Disfluency in Chinese L2 Spontaneous Speech: Patterns and Interactions

Jue Yu¹, Jiena Chen¹, Shengyi Wu², Ye Feng³

¹School of Foreign Languages, Tongji University, Shanghai, China
²Department of Chinese, Fudan University, Shanghai, China
³Department of Pediatric Rehabilitation, Shanghai Yangzhi Rehabilitation Hospital, China
erinyu@126.com, cjn0529@tongji.edu.com, wushengyi@yahoo.com, Fengye0509@163.com

Abstract
This paper mainly discusses the notion of disfluency within the context of language planning and production. The main purpose is to examine disfluency performance in Chinese L2 English spontaneous speech and figure out how different disfluency factors intertwine with each other, thus revealing the underlying processes and preferred strategies of language planning and repair utilized by Chinese L2 speakers. The results show that compared with English natives, Chinese L2 speakers are considerably more disfluent, in terms of time-related and performance-related aspects, have different preferences for particular disfluency markers in language planning and self-monitoring, and fail to strategically take advantage of pausing behavior to make a success of self-monitoring.

Index Terms: disfluency, language planning, Chinese L2 speakers, English L1 speakers, English spontaneous speech

1. Introduction
Second language production is of great complexity, involving linguistic and non-linguistic factors. The most prominent deviator of speech production in the second language learning environment is the fluency of speech. Actually all spontaneous speech production is a highly fragmented and discontinuous activity, characterized by frequent occurrences of disfluencies. However, for L2 speakers, the challenge seems to be greater [1].

Although early studies negatively regarded disfluency as a speech error or an indicator of poor proficiency [2] [3] [4], more and more researchers have viewed it as an informative ingredient in speech production [5] [6] [7]. To some extent, disfluency is a truthful reflection of speaker’s difficulty and the urgency of organizing a relevant linguistic constituent. Moreover, it can bring prosodic cues to indicate new messages and help speakers to buy some time for planning the upcoming utterance [8] [9] [10]. It is believed that fluent speech depends on the availability of well-established linguistic knowledge, routines of speech planning and articulation. A lack of speech fluency in L2 speech may point to a deficiency of these representations, due to incomplete acquisition, L1 transfer and partial language competition [11].

Recently, more attention has been devoted to comparative investigations of L1 and L2 disfluency performance. For example, studies show that L2 speakers are considerably more disfluent, make more self-corrections [12], produce two to three times as many hesitation markers as L1 speech of the same languages [13], differ in pausing behavior [14], produce shorter phrase lengths and slower speech rates [15]. Taken together, these results only suggest that on different measures, L2 speakers do not reach a native-like level of speech fluency. Few of them handle the specific distributions of these disfluency markers, let alone their functions and interactions in the L2 discourse speech plan.

This paper will use topic-driven spontaneous speech rather than read out or separated sentences, as speech material and will acoustically measure the utterance fluency. The main purpose is to investigate the prosodic characteristics and contributions of these disfluency markers to language planning and repair as well as their internal interactions in Chinese L2 English speech. Three research questions are proposed:
1. What are the differences in disfluency performance between English L1 and Chinese L2 speakers, in terms of time-related and performance-related factors?
2. What are the differences in the underlying processes and strategies of language planning and repair utilized by English L1 and Chinese L2 speakers? Are there any preferences for particular disfluency markers?
3. How do these disfluency factors interact with each other in English L1 and Chinese L2 speech?

2. Method

2.1. Data
Topic-driven spontaneous speech was used in this study. In order to facilitate speech production while ensuring spontaneity, subjects were instructed to give an English speech based on the topic “What Do You Do in Your Spare Time”. First, this topic is frequently used in TOEFL oral test, thus familiar to both native speakers and Chinese EFL learners; second, it is closely related to everyone’s daily life and subjects are able to produce connected speech by telling their own personal experiences.

Before the recording started, the subjects were given 5 minutes to prepare for a 1-minute speech and recordings were made after they were ready. To ensure spontaneity and naturalness, subjects were not interrupted for disfluency 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 6.6 utterances among Chinese L2 speakers to 7.8 utterances among English L1 speakers.

2.2. Subjects
Altogether, 20 Chinese L2 speakers (10 male and 10 female) were recruited and 10 American English speakers (5 male and 5 female) were used as a baseline. All speakers are at their twenties. Chinese L2 speakers are all college students, with different discipline background (e.g. arts, science, and engineering) but from the same dialect region—Shaoxing. Therefore, the influence of dialects and learning environments
can be minimized. Moreover, they all 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 [16], and then manually post-edited. For the disfluency labelling, all work was done manually by trained transcribers and double checked by the authors.

To categorize disfluency, the classification scheme described in [17] was used. Therefore, disfluency was labelled in terms of hesitation markers and self-monitoring markers (see Fig.1). According to [18] and the research aims of this paper, 100ms was chosen as the cut-off point of silent pause.

<table>
<thead>
<tr>
<th>Four Types of Self-monitoring</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Silent pause/silent pause: It refers to phonetically empty interruptions.</td>
<td></td>
</tr>
<tr>
<td>2. Filled pause: It refers to interruptions with sounds like “uh” or “ah”.</td>
<td></td>
</tr>
<tr>
<td>3. 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 my life.” (’W’ is coded as the location of silent pause).</td>
<td></td>
</tr>
<tr>
<td>4. Anticipatory retraction: Anticipatory retraction describes a situation where the speaker goes back to what he or she said before or something else. E.g.: “...reading some books about some books about the Internet...” (’W’ is coded as the location of silent pause).</td>
<td></td>
</tr>
<tr>
<td>5. 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 to have less space time.” (’W’ is coded as the location of silent pause).</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1: Classification of disfluency markers.](image)

2.4. Data analysis

For this paper, fluency is typically defined by some key concepts that are time-related and performance-related. Temporal variables of fluency were measured by the speech/pause relationship [19]. As for performance aspects, disfluency markers were first categorized into two major groups: hesitation markers and self-monitoring markers, and then the prosodic features of these markers in conjuction with their location and frequency were analyzed. Thus, the differences in the manner of discourse planning and repair between Chinese L2 and English L1 speakers can be observed. Besides, this study further provided some information on the basic grammar construction strategies of Chinese L2 speakers by measuring word count per T-unit (According to [20], a T-unit is a dominant clause and its dependent clauses). Finally, the correlations between all these disfluency factors were examined. So the interplay between individual factors can be observed.

In order to exclude the influence of speaker and speech rate, Z-score normalization was conducted for all the duration parameters.

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. Speech rates of English L1 and Chinese L2 speakers differ a bit. For most of the L2 speakers, the speech rate is 3–4 syllables per second, much lower than L1 speakers, with above 4 syllables per second. This overall difference is highly significant (F=5.367, P<0.01), which indicates English L1 speakers were considerably faster than Chinese L2 speakers in spontaneous speech. This is highly consistent with many previous studies [17] [21] [22].

Phonation time (or articulation time) is calculated by silent pause time divided by the total speech time. The results show that Chinese L2 speakers produced more pause time than English L1 speakers in discourse speech (English L1: mean ratio=0.15; Chinese L2: mean ratio=0.25), which is extremely remarkable in statistical analysis (F=6.494, P<0.01). It indicates that either more pauses or longer pause durations are involved in L2 discourse speech.

Moreover, based on standard deviations, in-group variabilities of Chinese L2 speakers were much higher than that of English L1 speakers, in terms of both speech rate and phonation time, which may be caused by the different language proficiency.

3.2. L1-L2 differences in performance-related fluency

Fluency should not be reduced to the speed of delivery [19]. It is equally important to consider other performance aspects. For second language speakers, processing tasks are much greater and therefore increase the cognitive load, resulting in good chances of producing disfluency markers. Furthermore, limitations in the learners’ second language proficiency causes patterns of disfluency markers different from those of native speakers [23]. In this part, patterns of disfluency markers were compared between Chinese L2 and English L1 speakers both at discourse level and sentence level; thus the underlying processes or strategies of language planning and repair can be revealed.

3.2.1. Global patterns of disfluency markers at discourse level

According to Levelt’s speech model [24], disfluencies can be solved through a process of self-monitoring, which involves repetitions, instant repair, restarts and anticipatory retracting, or the use of hesitation phenomena: silent pauses or filled pauses. Altogether, six disfluency markers were examined in this study.

Figure 2 shows the overall incidence of hesitation and self-monitoring markers in English L1 and Chinese L2 speech. Obviously, Chinese L2 speakers produce more disfluency markers, among which contains twice as many self-monitoring markers as L1 speakers (some maybe triple, e.g. repetitions). These results are consistent with previous studies. It implies that incomplete acquisition is often associated with increased disfluency. If errors are caught before articulation, more pauses and repetitions should be found, whereas errors that are only noticed after articulation would result in more instant repair, restart and anticipatory retracting.

Moreover, although both English L1 and Chinese L2 speakers prefer to use silent pause as a major manner of language planning, they have different preferences for self-monitoring. Compared with L1 speakers, L2 speakers list repetition rather than restart as a second major disfluency marker. This indicates 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. In addition, the other three disfluency markers are comparatively less utilized by either L1 or L2 speakers.
In order to reveal the differences in linguistic plans between English L1 and Chinese L2 speakers, pauses were further studied in terms of duration and distribution (see Fig.3). Table 1 displays the normalized durations and distributions of both silent and filled pauses. Generally, L2 speakers produce statistically longer pauses than L1 speakers, especially filled pauses \( (F=6.729, P<0.01) \). The possible reason is that Chinese L2 learners prefer timing mechanism to do linguistic processing and planning, due to incomplete acquisition.

![Figure 2: L1-L2 global incidence of disfluency markers.](image)

### Figure 3: Classification of pause distributions.

A speaker may stop for planning at grammatical junctures, constituent boundaries and before the first content word within a constituent [25]. Consequently, a number of disfluency markers may appear at these planning points. The L1-L2 distribution results show that most pauses are located at clause boundaries. However, compared with L1 speakers, Chinese L2 speakers generate more ungrammatical pauses in phrase-medial or word-medial position (see Table 1).

| Table 1: L1-L2 global pause durations and distributions. |
|-----------------|------------------|
| **Pause Location** | **L1** | **L2** |
| silent pause | 0.305 | 0.377 |
| filled pause | -0.755 | -0.575 |
| **Pause Duration** | | |
| inter-clause | 67.58% | 47.43% |
| inter-phrase | 32.42% | 45.06% |
| inter-word | 0 | 4.35% |
| intra-word | 0 | 3.16% |

According to [24], speech production includes macroplanning that generates preverbal messages and micro-planning which includes grammatical encoding and phonological encoding, articulation and self-monitoring. The distribution of pauses reveals that English L1 and Chinese L2 speakers apply different planning strategies when handling discourse speech. Taking the other disfluency markers into consideration, it can be concluded that more microplanning time is needed by L2 speakers and more macroplanning time by L1 speakers. This can partially explain why L1 speakers are more likely to produce larger chunks than L2 speakers do.

### 3.2.2. Local patterns of disfluency markers at sentence level

It is proposed that the underlying processing difficulties may be associated with the degree of syntactic complexity of the utterance and higher cognitive load necessary for processing it [26]. This section will look at the relationship between the occurrence of disfluency markers and more complex sentences.

First, the word count per T-unit was investigated as an index of sentence complexity [20]. Figure 4 shows that L1 speakers can produce less than 10 words per T-unit to more than 35 words per T-unit, which suggests greater variability in sentence length comparing to L2 speakers. Moreover, although the most frequent sentence length is around 14 words per T-unit in both L1 and L2 speech, L1 speakers can also produce sentence length of 28 words per T-unit frequently. This indicates L1 speakers can produce comparatively more complex and longer sentences than L2 speakers.

![Figure 4: Distribution of the word count per T-unit (Left: L1 speech, Right: L2 speech).](image)

A closer analysis of disfluency markers per T-unit in Figure 5 shows that all speakers prefer to use silent pause as the basic grammar construction strategy (especially Chinese L2 speakers). However, there appears to exist a preference for a particular disfluency marker among L1 and L2 speakers. Most L1 speakers prefer to resort to a greater use of filled pauses to facilitate lexical and syntactic planning [24], while L2 speakers prefer to use repetition to buy time for lexical search rather than carefully plan their utterance internally. Furthermore, L2 speakers do not seem to monitor their speech so often within a sentence. This may be partially because the pressure for showing fluency forces speakers to keep up the pace of their speech and overlook their mistakes.

![Figure 5: L1-L2 local incidence of disfluency markers.](image)
complexity. What’s more, L1 speakers will involve more silent pause and instant repair with the increasing of sentence complexity, while L2 speakers use more filled pause, repetition, anticipatory retracing and restart. This indicates L2 speakers have more difficulties in syntactic and semantic planning.

Table 2: Correlations between sentence complexity and disfluency markers.

<table>
<thead>
<tr>
<th></th>
<th>L1 T-unit Complexity</th>
<th>L2 T-unit Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silent pause</td>
<td>.748**</td>
<td>.633**</td>
</tr>
<tr>
<td>Filled pause</td>
<td>.401**</td>
<td>.468**</td>
</tr>
<tr>
<td>Repetition</td>
<td>.553**</td>
<td>.638**</td>
</tr>
<tr>
<td>Instant repair</td>
<td>.318**</td>
<td>.071</td>
</tr>
<tr>
<td>Anticipatory retracing</td>
<td>.085</td>
<td>.339**</td>
</tr>
<tr>
<td>Restart</td>
<td>.349**</td>
<td>.381**</td>
</tr>
</tbody>
</table>

** significant at the .01 level

3.3. L1-L2 differences in interactions between disfluency factors

This section will look at the interactions between time-related and performance-related factors. Meanwhile, it tries to specify how individual factors intertwine with each other in English L1 and Chinese L2 spontaneous speech respectively.

Figure 6 shows the correlation results in English L1 and Chinese L2 spontaneous speech. First, time-related disfluency factors are highly correlated with most hesitation markers but few self-monitoring ones (except anticipatory retracing) in L1 speech. For example, speech rate is highly correlated with filled pause frequency (FPF) and its duration (FPD), while phonation time is correlated with silent pause. This means different types of pause have a specific impact on temporal fluency in L1 speech. However, only silent pause duration (SPD) has an impact on temporal fluency in L2 speech, which evidences that it is not more but longer pauses that cause the slow L2 speech.

The interplay between individual hesitation markers and self-monitoring markers was further examined in Figure 7. The results show that for most of L1 speakers, silent pauses are correlated with repetition and restart, while filled pauses correlated with the rest two. However, there is lack of systematic strong correlation between hesitation markers and self-monitoring ones in L2 speech. This suggests Chinese L2 speakers strategically fail to take advantage of hesitation markers to make a success of self-monitoring.

4. Conclusion

This paper mainly discusses the notion of disfluency within the context of language planning and production. The main purpose is to examine disfluency performance in Chinese L2 English speech and figure out how different disfluency factors intertwine with each other, thus revealing the underlying processes and preferred strategies of language planning and repair utilized by Chinese L2 speakers. Due to the limited space, gender differences are not reported in this paper.

Some interesting findings are illustrated here: 1) Compared with English natives, Chinese L2 speakers are considerably more disfluent, in terms of time-related and performance-related aspects. 2) Chinese L2 speakers occupy more timing mechanism to do linguistic processing and planning, due to incomplete acquisition. 3) Although both English L1 and Chinese L2 speakers prefer to use silent pause as a major manner of language planning, they have different preferences for self-monitoring. 4) For syntactic planning, most English L1 speakers prefer to use filled pause to facilitate lexical and syntactic planning, while Chinese L2 speakers prefer to use repetition for lexical search. Moreover, with the increasing of sentence complexity, L1 speakers will involve more silent pause and instant repair, while L2 speakers use more filled pause, repetition, anticipatory retracing and restart. 5) Different types of pause have a specific impact on temporal fluency in English L1 speech but not in Chinese L2; what’s more, there is lack of systematic strong correlation between hesitation markers and self-monitoring ones in L2 speech.

5. Acknowledgements

The first author is sponsored by the National Social Science Foundations of China (Grant No. 15BYY183) and the Interdisciplinary Program of Tongji University (1100219144) for this research work. We are very grateful to Professor Dafydd Gibbon at Bielefeld University for his support in proofreading.
6. References


