



Discourse Planning Strategies in Chinese L2 English Speech: Chunking Preference and Disfluency Attributes

Yiyuan Liao¹, Jue Yu¹

¹ School of Foreign Languages, Tongji University, Shanghai, China
meredithliao@tongji.edu.cn, erinyu@126.com

Abstract

This paper mainly discusses the on-line planning strategies by observing chunking preference and disfluency attributes utilized by Chinese L2 learners in English discourse speech. The main purpose is to reveal the discourse-level speech planning shaped by the richness and density of the moment-by-moment planned message in Chinese L2 English. The results show that compared with English native speakers, 1) as a global property of the spoken discourse as a whole, Chinese L2 learners were temporally more disfluent, meanwhile with a greater in-group variabilities due to the insufficient L2 proficiency; 2) as for the semantic macro-planning strategies, Chinese L2 learners usually generated rather smaller chunks (especially 2-word chunks) and less flexible chunk size in on-line discourse speech, thus leading to a smaller planning size of speech output and a stable pattern of non-native-like ultimate attainment; 3) at the micro-planning level, Chinese L2 learners produced more hesitation markers (especially silent pauses), differed categorically in repair strategy and failed to strategically take advantage of certain disfluency markers to buy plan time so as to increase the complexity of utterance in on-line discourse speech.

Index Terms: discourse planning, disfluency, Chinese L2 learners, English spontaneous speech.

1. Introduction

According to Levelt's model, speech production is comprised of three main planning phases, namely conceptualization, formulation, and articulation [1]. When a speaker wants to convey a communicative message, he starts planning his utterance first through conceptual preparation. Then he formulates a linguistic plan in detail, including selecting lexicon and constructing grammatical frames of the content. Finally, he gives a phonological shape to this internally formulated content and uses the articulatory plan to produce it. Monitoring is believed to occur before and after speech production. This model of speech planning and production is originally for the monolingual speaker but it is believed that producing speech in a second language (L2) resembles the processes involved in speaking one's native language (L1) [2] [3]. To be more specific, some of the macroplanning processes are language-universal, such as the elaboration of the communicative intention at the conceptual level; while the construction of the preverbal message through microplanning, such as information structure, L1 vs. L2 lemma's, morpho-phonological codes etc. at the formulation and articulation level, is language-specific, which may very likely cause difficulties for L2 learners.

In recent years, there is a growing body of research examining the effect of planning on second language (L2)

production. Previous literature has demonstrated that task design features and planning time conditions exert great influence on complexity, accuracy and fluency in L2 learners' speech production performance. For example, under the time pressure, L2 speakers will have difficulties in conceptual planning and linguistic encoding, and accordingly fail to produce fluent contributions to a spoken discourse [4] [5] [6]. But most of such research has examined the effects of pre-task planning on the L2 performance, limited research has considered the effects of online planning. In fact, on-line planning requires L2 learners to plan their oral discourse within the task performance, which can be a greater challenge for L2 speakers. What's more, although many task-based planning studies confirmed there is a trade-off effect (that is, cognitively L2 learners are frequently regulated by one dimension of language performance at the expense of another) between either fluency and accuracy or accuracy and complexity in L2 speech performance [7] [8] [9] [10], so far there is no sufficient analysis with regard to the relationship between fluency and complexity.

This paper, different from most of previous studies, will use online spontaneous speech rather than read out or pre-task planning speech as speech material, and examine discourse planning strategies at both macro-and micro-planning level in Chinese L2 English speech. It is pointed out that the macro-planning deals with long range semantic organization of a sizable chunk of speech and the micro-planning concerns with purely local functions, like markers labelling clause boundaries and selecting words [11]. Therefore, we will mainly focus on the chunking preference and disfluency attributes at the level of discourse-level speech planning organization in L2 speakers. Although early studies negatively regarded disfluency as a speech error or an indicator of poor proficiency in L2 speech production, we extend its function, according to recent studies, as a character of speech revealing speech planning and production [12] [13] [14]. It is said that non-native speakers are quite ineffective in applying the chunking strategies and the proper disfluency marker to get the intended message across [15]. For example, L2 English speakers used more intermediate chunking units and fewer larger-scale planning units compared to L1 English speakers in discourse speech. Moreover, higher-level discourse planning through patterns of chunking and phrasing may further cause prosodic deviations in L2 English in the study by [16]. We hypothesize that L1 and L2 speakers differ in the utterance production planning mechanism they recruit for an on-line task, especially in the discourse context. Three research questions are proposed:

1. What are the differences in temporal fluency between English L1 and Chinese L2 speakers, as a global property of the spoken discourse as a whole?
2. How do Chinese L2 speakers differ from English L1 speakers in the semantic macro-planning strategies in

terms of chunk size? Are there any preferences for particular chunk size in L2 on-line discourse speech?

- How do Chinese L2 speakers and English L1 speakers differ strategically in utilizing disfluency at the micro-planning level in the discourse context? Are there any preferences for particular disfluency markers in relation to utterance complexity?

2. Method

2.1. Data

An online-planning speech task is elicited by the topic “What Do You Do in Your Spare Time”. The reason why this topic is chosen, is that it is closely related to everyone’s daily life, and therefore subjects are able to produce connected speech by telling their own personal experiences with a legitimate challenge. Although it is a sort of special manifestation of comprehensive input, there are several gaps between what they try to express and what they can actually express for most speakers, especially L2 learners, thus leading to unavoidable situations of disfluency in their speech performance. Disfluency in this paper, is regarded as a character of speech revealing speech planning and production, for example, bringing prosodic cues to indicate new messages or helping speakers to buy some time for planning the upcoming utterances [17].

Before the recording, the subjects were notified that they were required to make a 3-minute speech based on the topic given by the interviewer. During the recording, the interviewer first had a free chat with the subjects to make them feel relaxed and then gave the task. To ensure spontaneity and naturalness, disfluency was not controlled at all during their speech, so the disfluency markers can be completely investigated. On the account of various degrees of proficiency and different phonostylistic characteristics, the average length of the spontaneous passage varied from 6.1 utterances among Chinese L2 speakers to 7.8 utterances among English L1 speakers.

2.2. Subjects

Altogether, 20 adult Chinese L2 speakers (6 male and 14 female) were recruited and 6 American English speakers (2 male and 4 female) were used as a baseline. Chinese L2 speakers are all college students, with different discipline background (e.g. arts, science, and engineering) but from the same dialect region--Ningbo. Therefore, the influence of dialects and learning environments can be minimized. Moreover, they have studied English as a second language for at least 10 years, and have passed College English Test (CET 4 and CET 6), which serves as a reliable assumption that these L2 speakers can complete the oral task without so much difficulty.

2.3. Annotation

All the speech recordings of Chinese L2 and English L1 speakers were annotated with the help of SPPAS and Praat. The segmental labelling was first automatically annotated using the SPPAS tool [18], and then manually post-edited. For the disfluency labelling, all work was done manually by the trained transcribers and double checked by the authors.

Based on the classification scheme described in [19], two types of disfluency markers were used: hesitation markers and self-monitoring markers (see Fig.1). According to [20] and the research aims of this paper, 100ms was chosen as the cut-off point of silent pause.

Two Types of Hesitation Markers	
1. Silent pause/Unfilled pause:	It refers to phonetically empty interruptions.
2. Filled pause:	It refers to interruptions with sounds like “uh” or “er”.
Four Types of Self-monitoring	
1. Repetition:	Repetition occurs when the speaker say the whole or part of language unit (phoneme, word, phrase, etc.) again. E.g. “because I think taking photos can record # record my life” (“#” is coded as the location of silent pause)
2. Instant repair:	Instant repair happens when a different language unit is substituted for another language unit to avoid a troublesome or misused chunk said before. E.g. “and sometimes I enjoy play # playing games...” (“#” is coded as the location of silent pause)
3. Anticipatory retracing:	Anticipatory retracing describes a situation where the speaker goes back to what he or she said before or some previous point. E.g. “...reading some books about some # books about the literature...” (“#” is coded as the location of silent pause)
4. Restart:	Restart means that the speaker abandons the original linguistic structure to form a new start with little connection to the former one. E.g. “...I would like # I have less spare time” (“#” is coded as the location of silent pause)

Figure 1: Classification of disfluency markers.

2.4. Data analysis

Temporal fluency, as a global property of the spoken discourse, was first measured by the speech/pause relationship in L1 and L2 English speech [21]. Then the incidence and distribution of various chunk size were carefully compared between L1 and L2 speakers; thus the different semantic macro-planning strategies in terms of chunking pattern were observed. In this paper, speech chunk was defined by the number of words of each corresponding interpausal unit. Finally, disfluency attributes at the level of both discourse-level and utterance-level organization in L2 and L1 English speech were examined in order to unfold the different micro-planning strategies. Meanwhile, the relationship between syntactic complexity and fluency were further analyzed. The basic syntactic complexity was traditionally measured by word count per T-unit (According to [22], a T-unit is a dominant clause and its dependent clauses). But in the present study, we did not use this index. We believe this index can only reveal the length of utterance but not the real syntactic complexity. Therefore, we focused on the syntactic structure into consideration. Only those utterances with subordinate structure were considered as syntactic complex ones. By doing all these above measurements, the discourse speech planning shaped by the richness and density of the moment-by-moment planned message can be revealed both globally and locally in Chinese L2 English.

Besides, the statistical analysis of homogeneity variance and independent sample t-test have been applied in this paper.

3. Results and discussion

3.1. L1-L2 differences in temporal fluency

In this section, the temporal fluency of English L1 and Chinese L2 speakers will be compared. Figure 2 shows the difference in temporal fluency between English L1 and Chinese L2 speakers in terms of speech rate and articulation rate. Speech rate in this paper was calculated by semantic units (words) per second [23], while articulation rate was calculated in a similar way except for the total speech time minus the total pause duration.

The result shows that English L1 and Chinese L2 speakers are significantly different in speech rate ($t=7.222, p<0.001$). For English L1 speakers, the speech rate is averagely 3.11 words per second, much faster than L2 speakers, with an average speech rate of 1.80 words per second. The same result is found in articulation rate, with an average of 3.78 words per second in L1 speech and 2.54 words per second in L2 speech. Such temporal difference is also highly significant ($t=7.812, p<0.001$). The overall temporal fluency indicates that Chinese L2 speakers not only produced slower speech but also generate

more pauses or longer pause durations than English L1 speakers in on-line discourse speech.

In addition, based on standard deviations (especially the articulation rate), in-group variabilities of Chinese L2 speakers were much higher than that of English L1 speakers, which may be caused by different language proficiency.

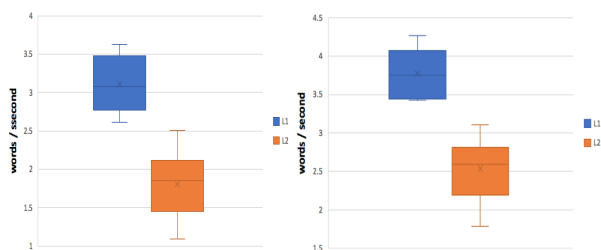


Figure 2: L1-L2 difference in temporal fluency (Left: speech rate, Right: articulation rate).

3.2. L1-L2 difference in macro-planning strategies

Chunk size is conducive to a detailed reflection of how speakers semantically organize their speech at the macro-planning level. In this section, speech chunk is defined by the number of words of each corresponding interpausal unit. All the chunks were further divided into three sub-groups according to the number of words it contained: small chunk (1-5 words per unit), medium chunk (6-15 words per unit) and large chunk (more than 16 words). The chunking preference will be closely studied between L1 and L2 speech.

3.2.1. Global patterns of chunk size

Figure 3 shows the different incidence of chunk size in L1 and L2 overall discourse speech. Statistically, a highly significant difference can be marked between L1 and L2 speech ($t=9.568$, $p<0.001$). The result shows that English L1 speakers were able to produce big chunks with the maximum of 26 words while Chinese L2 speakers only reached the maximum of 16 words. Moreover, Chinese L2 speakers frequently produced 2-word chunks, which may be the negative transfer from the mother language. In Chinese Mandarin, the most frequent prosodic chunk usually contains 2 words. With the increasing of chunk size, the incidence of the chunk declines both in L1 and L2 speech. Altogether, Chinese L2 speakers produced rather fewer words per chunk than English L1 speakers did, thus leading to a relatively smaller planning size of speech output.

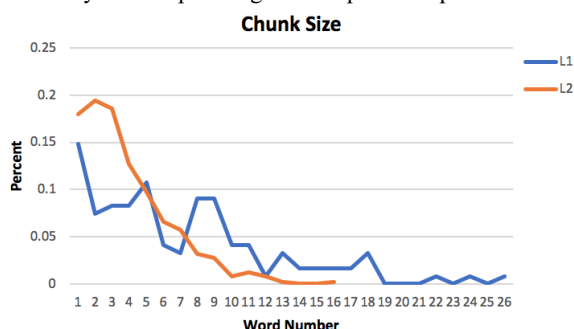


Figure 3: L1-L2 difference in global incidence of chunk sizes.

In addition, there is also a big gap in flexibility of macro-planning unit size between English L1 and Chinese L2 speakers. Specifically speaking, native speakers were able to vary more in chunk size than Chinese L2 learners did, especially in large chunks. For instance, L1 speakers can vary from 1 up to 26

words in a semantic chunk, while L2 speakers only from 1 to 16 words. This kind of distribution indicates that native speakers have a really high degree of freedom when they do on-line discourse planning at macro level due to their inherent linguistic knowledge. However, for Chinese L2 speakers, it is a big challenge.

3.2.2. Local patterns of chunking preference

The local distribution of various chunks will be further analyzed in this section. Figure 4, Figure 5 and Figure 6 show the L1-L2 difference in the distribution of small chunks, medium chunks and large chunks respectively. The detailed preference of certain chunk size under the three sub-groups can be also clearly observed. There are more conclusions than those English L1 speakers can produce more words within a semantic chunk.

In general, it can be seen that Chinese L2 speakers produced a lot of small chunks (78.44%), but rather fewer medium chunks (21.35%) and large chunks (0.21%), while on the contrary, English L1 speakers were more inclined to produce more medium chunks (41.32%) as well as large chunks (9.09%). This finding is inconsistent with the previous study which indicated that L2 English speakers used more intermediate chunking units [16].

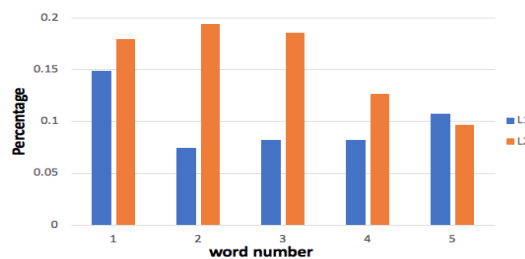


Figure 4: L1-L2 local incidence of small chunks.

Secondly, although L2 speakers used small-scale planning chunks more frequently than L1 speakers did at the macro-planning level, they preferred 1-word or 2-word chunks, rather than 5-word chunks as L1 speakers did (Figure 4). On the contrary, as shown in Figure 5, medium chunks are preferred by L1 speakers. Especially 8-word and 9-word Chunks frequently occurred in L1 speech. As for large-scale planning chunks, it is obvious that the incidence of large chunks in both groups drops, especially for the Chinese L2 speakers. What's more, L1 speakers preferred 18-word chunks as main large chunk size.

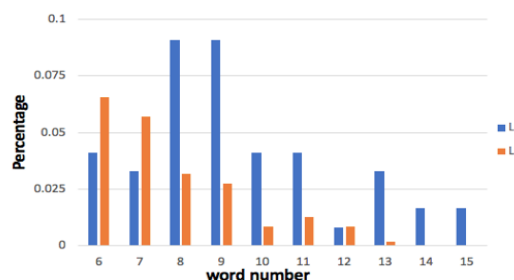


Figure 5: L1-L2 local incidence of medium chunks.

Taken the above results into consideration, for Chinese L2 learners, there is still a very long way to go in order to reach an almost-native-like competence. To be more specific, they should try to reduce the incidence of small chunks and reach the comparable proportion of medium chunks. In addition to that, a proper amount of large chunks is favored to reach a more

authentic attainment. Since language is a dynamic system, speaker groups encompass their own variation and Chinese L2 learners display an interminate linguistic competence.

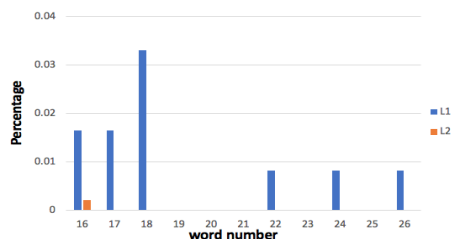


Figure 6: L1-L2 local incidence of large chunks.

3.3. L1-L2 differences in micro-planning strategies

Disfluency, as a character of speech revealing speech planning and production at the micro-level, is concentrated on in this section. First, different disfluency markers at the level of both discourse-level organization in L2 and L1 English speech were examined. Then the relationship between syntactic complexity and fluency were further analyzed.

3.3.1. Global patterns of disfluency markers at discourse level

According to the classification scheme described in Section 2.3, six disfluency markers were examined in this study. Figure 7 shows the overall incidence and distribution of the disfluency markers in English L1 and Chinese L2 on-line planning speech. Apparently, Chinese L2 speakers produced more hesitation markers (either silent or filled pauses), among which contains twice as many silent pauses as L1 speakers. This result is consistent with most previous studies (e.g. [24]). As for self-monitoring, different from L1 speakers, Chinese L2 speakers were intended to rely more on repetition and instant repair rather than anticipatory retracing and restart. This indicates that Chinese L2 and English L1 speakers differ categorically in repair strategy. For L2 speakers, repair occurs before the articulation, which means they have difficulties in the macroplanning for the discourse or lexical search; while L1 speakers think more about syntactic planning.

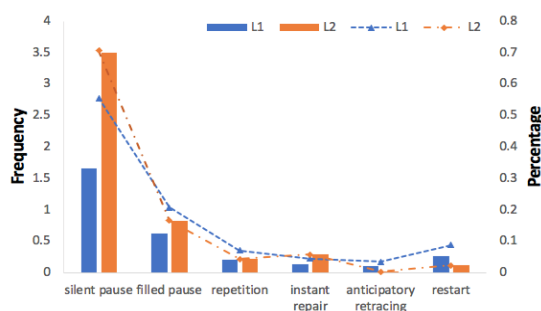


Figure 7: Incidence and distribution of the disfluency markers at discourse level

Statistical results suggest that the between-group difference in silent pauses is highly significant ($p < 0.001$), while instant repair, anticipatory retracing and restart is significant ($p < 0.05$). No significant difference was found in filled pause and repetition between L1 and L2 speakers.

3.3.2. Local disfluency attributes at utterance level

It is believed that there is a strong correlation between the occurrence of disfluency markers and syntactic complexity [21].

Therefore, in this section, we will focus on how the English L1 and Chinese L2 speakers differ from each other in the process of producing complex utterance.

From Figure 8, it can be seen that in the complex sentences, the planning strategies employed by the two groups showed statistical difference in silent pause and instant repair ($p < 0.05$) and highly significant difference in other disfluency makers ($p < 0.001$). This may reflect, to some extent, English L1 speakers can strategically take advantage of disfluency makers to make proper micro-planning so as to produce a more complex utterance in on-line planning discourse speech. But unfortunately, Chinese L2 speakers didn't seem to do such kind of strategic micro-planning. One reason to explain this phenomenon is that Chinese L2 learners are less proficient in English and they are more likely to be distracted by other disturbances in their mind like lexical retrieving, grammar checking, and syntactic construction etc.

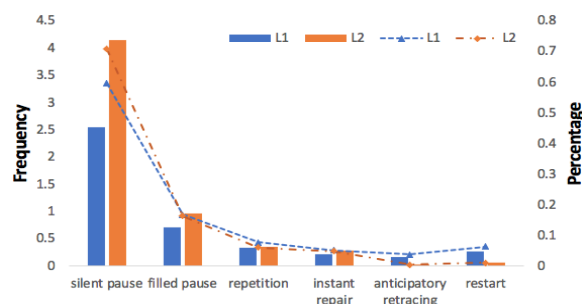


Figure 8: Incidence and distribution of the disfluency markers at complex utterances level

4. Conclusions

This paper mainly discusses the on-line planning strategies by observing chunking preference and disfluency attributes utilized by Chinese L2 learners in English discourse speech. The main purpose is to reveal the discourse-level speech planning shaped by the richness and density of the moment-by-moment planned message in Chinese L2 English.

Some interesting findings are illustrated here: 1) as a global property of the spoken discourse as a whole, Chinese L2 learners were temporally more disfluent than L1 speakers, in terms of speaking rate and articulation rate, meanwhile with a greater in-group variabilities due to the insufficient L2 proficiency; 2) as for the semantic macro-planning strategies, Chinese L2 learners usually generated rather smaller chunks (especially 2-word chunks) and less flexible chunk size than L1 speakers in on-line discourse speech, thus leading to a smaller planning size of speech output and a stable pattern of non-native-like ultimate attainment; 3) at the micro-planning level, compared with L1 speakers, Chinese L2 learners produced more hesitation markers (especially silent pauses), differed categorically in repair strategy and failed to strategically take advantage of certain disfluency markers to buy plan time so as to increase the complexity of utterance in on-line discourse speech.

5. Acknowledgements

This study was funded by the National Social Science Foundation of China, Project 15BY183.

6. References

- [1] W. J. M. Levelt, *Speaking: From Intention to Articulation*. Cambridge, Massachusetts: MIT Press, 1989.
- [2] K. De Bot, "A Bilingual Production Model: Levelt's 'Speaking' Model Adapted," *Appl. Linguist.*, vol. 13, no. 1, pp. 1–24, 1992.
- [3] N. Segalowitz, *Cognitive bases of second language fluency*. Routledge, 2010.
- [4] B. Vanpatten, "Attending to form and content in the input: An experiment in consciousness," *Stud. Second Lang. Acquis.*, vol. 12, no. 3, pp. 287–301, 1990.
- [5] H. H. Clark and J. E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, 2002.
- [6] F. Yuan and R. Ellis, "The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production," *Appl. Linguist.*, vol. 24, no. 1, pp. 1–27, 2003.
- [7] J. N. Wendel, *Planning and second language narrative production*. Temple University, 1997.
- [8] P. Skehan, "Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis," *Appl. Linguist.*, vol. 30, no. 4, pp. 510–532, 2009.
- [9] P. Foster and P. Skehan, "The influence of planning and task type on second language performance," *Stud. Second Lang. Acquis.*, vol. 18, no. 3, pp. 299–323, 1996.
- [10] R. Ellis, *Planning and task performance in a second language*. John Benjamins Publishing, 2005.
- [11] B. Butterworth, "Some constraints on models of language production," *Lang. Prod.*, vol. 1, pp. 423–459, 1980.
- [12] T. M. Derwing, M. J. Rossiter, M. J. Munro, and R. I. Thomson, "Second language fluency: Judgments on different tasks," *Lang. Learn.*, vol. 54, no. 4, pp. 655–679, 2004.
- [13] J. Bernstein, J. Cheng, and M. Suzuki, "Fluency changes with general progress in L2 proficiency," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2011, no. August, pp. 877–880.
- [14] S. J. Owens and S. A. Graham, "Thee, uhh disfluency effect in preschoolers: A cue to discourse status," *Br. J. Dev. Psychol.*, vol. 34, no. 3, pp. 388–401, 2016.
- [15] W. C. Mann and S. A. Thompson, "Rhetorical Structure Theory: Toward a functional theory of text organization," *Text*, vol. 8, no. 3, pp. 243–281, 1988.
- [16] C. Y. Tseng and C. Y. Su, "Where and how to make an emphasis?-L2 distinct prosody and why," in *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing, ISCSLP*, 2014, pp. 633–637.
- [17] M. Watanabe, K. Hirose, Y. Den, and N. Minematsu, "Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners," *Speech Commun.*, vol. 50, no. 2, pp. 81–94, 2008.
- [18] B. Bigi and D. Hirst, "SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody," in *Proceedings of Speech Prosody*, 2012, pp. 1–4.
- [19] L. García Amaya, "New findings of fluency measures across three different learning contexts," *Sel. Proc. 11th Hisp. Linguist. Symp.*, pp. 68–80, 2009.
- [20] A. E. Hieke, S. Kowal, and D. C. O'Connell, "The trouble with 'articulatory' pauses," *Lang. Speech*, vol. 26, no. 3, pp. 203–214, 1983.
- [21] F. Chambers, "What do we mean by fluency?," *System*, vol. 25, no. 4, pp. 535–544, 1997.
- [22] K. W. Hunt, *Grammatical structures written at three grade levels*. Champaign, Illinois: National Council of Teachers of English, 1965.
- [23] R. Towell, R. Hawkins, and N. Bazergui, "The development of fluency in advanced learners of French," *Appl. Linguist.*, vol. 17, no. 1, pp. 84–119, 1996.
- [24] J. Yu, J. Chen, S. Wu, and Y. Feng, "Disfluency in Chinese L2 spontaneous speech: Patterns and interactions," *Proc. Int. Conf. Speech Prosody*, vol. 2018-June, pp. 863–867, 2018.