

Prosody Features of Collaborative Construction in Mandarin Conversation

GUAN Yue

The Hong Kong Polytechnic University – Peking University Research Centre on Chinese
Linguistics, Beijing

guanyue92@sina.com

Abstract

This paper examines the projection of prosody and syntactic structure in conversation, while trying to find out which side is more powerful when there is a conflict between prosodic expression and syntactic structures. The central task of this paper is to look at the phonetic performance of the collaborative construction from the general indicators of prosodic features: duration (length), loudness (intensity) and frequency (pitch). It compares phonetic parameters of the collaborative construction with the adjacent turns and describes it in detail.

1 Introduction

In everyday face-to-face conversation, language forms are always accompanied by prosodic features of speech.

In this paper, the author examines the basic prosody phenomenon in collaborative construction. In daily conversation, a syntactically complete sentence can be co-constructed by different speakers in adjacent turns, which is referred to as collaborative construction (CC)¹ in this paper. The utterance produced by interactional participants

can project the remaining part of a sentence-in-progress, which is the basis of syntactic cooperation.

It has been sufficiently observed that speakers can express their stance and emotion prosodically (Benjamin and Walker 2013); in fact, prosody also plays a very crucial role in turn-taking and turn organization. For example, Walker (2010) examines the mechanism of rush-through in English as a phonetic design of turn-holding and points out that the articulation rate of the final foot at the end of a TCU (Turn Constructional Unit) roughly doubles as that of the preceding one. Local (2005) discusses some prosodic features of collaborative completion when speakers and listeners are undertaking interactional tasks in everyday talk. These studies focus on how prosodic features can be recognized by listeners as a projection of a possible position of turn-taking (Goodwin 1986; Ford 1993, 2004; Couper-Kuhlen and Selting 1996; Walker 2013).

2 Background

This paper takes an Interactional Linguistic perspective and adopts methodology of Conversation Analysis. Interactional Linguistics emphasizes on studying language in its home environment, manifested as conversations in everyday life. Conversations are “cooperatively achieved

¹ Here is an example quoted from Lerner (1987:16):

David: so if one person said he couldn't invest

(.)

Kerry: then I'd have ta wait

The two speakers co-produced one sentence.

objects”, which “need to be adaptable to the emerging and ever-changing trajectory of interaction” (Couper-Kuhlen and Selting 2018). In a conversation, “overwhelmingly, people talk in turns” (Sacks *et al.* 1974). In Conversation Analysis, turn-constructive units (TCUs) are the basic units of conversation that compose turns. TCUs are addressed to their speakers, make relevant the next action and select the next speaker (Lerner 2004). This phenomenon is prevalent in the “compound turn-constructive unit format” (Lerner 1991). Such format may be built with two clauses and sometimes delivered in two separate prosodic units, but is oriented to by participants as one single turn-constructive format (Lerner 1991, 1996). Some researchers have attempted to account for other issues in cross-linguistic perspective based on data from different languages, some of which focus on turns and increments (Ono & Couper-Kuhlen 2007; Luke & Thompson & Ono 2012). Despite the considerable number of research on collaborative TCUs in English (Goodwin and Goodwin 1987; Goodwin 1996; Ford *et al.* 1996), German (Auer 1996; Stivers *et al.* 2009) and Japanese (Hayashi 2005; Iwasaki 2009); studies on Chinese data are still scarce (Song 2019; Guan 2020).

In the field of pragmatics and sociolinguistics, Gumpertz (1982:100) mainly investigated six prosodic features, including intonation, loudness, stress, vowel length, phrasing and overall shifts in speech register. The prosodic features here refer to the pitch, intensity, duration and other qualities that can be recognized by language speakers rather than the acoustic features of the actual voice under the experimental condition. Computers can capture much more accurate acoustic features than human ears, but those are semantically nonsignificant for communication, for the prosodic function of linguistics cannot be directly explained from the experimental data (Gumpertz

1982:108). For example, studies have shown that people's perception of pitch is not the reflection of the actual pitch, but the result of pitch obtrusions as prosodic perception. This kind of consequence separates the utterances that achieves prosodic prominence from that of none saliency, and searches for an approximate preset pattern in the referential mental schemata according to the perceived voice.

With the development of CA, other related factors have been included in the study of prosody. One of them is phonation type. Phonation type refers to the state of glottis when speaking, such as "voiced" and "voiceless". Zhu Xiaonong (2010:66) summarizes 12 kinds of six types of phonation on the basis of previous studies. Phonation studies the change of pronunciation compared to the "default articulatory configuration"². In previous studies, phonation type is mainly used to describe the special pronunciation mechanism of other ethnic minority languages and dialects (Chao 1922, 1929; Zhu Xiaonong 2005, 2010). These studies of phonetics are major and important, but we still lack studies on the ability of human beings to master the sound and to use language. Recently, there are a few relevant research results. For example, Zhu Xiaonong (2004) points out that cracking voice can not only distinguish tones, but also have diminutive meaning; falsetto can also be used as a high-key diminutive tone. In this research on collaborative construction (CC), it is found that, in naturally occurring data, different phonation strategies are widely used in daily conversation, and they have conventionality, which is represented by prosodic projection³.

Prosodic factors sometimes overwhelm syntactic factors on influencing the utterances of speakers and listeners (Ford 1993; Ford and Thompson 1996; Couper-Kuhlen and Selting 2018). For example, if an utterance

² In this paper, the default articulatory configuration refers to that the vocal cord is in a normal state at the beginning of phonating. At this time, the tone value of the voice is "the default pitch". This pitch is the most basic and its value is [32], which is calculated according to Zhu Xiaonong's four-degree system (Zhu Xiaonong 2010).

³ Projection means that the earlier part of a structure foreshadows its later trajectory and thus makes its completion predictable (Couper-Kuhlen & Selting, 2018:39).

does not end in the form of language, but there is a pause caused by hesitation, the listener will often have the opportunity to take the turn. In collaborative-constructed sentences, turns that the preliminary speaker wants to transfer are differently designed in phonetic factors compared with the turns he wants to keep. When the subsequent speaker continues a sentence, it shows an on-going intonation in the prosodic characteristics of the turn, in comparison of pitch, comprehensive intensity and other phonetic features.

3 Data and Methods

This research uses visible data from face-to-face daily conversations in Beijing Mandarin, which were collected from July 2018 to May 2019. These dyad conversations were all naturally occurring interactions between friends and/or acquaintances. During recordings, researchers were absent and thus non-interactive factors could be diminished to the greatest extent. The data were transcribed and annotated to reflect the features of talk-in-interaction including co-existing movements as faithfully as possible.

We adopt an empirical method with a focus on language use from the perspective of Interactional Linguistics. It is widely believed in usage-based approach that ordinary conversation is a primordial site for language and thus constantly shapes forms and meanings of language (Couper-Kuhlen and Selting 2001, 2018). In addition, this study integrates a conversation analytical method to examine cases from our own data. We analyze in detail with reference to the observable orientation of both speakers and recipients within the moment-by-moment unfolding of interaction. Research has shown that multimodal resources jointly work together to build turns and courses of action (Li 2014), and that the turn-taking process is composed of

syntax, prosody, body movements and other pragmatic factors.

4 Findings

Shen Jiong (2003) points out that the hearing system of humans do process phonetic quality, duration, intensity and pitch separately, but they become recognizable and meaningful only after putting together in human brains. This paper is not a special study of phonetics. The discussion in this paper is mainly designed to explore the prosodic performance of CC and its interactional function. This research tries to avoid talking about some blurred situations such as long gap or overlapping turns⁴ just because of difficulties of measurement. Even though, there are still inescapable problems such as dealing with the relationship between "big wave and small wave"⁵ between tone of words and intonation. In addition, there are also the effects of phonation on the pronunciation and the influence of non-verbal factors.

It should be noted that from the perspective of prosody, two types of CC can be clearly distinguished. One is peacefully co-constructed and the other is competitive "collaborative completion". These two types of prosodic features have systematic differences. However, they intertwined with the syntactic categories.

4.1 Duration

Previous researchers on prosodic in Chinese have done a lot of experiments on the length of sound, and have drawn many important conclusions (see Zhu 2010). They find out that the duration of words has an important impact on the recognition of tones and stresses. Through the analysis of the CC data, the preliminary hypothesis is that: compared with the normal turn-taking model formulated by the speaker, the total duration of turns of the CC is smaller. On

⁴ Overlap means more than one speaker talk in a time. This situation will always happen in everyday conversation but it won't keep to long before one party take the turn.

⁵ Chao (1922) describes the relationship between tone of words and intonation as 'small wave rides on the big wave'. That is, each Chinese

character has its own tone. When it is put in a sentence, the intonation of the sentence influences the tone while it still remains the feature of the contour of the tone.

the one hand, this is due to the fact that the words at the end of CC is faster than that of the adjacent turn; on the other hand, the jointly-constructed utterances tend to be shorter, that is, have fewer words. As mentioned above, there are differences between peaceful co-constructions and competitive co-constructions. We randomly selected 30 examples of CC to measure.

The data we measured include but are not limited to: 1) total duration of conversation, 2) number of turns, 3) duration of each speaker's utterance and 4) number of words/characters of each speaker's utterance in a sequence of sub topics.

See excerpt (1):

Excerpt 1

- 1 R 这是高晓松说的,
zheshi Gao Xiaosong shuo de
It's what Gao Xiaosong⁶ said
- 2 就说,
jiushuo
it is said
- 3 历史上纪晓岚这个人就是这样的一个人。
lishi shang Ji Xiaolan zhege ren jiushi zheyang de
yigeren
Ji Xiaolan⁷ is just this kind of person in history
- 4 L 但是我觉得高晓松其实好多话都[有
danshi wo juede Gao Xiaosong qishi haoduohua
dou you...
but I think Gao Xiaosong's words often...
- 5 R [不能信,
buneng xin
can't be trusted
- 6 是吧?
shiba
right?
- 7 L 对对对。
duiduidui

- 8 right
比如说他骗你的那个,
biru shuo ta pianni de nage
such as what he once cheated you
- 9 R 就,
jiu
just...
- 10 骗我的啥来着?
pian wo de sha laizhe
what did (you refer to by saying) he cheated me?

According to the research needs, we put the audio of the above transcribed corpus into Praat software to get a series of data.

Table 1 Parameter in excerpt (1)

Starting time	Duration (s)	Sylla -bles	Lines	Speaker
00:34:59	1.006	7	1	R
00:35:00	0.506	2	2	R
00:35:00	2.089	17	3	R
00:35:03	2.660	15	4	L
00:35:06	0.524	3	5	R
00:35:07	0.386	2	6	R
00:35:07	0.507	3	7	L
00:35:07	1.173	9	8	L
00:35:09	0.310	1	9	R
00:35:10	1.088	6	10	R

Basic parameters:

Total time: 10.249(s); Total characters: 65; turns: 10;
Speakership changes: 4; Voiced time: 10.328(s)

Calculated items:

Average length: 0.158; Length of the preliminary part
(line 4): 0.177s; Length of the latter part (line 5): 0.175s

By making a statistical analysis of the examples of cooperation and co-construction we have collected, we get the following table:

Table 2 Duration parameters of random CC excerpts

⁶ A Chinese famous talk show host who is good at talking about the history.

⁷ Ji Xiaolan is a prime minister in feudal China. He is famous for his loquacious speaking-talent and knowledgeable, which is a positive

figure admittedly. However, in Gao's talk show, he is said to be a bad person.

Excerpt ⁸	total time duration (s)	total characters	speed (word/s)	speed of former part	speed of latter part	relation of the two part	overlap or not	type of response in third-position	type of CC
O-06	5.438	30	0.187	0.227	0.154	>	Y	positive	rush-though
S-11	5.639	33	0.171	0.145	0.209	<	N	neutral	collaborative completion
AE-52	6.603	41	0.161	0.181	0.301	<	N	positive	collaborative completion
J-03	5.759	31	0.186	0.246	0.201	>	Y	neutral	rush-though
AA-10	11.343	73	0.155	0.143	0.192	<	N	neutral	after-thought
B-12	3.566	24	0.149	0.225	0.147	>	Y	positive	rush-though
A-01	23.765	144	0.165	0.182	0.178	>	Y	positive	rush-though
.....									

Base on the analysis of the data in the table, it is found out tendentious rules at least: 1) the speaking speed of the former part of CC is slower than that of the adjacent turn of the same sequence; 2) when the speaking speed of the latter part of CC is higher than that of the former one, they often overlap. These two conclusions are applicable to all kinds of CC. Moreover, it is found that CC types are partly determined by the time difference between the two parts of CC. More precise results are still being calculated.

4.2 Stress

According to our observations of the data, the collaborative completion of a CC would be produced in a mild way by speaker within the controllable range, i.e., the collaborative speaker often artificially packages his own utterances as the completion of the previous turn of the other party by using a similar pitch as the previous speaker in the sense of hearing. The prosodic operation is mainly manifested in the special design on making the following utterance sound like a continuation rather than a disjunctive one. This strategy of

"loudness-matched" has been studied by some scholars (Local 1992, 2005; Szczepek 2000).

One technical problem is that, when collecting data, researchers should pay attention to the locating place of the recording equipment. The distance from the device to the two speakers should be equidistant. After verifying these, the results still show that there is no obvious difference of higher intensity or stress on the sound spectrogram. The speaker's voice intensity is always maintained at a relatively stable level. It is the pitch and duration that determine the stress in listening comprehension (Chao Yuen-Ren 1968; Lǚ Shuxiang 1979). Furthermore, stress is not the only way to highlight foreground information. In Chinese, words are often disyllabic and have their own cadence. In addition, due to the reason of the flow of speech, stress mainly relates to syntactic units rather than interactive units.

However, in the cases of less cooperative type of CC, it tends to use a recognizable stress using for preempting turns. For example, in the following example, the voice intensity of the co-constructed speaker is intentionally higher than that of the previous speaker in line 14:

⁸ The numbers are original ones.

Excerpt 2

1	L	挣得太多了钱。 zhengde taiduo le qian they earn too much money			
2		他们挣得太容易了那钱。 tamen zhengde tairongyi le na qian the way they earn money is too easy		10	R
3		他占太多了。 tamen zhan tai duo le they occupy too much		11	
4		演一部电视剧这么多钱。 yan yibu dianshiju zheme duo qian playing movies can earn a lot of money		12	
5		老百姓科学家能这么多钱啊， laobaixing kexuejia neng zheme duo qian a Can common people or scientists earn so much money?		13	
6		一辈子都拿不来这么多钱来。 yibeizi dou nabulai zheme duo qian lai they won't get so much with all their life		14	L
7		尤其是科学家， youqi kexuejia especially those scientists		15	R
8		上个学， shang ge xue highly educated		16	
9		不如他们那不上学的都。 buru tamen na bushangxue de dou not as good as those who haven't educated		17	

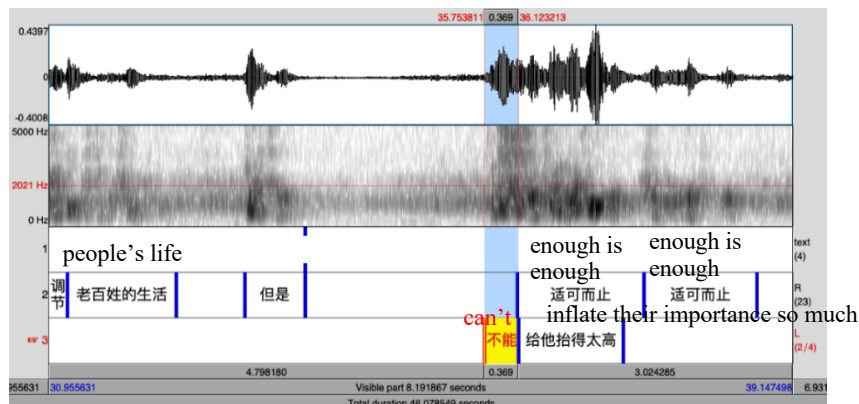


Figure 1 Sound spectrogram in line 14-15

In Excerpt 2, two retired professors are talking about the income among different people. Both of them hold a negative stance to the high-income in the entertainment industries. However, L's negative emotion is much stronger. In line 13, the adversative conjunction *danshi* "but" produced by R indicates his reservation on popular entertainment. However,

the CC completion produced by L in line 14 shows that she thinks that R's speech is not fierce enough to express L's mood. Therefore, the sound intensity of L in this turn is obviously enhanced, and the voice of the negative word "no" and degree adverb "too" is significantly enhanced.

4.3 Pitch

In linguistic prosodic analysis, pitch is another important indicator of prosodic features in addition to the above-mentioned duration and loudness. As we all know, for the same individual's pronunciation, relative pitch not only distinguishes tones in words, but also distinguishes the intonation in sentences. Pitch is a core acoustic feature of the perception of stress in phonetics. Somewhat differently, in Mandarin Chinese, the main representation of word stress is duration, which is due to the fact that Chinese has taken tone (presented mainly by pitch) as an important means to distinguish meaning of words (Zhu Xiaonong, 2010). However, the pitch line in the sound spectrogram is still the most obvious evidence.

Shen et al. (1994) point out that the effect on focus stress of duration is not significant, but that of pitch is very important. Wang Yunjia et al. (2016) also point out that the initial pitch, focus pitch and final pitch of intonation are relatively constant. These studies on focus stress and

intonation reflect the particular emphasis has been placed on information structure in linguistic research. But these studies are mainly involved in one speaker (monologue). This paper researches everyday conversation in its natural habitat and chooses a perspective of interaction to examine how different speakers negotiate and collaboratively achieve language tasks.

The relationship between Chinese intonations and tones has been fully discussed in the existing Chinese phonetic studies. Chao (1922, 1968) proposes theories like "rubber band effect", "algebraic sum" and "big wave and small wave" to describe this phenomenon. The outline of Chinese intonation concluded by these outstanding previous studies and experiments also has a high consistency in the data of CC, which indicates that the collaboratively constructed design of prosody acts as a part of coding and recognition of CC.

Shen (1985) comes up with the intonation structure of declarative sentence in Mandarin Chinese.

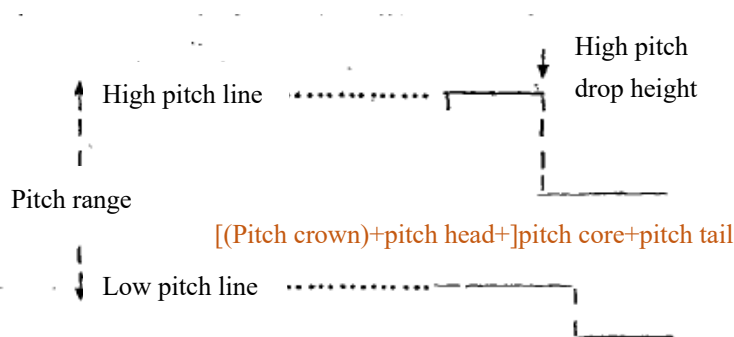


Figure 2 Intonation structure of declarative sentence in Mandarin Chinese (Shen 1985:21)

The following example of CC is a declarative sentence.

Excerpt 3

- 1 R 我感觉你对这个东西并没有那么深入的热爱()和喜欢,
wo ganjue ni dui zhege dongxi bingmeiyou name shenru de
re'ai he xihuan
I feel you to this thing not really very deep
love and like
- 2 L 所以,
Suoyi

So

- 其实我深度热爱和喜欢的东西就是,
qishi wo shendu re'ai he xihuan de dognxi jiushi
in fact I deep love and like thing COP
In fact, the thing I love most is,
- 4 R 画画儿。
Huahuar
Drawing
- 5 L 嗯
En

6 Hm
 (1.0)
 7 R p<那你学画画儿可以啊.
 na ni xue huahuar keyi a
 So you can learn to draw

Speakers collaboratively construct the utterance
 in line 3-4. The sound spectrogram shows the pitch
 inline 3-5 as below.

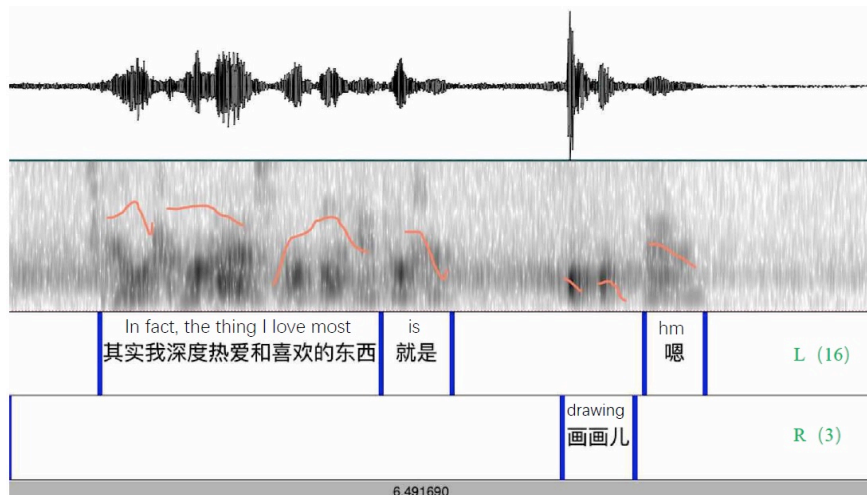


Figure 3 Sound spectrogram in line 3-5

As Figure 3 shows, although there is a gap between the collaboratively completed object produced by R from the previous turn, the intonation of the whole sentence still conforms to the pattern of Mandarin declarative sentence as shown in Figure 2. On the other hand, R's turn does not have the prosodic feature of one-member sentence, showing a phonetic design of collaborative completion.

Another function of pitch design for turn organization is 'recall'. The speaker suggests that the content of the speech is a recall of another past scene or an image of the virtual world via a significant tone change (usually using a high tone). In our data, there are such examples of using prosodic methods to simulate conversational scenes and project subsequent turns.

Excerpt 4

1 R 你还记得我之前说过么,
 ni hai jide wo zhiqian shuo guo me
 you still remember I before say PAT
 Do you still remember what I have said before
 2 L 什么?
 Shenme
 What?
 3 R 就是我跟你说过的,
 jiushi wo genni shuoguo
 It is that I have said to you,
 4 我说我们有一男老师,

5 wo shuo women you yi nanlaoshi
 I said there was a male teacher in our school
 我想夸他,'您真年轻'.
 wo xiang kua ta 'nin zhen nianqing'
 I wanted to flatter him 'you look so young'
 6 L @哈哈@
 @haha@
 7 噢,我[记得啊]
 o wo jide a
 Yeah, I remembered
 8 R [我说]
 wo shuo
 I said:
 9 我说'您看着就跟三十多岁似的'.
 wo shuo nin kanzhe jiu gen sanshiduosui shide
 I said 'nin look like only thirties.
 10 L ['就是三十多']
 jiushi sanshiduo
 'it is thirties'
 11 R [然后他说,]
 ranhou ta shuo
 then he said
 12 '我就是三十多呀'
 wo jiushi sanshiduo ya
 'I AM thirties.'
 13 我说,我说'没有,'
 wo shuo,woshuo meiyou
 I said, I said 'no,'
 14 我说'您看着像三十多'
 wishuo 'nin kanzhe xiang sanshiduo'
 I said 'you just look like thirties'

- 15 ‘对,我就是三十多’
 ‘dui wo jiushi sanshiduo’
 ‘Yeah, I am thirties’
- 16 @就是他@
 jiushi ta
 it is him

This excerpt is about reconstructing the past scene. R describes an awkward conversation when she first met a colleague and wanted to flatter him by underestimating his age. Unfortunately, what she estimated is actually the colleague's real age. In line 9, she imitates the tone of "dialogic style" in prosody, which is shown as exaggerated high tone, so as to mark that the narrative behavior is a scene reappearance. The content projection is followed by the reply of the colleague. What L produces in line 10 is the responding utterance that the other person in the story answered.

Such examples show that in conversation, prosodic projection and syntactic projection occupy different channels, but they are closely related and finally work together. Moreover, one prosodic feature may not be projective in one context (speaker, conversation scene, etc.), but in another context, it becomes pretty relevant and gets highlightedly interpreted.

5 Conclusion

The main research method of this paper is to describe the characteristics of CC through Praat spectrogram. But the purpose of showing the spectrograms and parameters is not to extract numbers, but to verify what is consistent with the speaker's listening sense in communication. In fact, the sounds that the conversation participants "hear" are not same as the sounds "heard" by the machine. In addition, auditory discrimination of human ears has its own focuses, and this kind of 'selective' hearing has complex social base.

Prosody is a very important factor in CC. According to the data, the former speaker will have some special prosodic performance when projecting the continuing turns to be co-constructed, such as to slower speaking speed or to delay length of the syllable at the end of the turn. Also in the "reappearance" scene, the former speaker would use

some special prosodic performances like the exaggerated tones to project the following turns.

6 References

- Allie Hope King 2018 Collaborative completions in everyday interaction: A literature review. *Applied Linguistics and TESOL*, 18(2), 1-14.
- Auer, Peter 1996 On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation. *Pragmatics* 6(3), 371-381.
- Auer, Peter 2005 Projection in interaction and projection in grammar. *Text* 25(1), 7-36.
- Benjamin, Trevor and Traci Walker 2013 Other-initiated repair. In Carole A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, Hoboken, NJ: Wiley-Blackwell.
- Bolden B. Galina 2003 Multiple modalities in collaborative turn sequences. *Gesture* 3(2), 187-212. Amsterdam: Benjamins.
- Chao, Yuen-Ren 1968 *A Grammar of Spoken Chinese*, Berkeley: University of California Press.
- Chao, Yuen-Ren 2002[1922] *Zhongguo Yuyan Zidiao di Shiyan Fangfa* [Experimental study of Chinese word tones]. *Sciences*, 7:9, 871-882.
- Couper-Kuhlen, Elizabeth 2001 Interactional prosody: High onsets in reason-for-the-call turns. *Language in Society* 30:29-53.
- Couper-Kuhlen, Elizabeth and Margret Selting (eds.) 1996 *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, Elizabeth and Tsuyoshi Ono (eds.) 2007 Turn continuation in cross-linguistic perspective. *Pragmatics* 17(4), Special issue.
- Couper-Kuhlen, Elizabeth and Selting, M. 2011 *Studies in Interactional Linguistics*, Amsterdam: Benjamins.
- Couper-Kuhlen, Elizabeth and Selting, M. 2018 *Interactional linguistics: studying language in social interaction*. Cambridge: Cambridge University Press.
- Du Bois J. W., Schuetze-Coburn, S., Cumming S. and Paolino D. 1993 Outline of discourse transcription. In Jane A. Edwards and Martin D. Lambert (eds.), *Talking Data: Transcription and Coding in Discourse Research*, 45-89. Hillsdale, NJ: Lawrence Erlbaum.

- Fang, Di and Yue Guan 2019 The function of final conjunctions in Mandarin Chinese conversation. Paper presented at *the 16th International Pragmatics Conference*, 9-14 June, Hong Kong Polytechnic University.
- Fang, Mei 2019 Collaborative completion as speech act alignment in Chinese conversation. Paper presented at *the 16th International Pragmatics Conference*, 9-14 June, Hong Kong Polytechnic University.
- Ford, Cecilia E. 1993 *Grammar in interaction*. Cambridge: Cambridge University Press.
- Ford, Cecilia E. 2004 Contingency and units in interaction. *Discourse Studies* 6:27-52.
- Ford, Cecilia E. and Sandra A. Thompson 1996 Interactional units in conversation: syntactic, intonational, and pragmatic resources for the projection of turn completion. In Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson (eds.), *Interaction and grammar*, 134-184. Cambridge: Cambridge University Press.
- Fox, Barbara A., Sandra A. Thompson, Cecilia E. Ford and Elizabeth Couper-Kuhlen 2013 Conversation analysis and linguistics. In Jack Sidnell and Tanya Stivers (eds.), *The Handbook of Conversation Analysis*, 726 – 740. Malden: Wiley-Blackwell.
- Goodwin, Charles 1981 *Conversational organization: interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, Charles 1986 Audience diversity, participation and interpretation. *Text* 6: 283-316.
- Goodwin, Charles and Marjorie Harness Goodwin 1987 Concurrent operations on talk: Notes on the interactive organization of assessments. *IPrA Papers in pragmatics* 1:1-54.
- Guan, Yue 2020 A Study of Syntactic Collaborative Construction in Mandarin Conversations. Doctoral Dissertation. The Graduate School of Chinese Academic of Social Sciences..
- Guan, Yue and Mei Fang 2020 Hanyu duihua zhong dejufa hezuogongjian xianxiang chutan [A preliminary study of syntactic collaborative construction in Mandarin conversations], *Yuyan Jiaoxue yu Yanjiu [Language Teaching and Linguistics Studies]*, 203(3):60-69.
- Hayashi, Makoto 2005 Joint turn construction through language and the body: Notes on embodiment in coordinated participation in situated activities. *Semiotica* 156:21-53.
- Iwasaki, Shinako 2009 Initiating interactional turn spaces in Japanese conversation: Local projection and collaborative action. *Discourse Processes* 46: 226-246.
- Lerner, Gene H. 1987 Collaborative turn sequences: entence construction and social action. Unpublished Doctoral Dissertation. University of California, Irvine.
- Lerner, Gene H. 1991 On the syntax of sentences-in-progress. *Language in Society* 20, 3: 441-458.
- Lerner, Gene H. 1992 Assisted storytelling: Deploying shared knowledge as a practical matter. *Qualitative Sociology*, 15, 247–271.
- Lerner, Gene H. 1995 Turn design and the organization of participation in instructional activities. *Discourse Processes* 19:111-131.
- Lerner, Gene H. 1996. On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker. In Elinor Ochs, E. A. Schegloff and Sandra A. Thompson (eds.), *Interaction and grammar*, 238-276. Cambridge: Cambridge University Press.
- Lerner, Gene H. 2004 On the place of linguistic resources in the organization of talk-in-interaction: Grammar as action in prompting a speaker to elaborate. *Research on Language and Social Interaction*, 37(2),151-184.
- Lerner, Gene H. and Takagi, Tomoyo 1999 On the place of linguistic resources in the organization of talk-in-interaction: A co-investigation of English and Japanese grammatical practices. *Journal of Pragmatics* 31, 49-75.
- Levinson, Stephen. C. 1983 *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen. C. 2013 Action formation and ascription. In Jack Sidnell and Tanya Stivers (eds.), *The handbook of conversation analysis*, 103-130. Malden: Wiley-Blackwell.
- Li, Xiaoting 2014 *Multimodality, Interaction and turn-taking in Mandarin Conversation*. Amsterdam: Benjamins.
- Local, John 2005 On the interactional and phonetic design of collaborative completions. In William J. Hardcastle and Janet Mackenzie Beck (eds.), *A figure of speech: A festschrift for John Laver*, 263-282. Hillside, New Jersey: Lawrence Erlbaum.
- Luke, Kang Kwong, Sandra A. Thompson and Tsuyoshi Ono 2012 Turns and increments: a comparative perspective. *Discourse Processes* 49:155-162.

- Lǚ Shuxiang 1979 *Hanyu Yufa Fenxi Wenti*, Beijing: Shangwu Yinshuguan.
- Ochs, Elinor, Emanuel A. Schegloff and Sandra A. Thompson (eds.) 1996 *Interaction and Grammar*. Cambridge: Cambridge University Press.
- Ono, Tsuyoshi and Elizabeth Couper-Kuhlen 2007 Increments in cross-linguistic perspective: Introductory remarks. *Pragmatics* 17(4):505-512.
- Ono, Tsuyoshi, Sandra A. Thompson and Yumi Sasaki 2012 Japanese negotiation through emerging final particles in everyday talk. *Discourse Processes* 49:243-272.
- Sacks, Harvey 1992 *Lectures on Conversation, 2 Volumes* (Fall 1964-Spring 1972). Oxford: Blackwell.
- Sacks Harvey, Emanuel A. Schegloff and Gail Jefferson. 1974 A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696-735.
- Schegloff, Emanuel. A. 1968 Sequencing in conversation opening. *American Anthropologist* 70:1075-1095.
- Schegloff, Emanuel. A. 1979 The relevance of repair to syntax-for-conversation. In Talmy Cívón (ed.), *Discourse Syntax*, 261-286. New York: Academic Press.
- Schegloff, Emanuel A. 1982 Discourse as an interactional achievement: Some uses of ‘uh huh’ and other things that come between sentences. In Deborah Tannen (ed.), *Analyzing Discourse: Text and Talk*, 71 -93. Washington, D.C.: George Town University Press.
- Schegloff, Emanuel A. 1996 Turn organization: one intersection of grammar and interaction. In Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson (eds.), *Interaction and Grammar*, 52 – 133. Cambridge: Cambridge University Press.
- Schegloff, Emanuel A. 2007 *Sequence Organization in Interaction: A Primer in Conversation Analysis, Vol. 1*. Cambridge: Cambridge University Press.
- Shen, Jiong 1985 Beijinghua shengdiao de yinyu he yudiao [The range of tones and intonation in Beijing Mandarin], In Tao Lin and Lijia Wang (ed.), *Beijing Yuyin Shiyuanlu* [Experimental Analysis on Beijing Mandarin] 73-125. Beijing: Beijing University Press.
- Shen, Jiong 1992 Hanyu yudiao moxing chuyi [On the pattern of Chinese intonation]. *Yuwen Yanjiu* 4:16-24.
- Shen, Jiong 2003 Language in running speech [Yunxing zhongde yuyan], *Linguistic Sciences [Yuyan Kexue]* 2:23-28.
- Song, Zixuan 2019 Collaborative Construction of Turn Units in Mandarin Conversation. Master Thesis. University of Alberta.
- Tao, Hongyin 1996 *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam: Benjamins.
- ten Have, Paul 2007[1999] *Doing Conversation Analysis*. London: Sage Publications Ltd.
- Thompson, Sandra A., Barbara A. Fox and Elizabeth Couper-Kuhlen 2015 *Grammar in Everyday Talk*. London: Cambridge University Press.
- Walker, Gareth 2010 The phonetic constitution of a turn-holding practice: Rush-throughs in English talk-in-interaction. In D. Barth-Weingarten, E. Reber and M. Selting (eds.), *Prosody in interaction*. Amsterdam: Benjamins. 51-72.
- Walker, Gareth 2012 Coordination and interpretation of vocal and visible resources: “Trail-off” conjunctions. *Language and Speech* 55(1): 141-163.
- Walker, Gareth 2013 Phonetics and prosody in conversation. In Jack Sidnell and Tanya Stivers (eds.), *The handbook of conversation analysis*. 55:141-163. Malden: Wiley-Blackwell.
- Wang Yunjia, Higashi Takahiro, Dayoung Jeong 2016 Stability and variation of sentence focus and ending pitches and relevant issues [Jiaodian he jumo yingao de hengding, bianyi jiqi xiangguan wenti], *Essays on Linguistics [Yuyanxue luncong]* 2:66-99.
- Wu Zongji 1982 Putonghua Yuju zhong de Shengdiao Bianhua [Tonal changes in sentences in standard Chinese]. *Zhongguo yuwen* 6:439-449.
- Zhu, Xiaonong 2004 Intimacy and high pitch [Qinmi yu gaodiao], *Contemporary Linguistics [Dangdai Yuyanxue]* 6(3):193-222.
- Zhu, Xiaonong 2005 Experimental phonetics and its contributions to Chinese linguistics [Shiyan yuyinxue he hanyu yuyin yanjiu], *Nankai Linguistics [Nankai Yuyanxuekan]* 1:1-17.
- Zhu, Xiaonong 2010 *Yuyinxue [Phonetics]*. Beijing: Shangwu Yinshuguan.

Appendix

A. Transcription convention

,	mid-rise final pitch
.	low-fall final pitch
ˊ	slightly rise question intonation
?	high rise pitch
(0.5)	paused for 0.5 seconds
(.)	micro pause, often less than 0,2 seconds.
=	latching
-	sudden stop
: : :	voice lengthening, the more colons, the longer the voice
↑	pitch step up
↓	pitch step down
hhh	hearable outbreaths, the more “h”s, the longer breathing
<	inhaled sound, like p<, s, ts<, and so on
ʔ	glottal stop
> <	compressed or rushed
< >	slowed or drawn out
[]	overlap
((<i>movement</i>))	non-linguistic actions
(<i>word</i>)	uncertain transcription
@.....@	words with laughter
→	target line which is referred to in the text
bold	target words

Note: This transcription system basically adopts DT (Discourse Transcription) and its revised edition DT2 (Du Bois et al. 1993, 2006), with a few small modifications according to Chinese data.

B. Glossing convention

AUX	auxiliary word (<i>de hua</i>)
COP	copular (<i>shi</i>)
NEG	negatives (<i>bu</i>)

POSS	possessive (<i>de</i>)
PRT	particle
Q	question marker
3SG	singular