hearer that there is no more to add to the current topic.

(3)

1 A: 所以 我 覺得, / suǒyǐ wǒ juéde
2 現在 我 覺得, INHALE / xiànzài wǒ juéde INHALE
So now I feel that it is my responsibility to tell them what I know.

‘[Right.]’

‘[As for] whether they accept it or not, it is up to them.’

‘Right.’

‘I do my best.’

‘@’

‘That is it. It is actually quite fun.’

### 4.3 Annotation scheme of discourse functions

Describing the discourse functions of connectives has often proven a challenging task since the interpretation of the functional properties can be quite elusive and often context dependent. The exploration and initial annotation of discourse functions were carried out by the authors. Auditory information was used to aid the classification wherever necessary. According to previous work, we were able to identify sentential use and eight types of discourse use for our connectives. It was relatively straightforward to adopt the definitions of resuming and concluding topics and disfluencies. We could also identify a more coarse-grained type of function elaboration for the majority of the cases that provide elaboration or clarification on a current topic. However, for functions such as emphasis, downtoning, securing the addressee’s attention, and contrast, it was, in fact, truly challenging to operationalize and identify them. Therefore, we collapsed the above four discourse functions, along with elaboration, into Elaboration, as they all provide more information to a proposition. The exploration of the data also identified cases where connectives signaled repairs in discourse (Tseng, 2006). As such, we added repair to our discourse functions and collapsed it into Disfluency. This led to an annotation scheme for four function categories: Conclusion, Resumption, Elaboration, and Disfluency. For validation, two trained labelers were recruited for verifying sentential use and the four function categories. Each labeler annotated half of the dataset independently. As a result, the agreement between the authors’ and the labeler’s annotations achieved a Cohen’s
kappa of 0.92. Although judgment of sentential/discourse use is likely to be considered highly subjective, the agreement over the annotation of connective functions is surprisingly satisfactory. Table 2 summarizes the statistics and their respective definitions. Controversial cases were discussed to finalize the annotations.

<table>
<thead>
<tr>
<th>Function Category</th>
<th>Function</th>
<th>Definition</th>
<th>Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>Sentential</td>
<td>Functions as a conjunction or an adverb</td>
<td>639</td>
<td>46.64%</td>
</tr>
<tr>
<td>Resumption</td>
<td>Resumption</td>
<td>Resumes a new or previous topic</td>
<td>32</td>
<td>2.33%</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Conclusion</td>
<td>Introduces a conclusion or summarization</td>
<td>103</td>
<td>7.52%</td>
</tr>
<tr>
<td>Elaboration</td>
<td>Elaboration</td>
<td>Elaborates on a current or previous topic</td>
<td>469</td>
<td>34.23%</td>
</tr>
<tr>
<td></td>
<td>Emphasis</td>
<td>Emphasizes the significance of the proposition</td>
<td>14</td>
<td>1.02%</td>
</tr>
<tr>
<td></td>
<td>Downtoning</td>
<td>Downplays the significance of the proposition</td>
<td>7</td>
<td>0.51%</td>
</tr>
<tr>
<td></td>
<td>Securing the</td>
<td>Secures the addressee’s attention</td>
<td>3</td>
<td>0.22%</td>
</tr>
<tr>
<td></td>
<td>addresssee’s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>attention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>Expresses the speaker's disagreement with the previous speaker</td>
<td>12</td>
<td>0.88%</td>
</tr>
<tr>
<td>Disfluency</td>
<td>Filled pause</td>
<td>Signals filled pauses</td>
<td>81</td>
<td>5.91%</td>
</tr>
<tr>
<td></td>
<td>Repair</td>
<td>Signals repairs</td>
<td>10</td>
<td>0.73%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1370</strong></td>
<td><strong>99.99%</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Sentential/Discourse Functions of Connectives in Sinica MCDC8

4.4 Annotation examples

As shown previously in (2) and (3), connectives can be used to perform topic shifts in conversation: *qíshí* in (2) was found to resume a topic, and *jiùshì* in (3) may conclude a previous topic. Aside from topic shift, some connectives may introduce an elaboration on the current topic, illustrated by *ránhòu* on Line 6 in (4).

(4)

1 B: 那 他們, BREAK /
    nà tāmén BREAK
2 講話 就是 有 口音 吗. / $
    jiānghuà jiùshì yǒu kǒyīn ma
3 A: NHN, $ INHALE / $
    NHN INHALE
4 還好 BA $ 可是, BREAK /
    háihào BA kěshì BREAK
5 因為 他 教 兩 個 班 $
    yīnwèi tā jiāo liǎng ge bān
6 然後 教 我們 班 跟 另外 一 班, $ INHALE / $
    ránhòu jiāo wǒmén bān gèn lingwài yī bān INHALE
‘Did he speak with an accent?’
A: ‘The accent was not bad. But, since he taught two classes, which were my class and another class, every time… sometimes when I needed to rush to another classroom for class, I would see that half of the students in the back fell asleep while he was teaching in the front. [That is it. Right.]’
B: ‘[Right. It was the same way for us too.]’

Intriguingly, we have observed that some connectives can emphasize or downtone the importance of a proposition in discourse, as illustrated in (5) and (6), respectively. In (5), the two speakers were talking about modified cars. Speaker B pointed out a problem on Line 7 in that many people drive fast in their modified cars. He then argued that people should modify their cars only for safety. Prefacing his argument with the fact-introducing marker qíshí (Hsieh and Huang, 2005; Wang et al., 2010), Speaker B was able to suggest the proposition in his argument was factual and consequently emphasize its importance.

(5)
仗著，INHALE / zhàngzhe INHALE

他們自己車子很好，$INHALE/$ tāmen zìjǐ chēzǐ hěn hǎo INHALE

開很快，$INHALE/$ kāi hěn kuài INHALE

那些是他們心態上的問題，$INHALE/$ NA zhèxiē shì tāmen xīntài shàng de wèntí

可是，/ kěshì

其實，BREAK/ qíshí BREAK

改車子並不是，BREAK/ gǎi chēzǐ bìng bù shì BREAK

針對於飆車用$是對安全性，$ zhěnduì yù biāochē yòng shì duì ānquánxìng SWALLOW/$ SWALLOW

A: MHMHMHM. / $ MHMHMHM

B: ‘Let me ask you… there is surely a difference between making a turn at the speed of 100 km/hr and making a turn at the speed of 50 km/hr.’
A: ‘Right.’
B: ‘Nowadays many modified cars… those people may drive fast because they think their cars are good. That’s the problem with their attitude. In fact, modifying cars is not about racing. It is about safety.’
A: ‘Right.’

In contrast, connectives such as zhǐshì can downtone the importance of the proposition in discourse. In (6), Speaker A explained that she was stuck in traffic right before arriving at Academia Sinica, where the recording of the conversation took place. She, however, went on to clarify on Line 18 that the traffic on the way to Academia Sinica was not that bad and that she was late only because she did not estimate the travel time correctly (shíjiān shàng yǔgū kēnèng méiyǒu xiǎo yìxià, ‘I didn’t think about the estimated time to get here.’). Her reason for being late was prefaced by zhǐshì, which limited the validity of the proposition in discourse (Wang et al., 2013). The connective allowed Speaker A to downtone the importance of her reason and strengthen her point that the traffic to Academia Sinica was in fact not bad.

(6)

1 A: NA, / NA
2 在中央研究院 的 NE GE, / zài zhōngyāng yánjiùyuàn de NE GE
3 研究院一路 那邊, INHALE / yánjiùyuàn yīlù nà biān INHALE
4 那鐵道有點塞車 所以 我 [一路] nà tiědào yǒudiǎn sāichē suǒyì wǒ yīlù
5 B: [HEN HEN]. / $ HEN HEN
6 A: 上 過來 剛好, INHALE / shàng guólái gānhǎo INHALE
There was a little bit traffic around the railway on Academia Road Section 1. [On the whole way]

[Right.]

A: ‘here, there happened to be two… one place under construction on the way here.

[Then.]

B: ‘[Right.]

A: ‘I was also not very familiar with the roads in Nangang, [so]’

B: ‘[Right.]

A: ‘I was stuck in traffic a little on Academia Road Section 1.’

B: ‘Oh.’

A: ‘The traffic on the way here was still not bad. It is just that I didn’t think about the estimated time to get here.’

Attention-securing is another interlocutor interaction enabled by our connectives. The connective rúguǒ, for example, can be used to secure the attention of the hearer. In (7), Speaker A argued that Western democracy is not the best political system for all countries and that it will be better if a country has the freedom to figure out which system works best for it. He used China as the example for his argument, saying that although China went through a dark time
when Communism was first introduced to the nation, it is enjoying great economic development now. He then recalled a report about Shanghai that he saw on TV on Line 10, prefaced by a rúguǒ-led clause addressing Speaker B (rúguǒ nǐ yǒu kàn disitài, ‘if you watch the cable TV’). Instead of probing for a response from Speaker B, evidenced by a lack of wait time for Speaker B to say something, Speaker A wanted to bring Speaker B’s attention to his next utterance on the TV report.

(7)

1. A: NA, / NA PAT
2. NA, / NA PAT
3. 他們 帶 過來 的 時候 $ 你 說, /
tāmen dài guòlái de shíhòu nǐ shuō
they bring over PAT time you say
4. 對 $ 中共, PAUSE / $ dui $ zhōnggòng PAUSE
To Chinese Communist Party PAT
5. 中國 大陸 的確 是 走 過 一 段 很, / zhōngguó dàlù díquè shì zǒu guò yī duàn hěn
China mainland indeed PAT walk PAT one stage very
6. 蠻 黑暗 的. $ INHALE / $ mán hēi’àn de INHALE
quite dark PAT PAT
8. A: [但 你 看] 他們 現在 <X 一 發 X>, / dàn nǐ kàn tāmen xiànzài yī fā
but you see they now once develop
9. 一 發展 起來, $ INHALE / $ yī fāzhǎn qǐlái INHALE
once develop PAT PAT
10. O 那 真的 不得了 $ 如果./ O nà zhēnde bùdélìáo rúguǒ
PAT that really incredible if
11. 你 有 見 第四台 $ 我 常常 看
nǐ yǒu jiàn dìsìtái wǒ chángcháng kàn
you have see cable TV I often see
12. NE GE 他們 報告 上海, / $ NE GE tāmen báodào shànghǎi
PAT they report Shanghai
13. 經 batteries, $ INHALE / $, jīng INHALE PAT
14. 大 城市 來 講 真的. PAUSE / dà chéngshì lái jiāng zhēnde PAUSE
big city PAT speak really PAT
15. 不會 輸 我們. $ PAUSE / bùhuì shū wǒmén PAUSE
will not lose us PAT
When the Soviet Union introduced Communism to China, it is true that mainland China walked through a very dark time.'

['Right.]

'[But you can see] that now they have developed (economically) and it is really incredible. If you watch the cable TV... I often watch them reporting about Shanghai. If we are talking about (the development of) big cities (in China), it won't lose to us.'

'Right.'

'It definitely beats us.'

Connectives can also convey certain interactions between interlocutors. For instance, in (8), the speaker used qíshí on Line 5 to express her disalignment with the previous speaker.

(8)

‘Actually, if I didn’t say I’m a taxi driver, you probably can’t tell.’

‘Actually, [I think that nobody can]’

'[You can’t really tell.]'

‘tell what a person’s [profession is]’

'[Right.]'

‘because there are too many professions. [It is also difficult to guess.]’

’[(The profession) does make a difference.]’
Last, connectives can signal disfluencies such as filled pauses and repairs, as illustrated in (9) and (10), respectively. In (9), suǒyǐ functioned as a filled pause on Line 11. It was followed by a short pause, which hints at the speaker’s hesitation or word-searching.

(9)

1 A: 除此之外, $\quad$ INHALE / $
   chūxízhúwài
2 因為 已經=, $
   yīnwèi yǐjīng
3 要 開始, EXHALE /
   yào kāishǐ
4 就是 要跟老師做研究了這樣, $\quad$
   jiùshì yào gēn lǎoshī zuò yánjiū le zhèyàng
5 所以, INHALE /
   suǒyǐ
6 在, INHALE
7 最近的幾天或 最近的幾個月之內
   zuìjìn de jǐ tiān huò zuìjìn de jǐ ge yuè zhīnèi
8 都 會 下去, $\quad$ INHALE /
   dū huì xiàqù
9 就是 台北 台南這樣兩邊跑
   jiùshì táiběi tánán zhèyàng liǎngbiān pǎo
10 這樣, $\quad$ INHALE /
   zhèyàng
11 對, $\quad$所以=, INHALE /
   duì suǒyǐ
12 雖然說 離家很遠 LA $
   suīránshuō líjiā hěn yuǎn LA
13 可是=, INHALE /
   kěshì
14 看 以後 會不會 成長, $\quad$
   kàn yǐhòu huì bù huì chéngzhǎng

A: ‘In addition, since I’m about to start doing research with my supervisor, I’ll go down to Tainan in the next few days or in the next few months. That is, I’ll travel between Taipei and Tainan. Right. So… Although I’ll be far away from home, I’ll see if I will grow (through this experience).’

In (10), Speaker B asked Speaker A, on Line 6, whether the humanities majors in her school have to take summer lessons. She started her question off with nǐmén, ‘you’, referring to the humanities majors, before introducing the alteration (qítā jiù búyòng, ‘other (people) just don’t need to (have lessons)’) with jiùshì. It is clear that jiùshì signaled to the hearer that what follows is a repair to what was just said.

(10)

1 A: 我們 好像是 學校念, INHALE /
   wǒmén hǎoxiàng shì xuéxiào niàn
2 NE GE 理科的 才要回去學校
   lǐkē de cái yào huí xuéxiào
3 去 [上課]. $\quad$ INHALE /
   qù shǎngkè

106
4 B: [O]. /$
O$
5 A: 對 A $ 對. /$
duì A duì
6 B: NA 妹們 就是, /
NA nǐmén jiǔshì
7 就是 其他 就 不用, /
jiǔshì qítā jiù búyòng
8 回去 LO.
huíqù LO
9 A: 對 A $ 文科 就 不用 回去. $ INHALE /$
duì A wénkē jiù búyòng huíqù INHALE
10 B: 好好 O. /$
hǎohǎo O
A: ‘For us, it seems that only the students who major in sciences in our school need to go back to [have lessons.]’
B: ‘[Oh.]’
A: ‘Right.’
B: ‘So, you just... other people just don’t need to go back for lessons.’
A: ‘Right. Students who major in humanities don’t need to go back.’
B: ‘That is nice.’

4.5 Descriptive results of annotated connectives

This section presents the descriptive statistics of our target connectives based on the three major groups of features. Section 4.5.1 shows the positional features, which designate the position of a connective in relation to a DU/PU. Section 4.5.2 presents the phonetic features, including duration, F0, intensity, and reduction degree. Section 4.5.3 describes the contextual features, including the duration of preceding and following paralinguistic events as well as the speech rate of DU/PU in which a connective occurs.

4.5.1 Positional features

The position of the connective is operationalized as the initial, medial, and final position in relation to a DU/PU and the case in which a connective itself forms an isolated DU/PU, annotated as DU_initial, DU_medial, DU_final, and DU_isolated. Similar to DU, the positions in a PU are annotated as PU_initial, PU_medial, PU_final, and PU_isolated. As shown in Table 3, connectives of Sentential use do not seem to particularly occur in DU-initial or -medial positions, but when connectives deliver discourse functions of Conclusion, Elaboration or Resumption, a DU-initial position is generally preferred. When used in relation to Disfluency, connectives tend to take DU-medial or DU-initial positions. In contrast to DU, positional features related to PU show that the prosodic manifestation of connectives is diverse across discourse functions. In terms of prosodic segmentation, connectives seem more likely to occur in the form of a standalone unit than in the meaning-oriented segmentation of discourse. We will later conduct a multinomial logistic regression model to examine whether there is any significant effect in the comparison of DU and PU.
4.5.2 Phonetic features

We considered word duration, pitch, intensity, and reduction degree in our analysis, following previous studies’ suggestions of phonetic correlates for the discourse function of connectives (Didirková et al., 2018; Hirschberg and Litman, 1987, 1993; Horne et al., 2001; Liu et al., 2016; Rennie et al., 2016; Wang, 2018; Yang, 2006). For each connective token, Rate, Initial_F0, and IntensityMean represent the word duration in the form of speech rate (seconds per syllable), the initial F0 value, calculated using the firstPitch function in Praat (Boersma and Weenink, 2022), respectively, and the mean of the intensity values is calculated using Praat’s meanIntensity function. Figure 1 presents the results. For speech rate, connectives tend to be longer for Disfluency (Mean: 0.19 sec/syll) or Resumption (Mean: 0.17 sec/syll) use than for Sentential use (Mean: 0.14 sec/syll). All discourse uses (MeanConclusion: 198.07 Hz, MeanDisfluency: 200.53 Hz, MeanElaboration: 185.83 Hz, MeanResumption: 216.29 Hz) tend to have a higher pitch onset than Sentential use (Mean: 173.33 Hz). For intensity, there is no clear trend contrasting sentential/discourse use, but Disfluency use seems to have a weaker intensity (Mean: 68.15 dB) than the other discourse functions (MeanConclusion: 71.19 dB, MeanElaboration: 70.76 dB, MeanResumption: 71.44 dB).

Table 3: Distribution of Discourse Functions by DU and PU Positions of Connectives

<table>
<thead>
<tr>
<th></th>
<th>DU_initial</th>
<th>DU_medial</th>
<th>DU_final</th>
<th>DU_isolated</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>290 (45.38%)</td>
<td>343 (53.68%)</td>
<td>6 (0.94%)</td>
<td>0</td>
<td>639</td>
</tr>
<tr>
<td>Conclusion</td>
<td>96 (93.20%)</td>
<td>7 (6.80%)</td>
<td>0</td>
<td>0</td>
<td>103</td>
</tr>
<tr>
<td>Disfluency</td>
<td>31 (34.07%)</td>
<td>51 (56.04%)</td>
<td>3 (3.30%)</td>
<td>6 (6.59%)</td>
<td>91</td>
</tr>
<tr>
<td>Elaboration</td>
<td>397 (78.61%)</td>
<td>105 (20.79%)</td>
<td>1 (0.20%)</td>
<td>2 (0.40%)</td>
<td>505</td>
</tr>
<tr>
<td>Resumption</td>
<td>26 (81.25%)</td>
<td>6 (18.75%)</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PU_initial</th>
<th>PU_medial</th>
<th>PU_final</th>
<th>PU_isolated</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>219 (34.27%)</td>
<td>305 (47.73%)</td>
<td>70 (10.95%)</td>
<td>45 (7.04%)</td>
<td>639</td>
</tr>
<tr>
<td>Conclusion</td>
<td>68 (66.02%)</td>
<td>13 (12.62%)</td>
<td>3 (2.91%)</td>
<td>19 (18.45%)</td>
<td>103</td>
</tr>
<tr>
<td>Disfluency</td>
<td>23 (25.27%)</td>
<td>5 (5.49%)</td>
<td>24 (26.37%)</td>
<td>39 (42.86%)</td>
<td>91</td>
</tr>
<tr>
<td>Elaboration</td>
<td>284 (56.24%)</td>
<td>81 (16.04%)</td>
<td>46 (9.11%)</td>
<td>94 (18.61%)</td>
<td>505</td>
</tr>
<tr>
<td>Resumption</td>
<td>17 (53.12%)</td>
<td>4 (12.50%)</td>
<td>1 (3.12%)</td>
<td>10 (31.25%)</td>
<td>32</td>
</tr>
</tbody>
</table>

Figure 1: Boxplot Representation of Discourse Functions by Speech Rate, Initial F0 Value, and Intensity Mean of Connectives (the red dot indicates the mean)
For word reduction, four degrees are distinguished for each connective token: the canonical form (CAN), the marginal segment deletion form (MSD), the nuclei merging form (NUM), and the syllable merger form (SYM), representing an increase in phonetic reduction from CAN to SYM (Liu et al., 2016). High-frequency disyllables in Sinica MCDC8 mostly take the most reduced form, SYM. The results shown in Table 4 are partly in accordance with this tendency. However, for Sentential or Disfluency use, the percentage of SYM is smaller than the percentage of other discourse functions.

<table>
<thead>
<tr>
<th>Discourse Function</th>
<th>SYM</th>
<th>CAN</th>
<th>MSD</th>
<th>NUM</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>382 (59.78%)</td>
<td>190 (29.73%)</td>
<td>7 (1.10%)</td>
<td>60 (9.39%)</td>
<td>639</td>
</tr>
<tr>
<td>Conclusion</td>
<td>91 (88.35%)</td>
<td>7 (6.80%)</td>
<td>0</td>
<td>5 (4.85%)</td>
<td>103</td>
</tr>
<tr>
<td>Disfluency</td>
<td>50 (54.95%)</td>
<td>20 (21.98%)</td>
<td>1 (1.10%)</td>
<td>20 (21.98%)</td>
<td>91</td>
</tr>
<tr>
<td>Elaboration</td>
<td>350 (69.31%)</td>
<td>88 (17.43%)</td>
<td>16 (3.17%)</td>
<td>51 (10.10%)</td>
<td>505</td>
</tr>
<tr>
<td>Resumption</td>
<td>22 (68.75%)</td>
<td>7 (21.88%)</td>
<td>1 (3.12%)</td>
<td>2 (6.25%)</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 4: Distribution of Discourse Functions by Reduction Degree of Connectives

### 4.5.3 Contextual features

Previous research has suggested functional differences in the contextual cues occurring around connectives, such as silent pause (Didirková et al., 2018; Horne et al., 2001; Rhee, 2020) and laughter (Wang, 2018). We considered the position and duration of all paralinguistic events, such as pauses, coughs, laughs, etc., occurring around each connective token in terms of Previous_duration and Next_duration. Previous_duration is the duration of a preceding paralinguistic event, and if there is no immediately adjacent paralinguistic event, the absence is marked with a zero. The same definition applies for Next_duration. Figure 2 presents the results. Connectives of discourse uses ($Mean_{Conclusion}$: 0.32 sec, $Mean_{Disfluency}$: 0.21 sec, $Mean_{Elaboration}$: 0.26 sec, $M_{Resumption}$: 0.39 sec) seem to be more likely to be accompanied by a preceding paralinguistic event than Sentential use ($Mean$: 0.10 sec). However, paralinguistic events following the occurrence of a connective do not seem to be as influential as those that precede such occurrences, except for Disfluency use ($Mean_{Disfluency}$: 0.33 sec).

![Figure 2: Boxplot Representation of Discourse Functions by Duration of Paralinguistic Events in the Previous and Next Position of Connectives (the red dot indicates the mean)](image-url)
Another contextual property of connectives we considered is the speech rate of the entire DU/PU in which the connective occurs. We posited that the overall speech rate of DU/PU may correlate with the discourse function by reflecting predominant rhythmic patterns. DU_rate is calculated by DU/PU duration divided by the number of syllables in DU/PU after removing paralinguistic events and fillers. Figure 3 shows the results of DU_rate and PU_rate. Connective-occurring DUs seem to be articulated in a slower tempo when used for Disfluency (Mean: 0.23 sec/syll) or Resumption (Mean: 0.21 sec/syll) purpose than when used sententially (Mean: 0.19 sec/syll). On the other hand, connective-occurring PUs seem to be articulated at a slower tempo for all discourse uses (MeanConclusion: 0.22 sec/syll, MeanDisfluency: 0.35 sec/syll, MeanElaboration: 0.23 sec/syll, MeanResumption: 0.27 sec/syll) than for Sentential use (Mean: 0.21 sec/syll).

![Figure 3: Boxplot Representation of Discourse Functions by Speech Rate of the DU and PU Where Connectives Occur (the red dot indicates the mean)](image)

5. Analysis of form and function of Mandarin connectives

We have identified a number of tendencies for the three groups of features in connective use. Seemingly, all these features play certain roles in the production of connectives. Our next study fits a multinomial logistic regression model to test whether these features statistically show an inclination for sentential/discourse uses of connectives.

5.1 Multinomial logistic regression

Multinomial logistic regression (MLR) makes inferences about category memberships. It describes the probability of a comparison category being chosen over a reference category in a dependent variable as an outcome based on multiple independent variables. In this study, we built regression models to determine which features are more negatively or positively critical than others in predicting discourse functions. The independence of irrelevant alternatives (IIA) is crucial in MLR modeling since the probabilities for any pair of categories should be determined without reference to the other categories that might be available. The IIA can be tested by the Hausman specification test (Hausman and McFadden, 1984). We calculated $H_{IIA}$ using the $hmftest$ function from the mlogit package (Croissant, 2020) and obtained an
insignificant $H_{11a}, p > .05$, suggesting that we can apply MLR modeling to our data.

To build the MLR models, we used the `multinom` function from the `nnet` package for R (Venables and Ripley, 2002). For the predictor variables, we considered all the features described in Section 4.3. For the response variable, we considered our Level 1 functions, consisting of five categories: Sentential, Conclusion, Disfluency, Elaboration, and Resumption. We set Sentential as the reference category in the response variable. For each outcome pair of the dependent variable, the `multinom` function returns four values: the regression coefficients, the standard errors, the residual deviance of a model and the Akaike Information Criterion (AIC). The coefficients associated with a given predictor variable are the log-odds for a given probability of having a comparison category as the outcome per unit change in the predictor variable. The residual deviance of a model and the AIC both entail goodness-of-fit. Roughly speaking, residual deviance is an indicator of the amount of data not accounted for by a model. AIC is sensitive to the number of parameters and increases with the number of variables used in a model or with their levels. For both measures, lower scores are better.

We calculated four additional values for each outcome pair of the dependent variable using various functions in R: the $z$ test scores, the $p$ values, the odds ratios, and the confidence intervals (CIs). We calculated the $p$ values using $z$ tests. The $z$ test scores for a given predictor variable were obtained by dividing the predictor’s coefficients by its standard errors, which were then transformed into $p$ values. An odds ratio $> 1$ indicates that the risk of the outcome falling into the comparison category relative to the risk of the outcome falling into the reference category increases as the variable increases. The CI for a given odds ratio informs us of the lower and upper limit of the interval for the odds ratio for the outcome relative to the reference category, given the other predictors are in the model. This is evaluated with a 95% confidence level. For the odds ratios, we used the `exp` function from the `base` package to obtain the exponentiation coefficients. For CI, we used the `tidy` function from the `broom` package (Robinson et al., 2022).

In addition to goodness-of-fit, we also performed some prediction with the models to see how well our considered features can predict discourse functions. We adopted a 70-30 split (70 for the training dataset and 30 for the test dataset) for the data and used the test dataset for prediction. We indicated the model performance with accuracy, which is calculated by the sum of true positives and true negatives over the number of all tokens.

5.2 Overview of the models

Since the gradient variables were each measured by different units, we first performed $z$ score standardization using the `scale` function in R (Becker et al., 1988). Since our goal is to compare Sentential use tokens, which typically occur in the initial position of a DU, we defined `DU_initial` and `PU_initial` as the reference levels for positions in DU/PU. As high-frequency disyllabic words are often contracted or merged (Tseng, 2005), we defined `SYM` as the
reference level for reduction degree. We did this by using the `relevel` function for R (R Core Team, 2021).

To explore how each group of features fits our data, we built separate models, models [1] to [4], and a complex model of all features, model [5]. We looked for the lowest residual deviance and AIC scores returned by the `multinom` function for each model to find the best-fitting one. We then examined the coefficients associated with the predictor variables in the best-fitting model for different outcome pairs of the dependent variable. Table 5 summarizes how each model fits our data.

Model [5] had the lowest residual deviance and AIC score and achieved an accuracy of 55.47% in predicting discourse functions in the test dataset. The degree to which other models fit the data, however, did not seem to be proportionally reflected in the prediction accuracy. For example, although model [5] fits the data much better than model [1b], as indicated by its smaller residual deviance and AIC scores, their accuracies are the same; while the model with the acoustic features (model [2]) has a better fit than the model with the reduction degree features (model [3]), model [3] predicts outcomes slightly better than model [2]. Since model [5] seems to be the best-fitting one, we will report the outcomes of model [5] below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Res. Deviance</th>
<th>AIC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a)</td>
<td>positionalDU</td>
<td>2996.85</td>
<td>3028.85</td>
</tr>
<tr>
<td>(1b)</td>
<td>positionalPU</td>
<td>2958.28</td>
<td>2990.28</td>
</tr>
<tr>
<td>(1c)</td>
<td>positionalDU+PU</td>
<td>2819.31</td>
<td>2875.31</td>
</tr>
<tr>
<td>(2)</td>
<td>acoustic</td>
<td>3147.61</td>
<td>3179.61</td>
</tr>
<tr>
<td>(3)</td>
<td>reduction degree</td>
<td>3176.55</td>
<td>3208.55</td>
</tr>
<tr>
<td>(4a)</td>
<td>contextualDU</td>
<td>3066.90</td>
<td>3098.90</td>
</tr>
<tr>
<td>(4b)</td>
<td>contextualPU</td>
<td>3075.65</td>
<td>3107.65</td>
</tr>
<tr>
<td>(4c)</td>
<td>contextualDU+PU</td>
<td>3058.37</td>
<td>3098.37</td>
</tr>
<tr>
<td>(5)</td>
<td>positionalDU+PU + reduction degree + acoustic + contextualDU+PU</td>
<td>2688.49</td>
<td>2824.49</td>
</tr>
</tbody>
</table>

Table 5: Goodness-of-fit of the Models

### 5.3 Variable performance

The statistical details for each feature of model [5] are summarized in Table 6 in terms of the regression coefficients, the standard errors, the z test scores and p values, the odds ratios, and the confidence intervals.
The implausibly large standard errors, z-test scores, odds ratios, and confidence intervals for the independent variable DU_final, DU_isolated, and MSD may be caused by the presence of empty and small cells in Table 3 and Table 4. It is generally assumed that for independent variables in Multinomial Logistic Regression the cell frequencies should be greater than 1.
Table 6: Statistical Details of Model [5] (with Sentential Use as the Reference Category)

<table>
<thead>
<tr>
<th></th>
<th>CAN</th>
<th>MSD</th>
<th>NUM</th>
<th>Rate</th>
<th>F0_initial</th>
<th>IntensityMean</th>
<th>Previous_duration</th>
<th>Next_duration</th>
<th>PU_rate</th>
<th>DU_rate</th>
<th>Resumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Intercept</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.26</td>
<td>0.42</td>
<td>-0.07</td>
<td>-0.01</td>
<td>0.17</td>
<td>0.12</td>
<td>0.26</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.05</td>
<td>-2.55</td>
</tr>
<tr>
<td>F0_initial</td>
<td>-1.40</td>
<td>0.86</td>
<td>-0.32</td>
<td>-0.11</td>
<td>2.49*</td>
<td>1.78</td>
<td>2.95**</td>
<td>-1.28</td>
<td>0.21</td>
<td>-0.59</td>
<td>-7.64***</td>
</tr>
<tr>
<td>IntensityMean</td>
<td>0.76</td>
<td>1.53</td>
<td>0.92</td>
<td>0.98</td>
<td>1.19</td>
<td>1.12</td>
<td>1.29</td>
<td>0.85</td>
<td>1.02</td>
<td>0.94</td>
<td>0.07</td>
</tr>
<tr>
<td>Previous_duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next_duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU_rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DU_rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.55</td>
<td>0.33</td>
<td>-1.01</td>
<td>0.83</td>
<td>-1.03</td>
<td>-1.03</td>
<td>-11.69</td>
<td>-1.03</td>
<td>-2.19</td>
<td>-0.14</td>
<td>-0.14</td>
</tr>
<tr>
<td>PU_medial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.55</td>
</tr>
<tr>
<td>PU_final</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.01</td>
</tr>
<tr>
<td>PU_isolated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.01</td>
</tr>
<tr>
<td>DU_medial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.03</td>
</tr>
<tr>
<td>DU_final</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-11.69</td>
</tr>
<tr>
<td>DU_isolated</td>
<td>0.11</td>
<td>0.11</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>CAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0_initial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IntensityMean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous_duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next_duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU_rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DU_rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Statistical Details of Model [5] (with Sentential Use as the Reference Category)

5.3.1 Positional variables

Our first analysis investigated the effects of the connective position in relation to a DU/PU on the prediction of discourse functions. The coefficients of the outcome for Conclusion, Elaboration, or Resumption use reached significance in the z test, decreasing relative to Sentential use for DU-medial connectives. The coefficient reached significance only in Conclusion and Resumption uses for DU-final connectives and showed a decreasing likelihood trend. For connectives forming standalone DUs, the coefficients of the outcome being Conclusion or Resumption use are also expected to decrease relative to Sentential use, while those of the outcome being Disfluency or Elaboration use expected to increase relative to that of Sentential use. Particularly, different from Conclusion, Elaboration, or Resumption use, the coefficient of the outcome for Disfluency use is expected to increase relative to that for Sentential use for DU-medial connectives as well as for connectives forming standalone DUs.

For PU-based predictors, the coefficients of the outcome for Conclusion, Disfluency, or Elaboration use are expected to decrease relative to that for Sentential use for PU-medial connectives. For connectives that form standalone PUs, the coefficients of the outcome for Conclusion, Disfluency, or Elaboration use are expected to increase relative to that for Sentential use. Interestingly, we found that Disfluency is the only category for which the
coefficient of the outcome for such use is expected to increase relative to that for Sentential use of PU-final connectives.

In summary, our results show that positions in DU/PU are sensitive to the prediction of discourse functions. Considering DU-related results, we can clearly see that connectives used to indicate a Conclusion or Resumption use seem to prefer initial positions in DU. This is an important finding because connectives used sententially do not always take an utterance-initial position in conversation, even if their conventional position is utterance-initial. On the other hand, PU-related features do not seem to be involved in the prediction of the sentential/discourse use of connectives for Conclusion and Resumption purposes as obviously as DU-related features. Nonetheless, it seems that PU-related features are more effective in predicting a Disfluency or Sentential use in our complex model, as connectives used to indicate a Disfluency use are shown to prefer final positions in PU or standalone PUs.

5.3.2 Phonetic variables

We then investigated the effects of the phonetic features of connectives on the prediction of discourse functions in our model. The coefficients of the outcome for Conclusion or Disfluency use are expected to decrease relative to those of Sentential use when connectives are produced in canonical form, and the coefficients reached significance. That is, when a disyllabic connective is pronounced with more phonetic details, it is more likely to be used sententially. However, this is not the case for Elaboration or Resumption use. The coefficients of the outcomes for all four discourse functions are expected to increase relative to that for Sentential use per point increase in word-initial F0 values, and the coefficients all reached significance. This significant result is clear evidence that the sentential and discourse differences in the use of connectives is reflected in the phonetic representation of the connectives. For word duration, the coefficient of the outcome for Resumption use is expected to increase relative to that for Sentential use per point, and those for Conclusion, Disfluency, or Elaboration use showed a tendency to decrease. Neither of the coefficients, however, was significant. Last, the coefficients of the outcome for Conclusion, Elaboration, or Resumption use showed a tendency to increase relative to that for Sentential use per point of increase in intensity, but without significance, either. In contrast, the coefficient of the outcome for Disfluency use is expected to significantly decrease relative to that for Sentential use per point of increase in intensity.

5.3.3 Contextual variables

Our final analysis investigated the effects of the contextual cues around the connective on the prediction of discourse functions. Our model showed that the coefficient of the outcome for any of the four discourse functions would be expected to increase relative to that for Sentential use per point of increase for the duration of paralinguistic events preceding connectives. All coefficients, except that for Disfluency use, were significant. The coefficients of the outcome for Conclusion, Disfluency, or Elaboration use showed a tendency to decrease relative to that
for Sentential use per point of increase for the duration of paralinguistic events following connectives, but without significance. The coefficient of the outcome for Resumption use also showed a tendency to increase relative to that for Sentential use per point of increase for the duration of paralinguistic events following connectives, without significance. For the overall speech rate of DU, the coefficient of the outcome for Conclusion use is expected to decrease relative to that for Sentential use per point of increase. For PU, the coefficient of the outcome for Disfluency use is expected to increase relative to that for Sentential use per point of increase. Both reached significance.

5.4 Lexico-semantic property matters

Making use of all connective tokens in the corpus, we have observed a number of tendencies between the form and the discourse function of Mandarin connectives. However, we would also like to mention that individual differences nevertheless exist among connectives, as the lexico-semantic meaning of connectives inevitably affects the form and the function of connectives to various degrees. We illustrate this point by taking ránhòu as an example, as its phonetic patterns clearly differ from the general patterns we identified for all connective tokens. This is shown in Table 7 and Figures 4 and 5. Although ránhòu is similar to other Mandarin connectives in the sense that these connectives all canonically indicate a semantic relationship between two adjacent clauses, discourse uses of ránhòu occur predominantly in the DU-initial position, while the distribution of discourse uses of all connective tokens is mostly DU-initial or DU-medial. Furthermore, discourse uses of ránhòu tend to occur more often in standalone PUs than discourse uses of all connective tokens, which have a more even distribution across PU positions. These tendencies seem to correspond to the observations in Wang (2018), where ránhòu exclusively occurs in the turn-initial position when signaling topic shifts. In tokens of ránhòu that close the current turn and invite the hearer’s response, Wang also found lengthening of the connective and occurrence in only standalone intonation units. While more data and a more sophisticated taxonomy of the discourse functions of connectives are needed to better understand the degrees to which the lexico-semantic meaning of connectives may affect the form and the function of connectives, these tendencies suggest that different lexico-semantic properties may render different configurations of positional, phonetic, and contextual encodings in connectives.
Table 7: Distribution of Discourse Functions by DU and PU Positions of ránhòu

<table>
<thead>
<tr>
<th>Discourse Function</th>
<th>DU_initial</th>
<th>DU_medial</th>
<th>DU_final</th>
<th>DU_isolated</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>92 (100%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Conclusion</td>
<td>3 (100%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Disfluency</td>
<td>3 (100%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Elaboration</td>
<td>186 (95.88%)</td>
<td>8 (4.12%)</td>
<td>0</td>
<td>0</td>
<td>194</td>
</tr>
<tr>
<td>Resumption</td>
<td>4 (66.67%)</td>
<td>2 (33.33%)</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discourse Function</th>
<th>PU_initial</th>
<th>PU_medial</th>
<th>PU_final</th>
<th>PU_isolated</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential</td>
<td>57 (61.96%)</td>
<td>20 (21.74%)</td>
<td>4 (4.35%)</td>
<td>11 (11.96%)</td>
<td>92</td>
</tr>
<tr>
<td>Conclusion</td>
<td>0</td>
<td>2 (66.67%)</td>
<td>0</td>
<td>1 (33.33%)</td>
<td>3</td>
</tr>
<tr>
<td>Disfluency</td>
<td>1 (33.33%)</td>
<td>0</td>
<td>0</td>
<td>2 (66.67%)</td>
<td>3</td>
</tr>
<tr>
<td>Elaboration</td>
<td>102 (52.58%)</td>
<td>30 (15.46%)</td>
<td>17 (8.76%)</td>
<td>45 (23.20%)</td>
<td>194</td>
</tr>
<tr>
<td>Resumption</td>
<td>2 (33.33%)</td>
<td>1 (16.67%)</td>
<td>0</td>
<td>3 (50%)</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 4: Boxplot Representation of Discourse Functions by Speech Rate, Initial F0 Value, and Intensity Mean of ránhòu (the red dot indicates the mean)

Figure 5: Boxplot Representation of Discourse Functions by Duration of Paralinguistic Events in the Previous and Next Position of ránhòu and Speech Rate of the DU and PU Where ránhòu Occurs (the red dot indicates the mean)

6. Discussion and conclusion

In our exploratory analysis of all connective tokens in the corpus, we found that these connectives may resume a new or previous topic (Biq, 1994; Didirková et al., 2018; Horne et al., 2001; Rennie et al., 2016; Wang and Huang, 2006; Wang, 2018; Wang, 2005; Wang and Tsai, 2007; Wang et al., 2010; Yang, 2006), introduce a conclusion or summarization of the current topic (Rennie et al., 2016; Wang and Huang, 2006), elaborate on a current or previous
topic (Biq, 2001; Didirková et al., 2018; Rennie et al., 2016; Wang and Huang, 2006; Wang, 2018; Wang, 1998, 2005; Wang et al., 2013; Wang and Tsai, 2007; Wang et al., 2010; Yang, 2006), signal a filled pause or floor holding (Biq, 2001; Rennie et al., 2016; Rhee, 2020; Yang, 2006), or signal repairs (Tseng, 2006). We also observed that certain connectives may emphasize the significance of the proposition, downplay the significance of the proposition, secure the addressee’s attention, or express the speaker's disagreement. These observations seem to be in line with the findings from other Mandarin connective research (Hsieh and Huang, 2005; Wang, 2005; Wang et al., 2013; Wang and Tsai, 2007; Wang et al., 2010; Yang, 2006).

In this paper, we have proposed an applicable scheme that annotates the Sentential, Conclusion, Disfluency, Elaboration, and Resumption uses of Mandarin connectives. Using statistical modeling, we revealed rates of an array of encodings of connectives for the associated discourse functions. This issue related to the interaction between form and function has not been empirically examined in previous research on connectives.

Concerning the positional manifestation of connectives in conversation, tokens that introduce a new topic unit or return to a previous topic most often occur at the beginning of a prosodic phrase or constitute a standalone prosodic phrase (Hirschberg and Litman, 1993; Horne et al., 2001). We have identified differences in positional manifestation in prosodically oriented PUs and semantically oriented DUs. We found that Disfluency use of connectives is oriented more in terms of prosodically determined production units, while Conclusion, Elaboration, and Resumption uses of connectives are oriented more in terms of semantically determined production units. Disfluency use is more likely to occur in the PU-final position and in standalone PUs than Sentential use. Conclusion and Resumption uses are more likely than Sentential use to occur in the initial position of DUs. Elaboration use has similar patterns to those of Conclusion and Resumption uses, except that Elaboration use is more likely than Sentential use to occur in standalone DUs. This may be due to the many subtypes of Elaboration tokens of connectives in our corpus. A larger dataset may provide further insights into this issue. In contrast to the other discourse functions, Disfluency use is more likely than Sentential use to occur in DU-medial positions or form standalone DUs. Our results confirm that connectives that indicate a topic shift in our study, such as Conclusion and Resumption uses, may be better observed in terms of a meaning-oriented segmentation of discourse, as reported in previous research (Rennie et al., 2016; Wang, 2018), rather than in prosodically segmented units.

Lengthened word duration often appears to introduce a new topic or to return to a previous topic (Didirková et al., 2018; Horne et al., 2001; Wang, 2018) and can be used for the automatic detection of discourse markers (Zufferey and Popescu-Belis, 2004). Although our result did not reach significance, it shows a tendency for a lengthened duration in Resumption use and a tendency for a shortened duration in Conclusion, Disfluency, and Elaboration uses relative to Sentential use. While we report that prosodic patterns of connectives may be determined by the associated discourse functions, we require a larger set of spoken connectives to statistically
verify the tendencies in the future. For pitch-related features, we found that all four discourse functions tend to have a higher initial F0 value than Sentential use. If a subject were to increase the initial F0 value, Resumption and Disfluency uses are expected to have the largest increase, followed by a medium increase in Conclusion use and a much smaller increase in Elaboration use. This tendency corresponds to previous findings suggesting that when the topic shifts in conversation, the average onset F0 is higher than for other types of topic boundaries such as topic continuation, elaboration, and speech-act continuation and that elaborating utterances are characterized by the lowest onset F0 among all types of topic boundaries (Nakajima and Allen, 1993). On the other hand, the observation of increased initial F0 value in Disfluency use may be accounted for by the fact that phrase-initial filled pauses have a higher mean pitch than phrase-medial tokens (Swerts, 1998) and that error repairs tend to be marked by increased intonational prominence on the correcting information (Howell and Young, 1991; O'Shaughnessy, 1992). There is also a pitch reset in restarting Mandarin repairs, and the initial F0 in the repaired items is higher than that of its counterpart (Tseng, 2006). Finally, the mean intensity values seem not to be a salient phonetic factor in the context of our connective study. Nonetheless, we observed that Disfluency use tends to have lower mean intensity values than Sentential use. This may be related to the prosodic characteristics of filled pauses described in the literature, where filled pauses are characterized by low intensity as well as low, flat F0 with reduced articulation (Cole et al., 2005).

We have also observed some interesting interactions between the contextual properties and the functions of connectives. We found that Conclusion, Elaboration, and Resumption uses are more likely to have a longer duration of paralinguistic events preceding the connective than Sentential use. This observation is not only in line with the finding that a relatively long pause is relevant for the automatic detection of discourse markers (Zufferey and Popescu-Belis, 2004) but also lends empirical support to the prosodic characteristics of connectives that introduce a new topic or a specification (Didirková et al., 2018). We, however, did not find any significant effect for the duration of the paralinguistic events preceding connectives in the distinction of Disfluency and Sentential use of connectives, even though a clear durational difference can be observed in Figure 2. Furthermore, we did not find any significant effect for the duration of the paralinguistic events following connectives in any of the distinctions of the sentential/discourse uses of connectives. These findings seem to contrast the observations that speakers tend to produce a pause before repair to indicate that the forward flow of speech is being interrupted (Howell and Young, 1991) and where discourse uses of connectives are more often followed by a pause than sentential uses of connectives (Gao and Tao, 2021). The discrepancies may be attributed to two factors. First, the multiple features in our multinomial logistic regression model may have overlapping predictive powers. The effects of the paralinguistic event-based features in predicting discourse uses may be partially, if not fully, captured by other features, resulting in insignificant coefficients for the paralinguistic event-based features in the model.
Second, paralinguistic event-based features may suffer from the problem of data sparseness. While they had similar durational means in our data, only 246 connective tokens (out of a total of 1370 connective tokens) had a Next_duration value larger than 0, and only 518 connective tokens had a Previous_duration value larger than 0. Sparse data could result in a lack of generalization performance. A larger dataset may allow for a more informative look at the association between paralinguistic events and discourse functions. Finally, we found that Conclusion use is more likely to occur in DUs with a faster speech rate than Sentential use, while Disfluency use is more likely to occur in PUs with a slower speech rate. This is partially consistent with the positional distribution of connectives in our data, where Disfluency use was more oriented at the properties related to prosody and Conclusion use was more sensitive to semantically determined production units of discourse.

This study has several limitations. One important issue relates to the fine-grained categories of our taxonomy of discourse functions. Although we collapsed categories that share similar functional properties, such as Emphasis and Downtoning, for the analysis, it is worth acknowledging that these categories also have significant implications for speaker-hearer interactions in discourse and require validation from additional empirical data. As our model can be extended to analyze higher-level functional categories, a fine-grained taxonomy validated by empirical data may give us a better picture of the phonetic orientations for the associated discourse functions and ensure the applicability of our proposed annotation scheme to various speech data. Another issue concerns the effect of the lexico-semantic meaning of a connective on the form of the connective. We mentioned that the intrinsic lexico-semantic property of connectives poses differing degrees of variability to their phonetic orientations, such as in the case of ránhòu. It would be interesting to see how the semantic configuration of a connective may contribute to the overall phonetic forms of the connective. For future studies, a large set of well-annotated speech data and a sophisticated taxonomy of connectives are needed to achieve a deepened understanding of the use of connectives in conversation.

In summary, in this study, we systematically investigated many dimensions of the uses of Mandarin connectives in a large Chinese conversational speech corpus, including their discourse functions and production. Incorporating findings from a linguistic modeling perspective, we have further highlighted the connection between the discourse function and the phonetic form of connectives. We believe that the proposed methodology of integrating discourse function annotation and phonetic analyses can shed more light on the way discourse connectives contribute to the dynamic organization of discourse in conversational speech.

Acknowledgements

This work was supported by the Institute of Linguistics, Academia Sinica and the National Science and Technology Council, Taiwan, with projects [NSC 103-2410-H-001-063-MY2, MOST 106-2410-H-001-045-MY2] granted to the second author. We would like to thank the anonymous reviewers of Dialogue & Discourse for their invaluable feedback.
References


Shuxiang Lü. *Xiàndài hànyǔ bābǎi cí [800 words in Modern Chinese]*. Commercial Press, Hong Kong, 1980.


Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and


124