Using Corpora to Study L2 Acquisition of English Tense of Chinese non-English-major university students — A Study Based on Chinese Learner English Corpus

He Qiong
School of Foreign Language and Literature
Wuhan University

Abstract

Numerous studies on L2 acquisition of English tense have been conducted home and abroad over the past 30 years, however, since corpus-based study on the acquisition of tense of Chinese college English students is still lacking, the hypotheses drawn upon previous studies haven’t been testified with large quantities of written texts of Chinese college English students. The present paper seeks to testify the hypotheses on the basis of Chinese Learner English Corpus (CLEC) and then sketch out the implication of the study for College English Education.

Key words: CLEC, L2 Acquisition of tense, Chinese college English students

Introduction

There are more than ten tense types in English, according to College English Syllabus (1999:144, Chinese non-English-major university students should master as many as 11 English tense types, i.e. simple present, simple past, present perfect, present continuous, present perfect progressive, past perfect, past continuous, past perfect progressive, simple future, past future and future perfect. Moreover, English tense usage is very complicated, and the choice of a particular type of tense is always determined by the context, no wonder Leech (1987) identifies tense and aspect as two difficult areas in the L2 acquisition of English. Over the past 30 years, Numerous studies on L2 acquisition of English tense have been conducted home and abroad (Bronckart & Sinclair 1973, Guiora 1983, Robinson 1990, Ramsay 1990, Hinkel 1992, Shirai 1995,Fan&Lin 2002), however, the corpus-based study on L2 acquisition of English tense of Chinese non-English-major university students is still lacking. Therefore, the present paper aims to make an initial effort in using corpora to study L2 acquisition of English tense of Chinese non-English-major university students, so as to bring new insights into the inter language tense usage of Chinese non-English-major university students, hence giving some implications to Chinese College English Education.
Research Questions of the Study
Research Question 1: Is tense difficult for Chinese non-English-major university students to acquire?
Research Question 2: Which tense types are error-prone or improvement-resistant to Chinese non-English-major university students?
Research Question 3: Which “erroneous tense-target tense” pair is most likely to be confused by Chinese non-English-major university students?

Related Hypotheses
Tense is a very complex grammatical phenomenon in English, while in Chinese language, there’s no tense but aspect (cf. Li & Thompson 1981: 184-237). As Coppierers (1987: 544-573) points out, the understanding of L2 tense usage is greatly affected by L1 temporal structure and tense usage, so it’s not difficult for us to infer that the discrepancy in tense between English and Chinese may affect L2 acquisition of English tense of Chinese English learners, and L2 acquisition of English tense is very likely to be the difficult area for Chinese English learners. The difficulties, as Jacobs (cf.1995:194) puts it, lies in the complicated relationship between English tense and temporal references as well as the gap between English and Chinese tense system. According to the study of Hinkel (1992), there is a significant difference in the understanding of Present Perfect between Chinese English learners and native speakers. And in Lardiere’s (1998:1) 8-year investigation of Patty, an adult native Chinese speaker who has lived in the U.S. for 18 years, the results show the fossilization of English L2 past tense morphology at a consistently very low rate of suppliance (approximately 34%) in obligatory contexts, which to some extent proves that English tense is really very difficult to Chinese English learners, even those on near-native level. Fan & Lin (2002: 414-420) studied 87 EFL and ESL Chinese students with various levels, ranging from beginning to advanced levels, only to find that Chinese students tend to associate Present Perfect with time and place in the past. Although the tendency diminishes as the English proficiency level increases, it still exists among advanced learners. Based on previous studies, the following hypotheses can be made in line with the above three research questions,
Hypothesis 1: Tenses are difficult to Chinese English learners, including Chinese non-English-major university students.
Hypothesis 2: Simple past and present perfect are error-prone to Chinese non-English-major university students.
Hypothesis 3: “Simple past-present perfect” are among the “erroneous tense-target tense” pair that are most likely to be confused by Chinese non-English-major university students.

However, since corpus-based study on the L2 acquisition of English tense of Chinese non-English-major university students is still lacking, these hypotheses are still waiting to be testified with large quantities of written texts of Chinese non-English-major university students.
Corpus and Corpus tool in this Study

The present paper seeks to testify the above hypotheses on the basis of Chinese Learner English Corpus (CLEC). CLEC contains 1 million words of English compositions collected from Chinese learners of English with differing levels of proficiency, covering senior secondary school students, English-major, and non-English-major university students in China.

<table>
<thead>
<tr>
<th>Student Type</th>
<th>Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST2 senior secondary school students</td>
<td>208088</td>
</tr>
<tr>
<td>ST3 band 4 Non English-major university students</td>
<td>209043</td>
</tr>
<tr>
<td>ST4 band 6 non English-major university students</td>
<td>212855</td>
</tr>
<tr>
<td>ST5 Junior English-major university students</td>
<td>214510</td>
</tr>
<tr>
<td>ST6 Senior English-major university students</td>
<td>226106</td>
</tr>
<tr>
<td>Total</td>
<td>1070602</td>
</tr>
</tbody>
</table>

(http://www.clal.org.cn/baseinfo/achievement/Achievement1.htm)

The corpus is error tagged according to a hierarchical error marking scheme of 11 categories and 50 subcategories of errors, including various lexical, grammatical, semantic and sentence level errors. Moreover, the error marking scheme is flexible, subcategories can be added or deleted or even further subcategorized according to different research purposes. Once inserted into the learner corpus, the error tags can be searched like any other elements in the corpus with concordancing programs, and comprehensive lists of specific error types can be drawn up. For this investigation, I first manually subcategorized all verb tense errors (VP6) into 8 kinds of frequent verb tense errors from two error tagged learner sub-corpora of argumentative essay writing in CLEC. The first corpus contains writings by College English students in the Band 4 exam, and the second, writings by College English students in the Band 6 exam. The 8 subcategories of verb tense errors (VP6) are as follow, VP61 (Simple Present), VP62 (Simple Past), VP63 (Present perfect), VP64 (Present Continuous), VP65 (Past Perfect), VP66 (Past Continuous), VP67 (Simple Future), VP68 (Past Future). Then with a corpus tool called Wordsmith 3.0, I respectively retrieve these verb tense errors from CLEC. The first sub-corpus contains writing by band 4 Non English-major university students and the second, writing by band 6 Non English-major university students. I will refer to the two sub-corpora as Band 4 corpus and Band 6 corpus, which can be roughly regarded as intermediate and advanced respectively. Though strictly speaking, the study is not longitudinal since it involves two different groups of learners, it
seems legitimate to make inferences in terms of progress as the two groups have followed the same building principles in the same program of study. However, as the two corpora are of different size, they can’t be compared until the frequencies provided by Wordsmith 3.0 is standardised. In this study, the standardized frequency of the sub-category of VP6 can be obtained by using the formula as follows,

\[
\text{Raw frequency/size of the sub-corpus} \times 100000
\]

Based on the standardized frequency information, progress rate can be calculated, thereby can be examined whether or not the findings of this study are consistent with the above hypotheses about L2 acquisition of English tense of Chinese non-English-major university students

**Erroneous tenses**

VP6 is a subcategory of the VP category in CLEC, which includes all grammatical verb errors. The other subcategories are VP1(pattern), VP2(set phrase), VP3(articulation), VP4(finite), VP5(non-finite), VP7(voice), VP8(mode) and VP9(modal/auxiliary). Table I gives the breakdown of the VP category in CLEC.

<table>
<thead>
<tr>
<th>Category</th>
<th>C</th>
<th>ST3</th>
<th>ST</th>
<th>Progress rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>1387.9</td>
<td>325.9</td>
<td>498.4</td>
<td>-53%</td>
</tr>
<tr>
<td>VP2</td>
<td>505.8</td>
<td>139.3</td>
<td>61.2</td>
<td>56%</td>
</tr>
<tr>
<td>VP3</td>
<td>2283.9</td>
<td>524.6</td>
<td>785.2</td>
<td>-49.7%</td>
</tr>
<tr>
<td>VP4</td>
<td>526.2</td>
<td>159.1</td>
<td>110.8</td>
<td>30.4%</td>
</tr>
<tr>
<td>VP5</td>
<td>502.7</td>
<td>118.7</td>
<td>107.4</td>
<td>9.5%</td>
</tr>
<tr>
<td>VP6</td>
<td>2428.7</td>
<td>356</td>
<td>311.6</td>
<td>12.5%</td>
</tr>
<tr>
<td>VP7</td>
<td>485.8</td>
<td>104.1</td>
<td>98.4</td>
<td>5.5%</td>
</tr>
<tr>
<td>VP8</td>
<td>88.4</td>
<td>16.3</td>
<td>8.3</td>
<td>49.1%</td>
</tr>
<tr>
<td>VP9</td>
<td>793.2</td>
<td>274.3</td>
<td>278.5</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

The VP6 (tense error) category is the largest subcategory of verb errors in the whole one-million-word CLEC, accounting for 27% of the VP errors. A comparison of the Band 4 and Band 6 sub-corpora shows that the tense errors displays the sixth lowest progress rate: an improvement of 12.5% against the more encouraging 56% for the VP2(set phrase) category. This result confirms hypotheses 1, i.e. tenses are difficult to Chinese English learners, including Chinese non-English-major university students.
The Most-erroneous Tense

Table 2 gives the breakdown of overall standardized frequencies of erroneous tenses in both sub-corpora of band 4 and band 6 Non English-major university students. The following categories are distinguished: SP (simple present), SA (simple past), P (present perfect), P (present continuous), P (past perfect), and P (past continuous).

Table 2 Breakdown of overall standardized frequencies of erroneous tenses in both sub-corpora of band 4 and band 6 Non English-major university students

<table>
<thead>
<tr>
<th>Tense category</th>
<th>N. of tense errors</th>
<th>% of tense errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP62(SP)</td>
<td>101</td>
<td>49%</td>
</tr>
<tr>
<td>VP61(SPr)</td>
<td>70</td>
<td>33.4%</td>
</tr>
<tr>
<td>VP65(PaPerf)</td>
<td>7</td>
<td>3.15%</td>
</tr>
<tr>
<td>VP63(Pperf)</td>
<td>5</td>
<td>2.55%</td>
</tr>
<tr>
<td>VP64(PrC)</td>
<td>4.5</td>
<td>2.18%</td>
</tr>
<tr>
<td>VP67(SF)</td>
<td>4.4</td>
<td>2.1%</td>
</tr>
<tr>
<td>VP68(PaF)</td>
<td>3</td>
<td>1.45%</td>
</tr>
<tr>
<td>VP66(PaC)</td>
<td>1.5</td>
<td>0.73%</td>
</tr>
</tbody>
</table>

The two tenses which provoked the most errors were the simple present and the simple past, although this result may in fact simply be due to the higher overall frequency of these tenses in English (Kennedy 1998). The standardized frequencies of erroneous tenses in both sub-corpora of band 4 and band 6 Non English-major university students as presented in Table 2 may therefore not prove very revealing. However, a comparison of the frequencies in the two sub-corpora brings interesting insights into learners' progress. As the standardized frequencies of VP64(PrC), VP67(SF), VP68(PaF) and VP66(PaC) are less than 5, which are too low and therefore may make the result of this study invalid, so I ignore them when making the comparison. In Table 3, the erroneous tenses are listed in decreasing order of progress rate. This table indicates that as students progress, they make fewer mistakes with some tenses, such as VP61 and VP63 (with the nearly 52.03% improvement for VP61(SPr) and the 41.3% improvement rate for VP63 VP63(Pperf), while others, for instance, VP62(SP) and VP65(PaPerf), are most improvement-resistant (with the nearly -25.4% improvement for VP62 and the -80.2% improvement rate for vp65 the past Perfect). From this table, it’s not difficult for us to see that simple past and present perfect are among the erroneous tenses, which testifies hypothesis 2, i.e. simple past and present perfect are error-prone to Chinese non-English-major university students.
Table 3 Breakdown of standardized frequencies and progress rate of tense errors in ST3 and ST4

<table>
<thead>
<tr>
<th>Tense</th>
<th>ST3</th>
<th>ST4</th>
<th>Progress rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP61SPr</td>
<td>91</td>
<td>43</td>
<td>52.03%</td>
</tr>
<tr>
<td>Vp63PPerf</td>
<td>6</td>
<td>4</td>
<td>41.3%</td>
</tr>
<tr>
<td>VP62Spa</td>
<td>86</td>
<td>108</td>
<td>-25.4%</td>
</tr>
<tr>
<td>VP65PaPerf</td>
<td>4</td>
<td>8</td>
<td>-80.2%</td>
</tr>
</tbody>
</table>

Target tense-erroneous tense pair

For each tense error, the sub-corpora of band 4 and band 6 Non English-major university students in CLEC contain the erroneous tense, i.e. the wrongly selected tense, as well as the target tense, i.e. the tense that should have been selected. The following figure shows that there are significantly different proportions of target use / erroneous use across the range of tenses. A detailed analysis of the ‘target tense / erroneous tense’ pairs brings out the tenses that are most likely to make Chinese-speaking learners confused. The most error-prone tense pairs in both sub-corpora of band 4 and band 6 Non English-major university students are listed below in decreasing order of frequency. The figures in brackets indicate the number of instances, and the tense with a little star on the upper right hand represents erroneous tense.

Figure 1 the breakdown of the overall frequencies of Target vs. erroneous use in both sub-corpora of band 4 and band 6 Non English-major university students.

1. **Simple Present t-Simple Past** (527)
   - SPr* → SPa (343)
   - SPA* → SPr (184)

2. **Simple Past t- Present Perfect** (49)
   - SPA* → PPerf (7)
   - PPerf* → SPA (42)

3. **Simple Present t-Present Perfect** (33)
   - SPr* → PPerf (10)
   - PPerf* → SPr (23)

4. **Simple Present t-Simple Future** (23)
   - SPr* → SF (6)
   - SF* → SPr (17)
5. **Past Future t-Simple Future* (14)**

\[
\begin{array}{ccc}
\text{PaF*} & \text{SF(5)} \\
\text{SF*} & \text{PaF (9)}
\end{array}
\]

6. **Present Continuous t- Simple Present* (11)**

\[
\begin{array}{ccc}
\text{PrC*} & \text{SPr (1)} \\
\text{SPr*} & \text{PrC (10)}
\end{array}
\]

7. **Simple Present t- Past Perfect* (7)**

\[
\begin{array}{ccc}
\text{SPr*} & \text{PaPerf (4)} \\
\text{PaPerf*} & \text{SPr (3)}
\end{array}
\]

8. **Simple Future t-Simple Past* (7)**

\[
\begin{array}{ccc}
\text{SPa*} & \text{SF(2)} \\
\text{SF*} & \text{SPa (5)}
\end{array}
\]

9. **Simple Past t- Present Continuous* （2）**

\[
\begin{array}{ccc}
\text{PaC*} & \text{PrC (2)} \\
\text{PrC*} & \text{PaC (0)}
\end{array}
\]

The figures show that Chinese learners are much more likely to be confused about simple past and simple present, however, as discussed above, this result may in fact simply be due to the higher overall frequency of these tenses in English (Kennedy 1998). Therefore, it’s more reasonable to regard “Simple past- present perfect” as the “erroneous tense-target tense” pair that is actually most likely to be confused by Chinese non-English-major university students. Thus is testified Hypothesis 3 i.e. “Simple past- present perfect” are among the “erroneous tense-target tense” pair that are most likely to be confused by Chinese non-English-major university students.

The findings of this quantitative study are quite fruitful, because they identify the tenses and tense pairs that still ought to be emphasized while teaching Chinese-speaking EFL learners at tertiary level. In the following sections, I will probe into the source of the main tense problem found in this study, then sketch out the implications of the study for College English Education in China.
3. Source

As stated by McCarthy (1991: 62) “tense and aspect vary notoriously from language to language,” no doubt, L1-L2 discrepancies is an important source for tense errors in the learner corpus. In this section I will focus on the “Simple past- present perfect” pair, for they are among the “erroneous tense-target tense” pair that are most likely to be confused by Chinese non-English-major university students. As Figure 1 shows, Chinese learners tend to use an erroneous present perfect in place of a target simple past more than the vice versa (the instances are 42 and 7 respectively). For this tendency, as I see it, the root right lies in L1-L2 discrepancies.

In contrast with a tense language like English, Mandarin Chinese is exclusively an aspect language (cf. Wang 1943; Gao 1948:189; Gong 1991:252; Norman 1988), because in Chinese, the concept denoted by tense is indicated by content words like adverbs of time, and aspectual meanings, however, are conveyed systematically by aspect markers — grammaticalized function words that are semantically encoded to convey aspectual meanings (Xiao, 2004). The English present perfect is the combination of the present tense and the perfect aspect. With the inherent semantic feature of describing the lasting processes of things in a time duration, it refers to a state resulting from a single event that took place in the past (Comrie 1985:73). In other words, the present perfect relates a previous situation to the present, i.e. indicating the current relevance of a previous situation, therefore, the sentence “I have finished the wine ten minutes ago” is considered as ungrammatical in English (cf. Trask 1997:165). Nevertheless, Chinese does not have perfect constructions. Although the actual aspect marker -le and the experiential aspect marker -guo also indicate the current relevance state (cf. Li & Thompson 1981), it’s not restricted to the present. The following provides more detailed analyses of the usage of Chinese actual aspect marker –le and the experiential aspect marker–guo.

3.1 Actual Aspect marker –le

In Chinese, the actual aspect –le signals actuality. Actuality simply means that the situation denoted by a sentence actually occurs or materializes, i.e. the situation becomes a reality with respect to the relevant reference time (Xiao 2004: 113). It refers to the situation being actualised in relation only to the specified reference time, which may correspond to the past, the present or the future (cf. Smith 1981:214). When the reference time is anterior to the speech time, the actual -le indicates the actuality of a situation in relation to a past reference time, which behaves like the English simple past. E.g.

1) zuotian wo youlan-le dayingbowuguan.
   Yesterday I visit-le the British Museum.
   “Yesterday I visited the British Museum.”

   This usage can be modeled as follows,
   Time or place in the past + v + actual aspect –le…
On the other hand, when the reference time is unspecified, -le signals that the situation has been actualized in relation to the present, which is regarded as equivalent to the present perfect in English. E.g.

2) wo youlan-le dayingbowuguan.
   I visit-le the British Museum.
   “I’ve visited the British Museum.”

Sometimes, RT is posterior to ST, then -le signals the actuality of a situation in relation to a future RT, which correspond to the English future tense. E.g.

3) wo mingtian xia-le ban qu kan dianying (Dai 1997)
   I tomorrow off-le work go see movie
   “I will go to the cinema after work tomorrow”

3.2 Experiential aspect marker –guo

Unlike the actual aspect –le, the experiential aspect marker –guo focuses on experientiality. It “asserts a discontinuity between the final endpoint of the prior situation and the current state of affairs” (Smith 1997:82), which implies that its span “must include the prior situation and a post-final stage not part of the situation itself” (ibid:83). Therefore, it conveys the event experienced prior to a particular reference time. According to Li & Thompson (1981:228-229) and Tiee (1986:98), this feature emphasizes “the event’s having been experienced at least once” and “being over” with regard to a reference time. Thus, the experiential -guo rarely occurs in future situations. When the reference time is anterior to the speech time, –guo indicates a situation experienced in relation to a past reference time, which is equivalent to English past tense. E.g.

4) wo yinianqian quguo lundun de tangrenjie.
   I a year ago go-guo London Chinatown.
   “A year ago, I went to the Chinatown in London.”

This usage can be patterned as follows,

Time or place in the past + v + actual aspect –le…

When RT is unspecified, the experiential aspect –guo conveys the fact that the event has already happened in relation to the present, In this sense, the experiential aspect –guo marker behaves like the English present perfect. E.g.

5). Wo quguo lundun de tangrenjie.
    I go-guo London Chinatown.
    “I’ve been to the Chinatown in London.”

From the above analysis, we notice that When RT is unspecified (see example 2 and 4), the Chinese experiential aspect marker –guo and the actual aspect marker-le both behave like English present perfect, which tend to be extended by Chinese learners to situations patterned as “time or place in the past + v + actual aspect –le or experiential aspect -guo (see example 2 and 4),” where actually –le and -guo correspond with the
English Simple Past. No wonder in this study (see Figure one), as far as the “Simple past-present perfect” pair (see Figure 1) is concerned, the erroneous Present perfect in place of the Target simple past greatly outnumber the vise verse (i.e. 42 versus 7).

**Implications for College English Education in China**

One pedagogical implication of this study is that English tenses still need to be emphasized when teaching Chinese-speaking EFL learners who have reached a relatively high level of proficiency. As pointed out by Cook (1989: 127), tenses “has often been neglected in language teaching [...]” This neglect is taken for granted by college English teachers and learners in China, as nowadays, the trend in Chinese English college education is changing from grammar-translation-oriented to performed-culture-oriented. Numerous tense errors in the Band 4 and Band 6 sub-corpora in CLEC manifest the detrimental effect of this neglect. Another pedagogical implication of this study is that college English teachers should adopt a contrastive approach when teaching tenses. This approach highlight the difference of the tense uses in the target language and those of the inter-language of Chinese students, for it is believed to be able to enhance learners’ awareness of the English tense system in its own right with its own major underlying principles. Computer corpora are an excellent source of data for teaching tenses by this approach. In the classroom, teachers can make use of the collective electronic data which the students hand in as assignments, guiding them to compare the tense uses in their own interlanguage productions with those in such English native corpus as the Flob or BNC, so as to enable them to better notice gaps between their own tense uses and those in the target language, thereby help them correct errors in their own interlanguage. Once they have realized that most of their tense errors are L1-dependent, they’ll try their best to avoid this kind of errors in the future.

**Conclusion**

This is a study of L2 acquisition of English tense of Chinese non-English-major university students based on Chinese Learner English Corpus (CLEC). The study finds that tenses are difficult to Chinese learners, including Chinese college English students; and the root of this difficulty lies in the discrepancies between Chinese and English tense system. Since the Chinese system of tense has great impact on the acquisition of English tense of Chinese learners, therefore, it’s necessary for EFL teachers to emphasize the complicated relationship between English tense and temporal references to all Chinese English learners, including advanced English learners at tertiary level.

**Note:** Following Johansson and Lysvag (1987: 117), the present paper includes the tense category both the tenses proper (present and past) and the forms which are in effect combinations of tense and aspect (present perfect, past continuous, etc.).
Bibliography
http://www.clal.org.cn/baseinfo/achievement/Achievement1.htm

Acknowledgement
This paper is sponsored by the CSC and School of Foreign Language and Literature of Wuhan University.