A CORPUS-BASED COMPARATIVE STUDY OF THE USE OF
ENUMERATIVE CONNECTIVES IN WRITING: TAKING CHINESE
ENGLISH LEARNERS AND NATIVE SPEAKERS AS EXAMPLES

Ruiqi Peng
Northwestern Polytechnical University, China

DOI: 10.46609/IJSSER.2023.v08i11.015 URL: https://doi.org/10.46609/IJSSER.2023.v08i11.015

Received: 8 November 2023 / Accepted: 21 November 2023 / Published: 30 November 2023

ABSTRACT

Many second language acquisition (SLA) studies are conducted with large amount of language data. To explore the features and trends of SLA from the perspective of English learners, studies based on learner corpus are in great demand. Therefore, taking advantage of discourses in a learner corpus and a reference corpus, this research focused on the similarities and differences in the use of enumerative connectives in English writing between Chinese English learners and native speakers so as to identify the characteristics of the interlanguage of Chinese English learners. It was discovered that there was a significant difference in the use of enumerative connectives between Chinese English learners and native speakers, which was further believed to be intensely connected to the cognitive processes that shape interlanguage linguistic systems.

Keywords: interlanguage; enumerative connectives; corpus; SLA

Introduction

Interlanguage is originally a term proposed by an American linguist, Larry Selinker (1972), who defined it as the systematic knowledge of a second language that is independent of both the learner’s first language and second language. It gradually draws upon the target language with the continuous accumulation of language input. Based on this concept, interlanguage has three characteristics: systematic, dynamic, and permeable. Furthermore, 95% of language learners’ target language will fail to achieve the level of the native speakers and their interlanguage manifests a tendency of fossilization, which is referred to as the fixed state of some language items, grammatical rules, and systematic knowledge in the interlanguage of foreign language learners (Selinker & Lakshmanan, 1992). The growth of age and the variation of learning volume do not effect transforming this fixed state. From the perspective of Selinker, fossilization results because learners acquiring second languages use more general cognitive processes, which he
referred to as latent psychological structure, rather than an innate language-specific UG (Lenneberg, 1967). He identified five cognitive processes that constitute the latent psychological structure and shape interlanguage linguistic systems: language transfer, overgeneralization of target language rules, transfer of training, communication strategies, and learning strategies (Tarone, 2018). Transfer of training occurs when the second language learners apply rules learned from instructors or textbooks. The learners adopt communication strategies to get meaning across when the inter-language system does not provide the necessary forms of communication in a native way. In addition, learning strategies are the conscious attempts of learners to master the target language.

Writing, without a doubt, is one of the four basic skills in language learning. The requirement for writing in a second language is not to merely pile up words into sentences and sentences into paragraphs, but to take the cohesion between sentences and the structure of texts into consideration (Zhao, 2003). There are diverse types of cohesive devices in discourses. The significant ones consist of grammatical devices, lexical cohesion, logical connectives, and pragmatic and semantic implications. Since logical connectives distinctly indicate the connection between parts of text, they have become an important means of creating textual cohesion. Halliday and Hasan (1976) initially divided the logical connectives into 4 broad categories: additive, causal, adversative, and sequential. Later, scholars at home and abroad have generated more specific divisions of logical connectives. According to the previous categories, Zhao (2003) proposed 14 types of connectives: comparison, consequence, exemplification, space, enumeration, emphasis, concession, sequence, substitution, condition, cause, supplement, transition, adverse, and conclusion.

As the interlanguage of language learners, a continuum from one extreme of the native language to the other of the target language, is permeable and transitional (Adjemian, 1976), this study was carried out to figure out the features of the interlanguage of Chinese English learners, as well as the way cognitive processes of Selinker construct the interlanguage system by comparing the use of enumerative connectives in English writing.

**Literature Review**

Before the emergence of learner corpora, the analysis of the interlanguage was often limited to qualitative analysis due to the lack of a large number of representative data (Li, 1999). With the development of corpus linguistics, the establishment of learner corpora, and the continuous improvement of computers and corresponding software, qualitative analysis and quantitative analysis have been further and better combined, along with the analysis methods of interlanguage becoming more and more scientific. One of the advanced methods of interlanguage analysis is contrastive interlanguage analysis (CIA) coined by Granger (1996), which mainly consists of
comparisons of two aspects. The former is the comparison between the interlanguage and the target language to reveal the similarities and differences between the two, and reflect the characteristics of interlanguage such as over-use, under-use and misuse, which is the core part of the CIA. The latter is the comparison of learner corpora with different native language backgrounds to reflect the features of language output of learners with such backgrounds, which aims to investigate the characteristics of interlanguage in different learner groups. To apply it into practice, Granger initiated the International Corpus of Learner English and conducted relevant empirical studies. In recent years, CIA has been widely adopted in the field of SLA research at home and abroad to further explore the features of interlanguage (see, e.g., Howarth, 1998; Ağçam, 2014; Huang, 2015; Pan & Feng, 2004; Liu & Miao, 2011) and have achieved remarkable research results. For instance, Howarth (1998) discovered that compared with native English speakers, EFL learners rarely use restricted collocations through a comparative study.

With regard to the studies of cohesive devices, the empirical ones overseas can be divided into two trends: cross-stylistic and cross-disciplinary comparative studies based on the corpus of native speakers (Peacock, 2010) and cross-group comparative studies based on the corpus of learners and native speakers or studies on specific linking adverbials (see, e.g., Conrad, 1999; Charles, 2011; Kim & Yeates, 2019). In terms of domestic research, there are also two inclinations, which are the general studies of Chinese EFL learners’ use of linking devices in writing (see, e.g., Chen, 2002; Zhao, 2003; Du et al., 2013) and concrete studies of a type of connectives, some focusing directly on certain connectives (see, e.g., Zhang & Zhou, 2007; Zhang & Lv, 2019; Yang, 2019). Nevertheless, a great number of them laid emphasis on the description of the language phenomenon rather than in-depth exploration and interpretation of the result from the perspective of interlanguage system and its cognitive processes. Moreover, the enumerative connectives in the cohesive devices have scarcely been taken seriously in the related studies.

In terms of practical consideration, globalization requires fluent communication between different countries and districts to ensure the running of a series of economic and cultural activities. During this progress, English as a lingua franca has an important communicative function. As China nowadays is always playing a critical role in the rapid development of globalization, an enormous quantity of fluent English speakers is in urgent demand. It is thus rather crucial to figure out how to improve the quality of Chinese English education from research on SLA and, specifically, interlanguage.

Therefore, in order to address the gap mentioned above, the present paper combined Selinker’s interlanguage concept and an exhaustive empirical analysis of two existing corpora and endeavored to identify and interpret the similarities and differences in the use of enumerative
connectives between Chinese English learners and native speakers. Furthermore, since multitudinal studies on language transfer concentrate on the misuse of the target language rather than the other two manifestations (see, e.g., Zhou, 2007; Song, 2010; Dissington, 2018), this research laid more emphasis on under-use and over-use. It intends to benefit not only Chinese students, but also provide pedagogical implications for the advancement of Chinese English education. The research questions are as follows:

1. What are the similarities and differences in the use of enumerative connectives in writing between Chinese English learners and native speakers?

2. Do Chinese English learners tend to overuse or underuse enumerative connectives compared with native speakers?

**Methodology**

**Data Collection**

In this corpus-based study, a Chinese English learner corpus and a native speaker corpus were adopted, respectively, adopted to draw a comparison in the use of enumerative connectives in the English writing.

The learner corpus used is WECCL (Written English Corpus of Chinese Learners), a written sub-corpus of SWECCCL (Spoken and Written English Corpus of Chinese Learners), which was built by Wen Qiufang et al. in 2005, with a total capacity of approximately 2 million tokens. For the sake of investigating the general picture of the use of enumerative connectives of Chinese English learners with intermediate English level or above, the article employed the essays of non-English major students from freshmen to seniors, with a total of 591 texts, 144,681 tokens.

The native-speaker reference corpus selected is NESSIE (Native English Speakers’ Similarly-and Identically-promoted Essays), a corpus mainly composed of English writings from British and American native speakers according to the topics of writings in CET or TEM as well as other writings with similar topics chosen from BAWE, MICSUP, and some other corpora of British and American university students with a total of 256 texts, 193,844 tokens. It was established in 2012 by the team of Chinese scholar Xu Jiajin as a substitute for LOCNESS (Louvain Corpus of Native English Essays) since LOCNESS is inaccessible to Chinese language researchers and incomparable to the corpus of Chinese university students with regard to the different topics of their writings.

**Tools**

To analyze the discourses in the two corpora, WordSmith Tools 7.0 and SPSS 22 were opted as
the research instruments.

WordSmith Tools 7.0 is invented as a retrieval tool especially for corpus analysis. There are WordList, Concord, KeyWords and many other functions. In this study, the software was mainly used to conduct concordance analysis and collect frequency statistics.

SPSS 22 (Statistical Package for the Social Sciences) is a powerful statistical software that covers a vast range of advanced statistical procedures. It offers a user-friendly interface and robust features that enable researchers’ organization to rapidly extract actionable insights from their data. In this study, the research used SPSS to identify whether there were significant differences in the use of enumerative connectives between Chinese English learners and native speakers.

**Data Analysis**

Above all, based on the classification of logical connectives provided by previous scholars (Zhao, 2003; Zhang & Lv, 2019), the study selected the following 18 types of enumerative connectives from the two corpora: first, second, firstly, secondly, for one thing, for another, in the first place, in the second place; first of all, to begin with, to start with, for a start, next, then, last, finally, on (the) one hand, on the other hand.

With these enumerative connectives being the typical samples, the study used WordSmith Tools 7.0 to retrieve all these connectives in WECCL and NESSIE respectively and examine the frequency. To answer the first question, the normalized frequency of enumerative connectives in two corpora was input into SPSS 22 to test whether there are significant differences in using enumerative connectives between Chinese EFL learners and native speakers through Chi-Square Tests.

According to the data obtained above, the researcher compared the use of enumerative connectives in two corpora so as to provide answers for the second research question and determine the cognitive processes that impact the interlanguage of Chinese learners.

**Results and Discussion**

**Similarities and Differences of Enumerative Connectives Used in WECCL and NESSIE**

Through retrieving 18 types of enumerative connectives in WECCL and NESSIE and analyzing the frequency, it was found that there were 509 enumerative connectives in WECCL and 120 enumerative connectives in NESSIE. Since the two corpora differ in size, normalized data, i.e. the number of tokens per 1,000,000 words, were also listed for comparability. Then the normalized frequencies of enumerative connectives in the two corpora were respectively
detected to be 3518.08 and 619.05. In general, it is observed that Chinese English learners are inclined to use more enumerative connectives in their writing than native speakers. In order to determine whether the difference is significant, the Chi-Square Test was performed using SPSS 22. The result is demonstrated in Table 1.

Table 1: Chi-Square Test of Overall Enumerative Connectives in WECCL and NESSIE

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
<th>Exact Sig. (2-sided)</th>
<th>Exact Sig. (1-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>375.428</td>
<td>1</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Continuity Correction b</td>
<td>373.866</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>386.789</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisher's Exact Test</td>
<td>375.427</td>
<td>1</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>338525</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 0 cells (0.0%) have expected count less than 5. The minimum expected count is 268.83.

According to Table 1, the Chi-Square value is 375.428 and the significance is 0.000 (<0.05). From the result, it is proved that there is significant difference in the use of enumerative connectives between the two corpora.

To further explore the differences of relevant data in WECCL and NESSIE, the frequencies of 18 types of enumerative connectives are respectively illustrated for comparison as follows:

Table 2: Frequencies of Individual Enumerative Connectives in WECCL and NESSIE

<table>
<thead>
<tr>
<th>Type</th>
<th>Occurrence (n)</th>
<th>Frequency (n/1,000,000)</th>
<th>Type</th>
<th>Occurrence (n)</th>
<th>Frequency (n/1,000,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>first</td>
<td>96</td>
<td>663.53</td>
<td>first</td>
<td>28</td>
<td>144.45</td>
</tr>
<tr>
<td>then</td>
<td>82</td>
<td>566.76</td>
<td>then</td>
<td>22</td>
<td>113.49</td>
</tr>
<tr>
<td>second</td>
<td>61</td>
<td>421.62</td>
<td>finally</td>
<td>21</td>
<td>108.33</td>
</tr>
<tr>
<td>secondly</td>
<td>57</td>
<td>393.97</td>
<td>firstly</td>
<td>16</td>
<td>82.54</td>
</tr>
</tbody>
</table>
Table 2 has distinctly demonstrated the relevant usage in the two corpora. There is a group of similarities and differences between Chinese non-English majors and native speakers in the use of enumerative connectives.
Primarily, in the top five enumerative connectives in the two corpora, there are four overlapping, with the first two both “first” and “then”. At the same time, four out of the last five enumerative connectives are the same in WECCL and NESSIE. This similarity between the two corpora is consistent with the findings of Zhang and Lv (2019). Additionally, “first”, “firstly”, “second”, and “secondly” all rank high in the two corpora, which can be interpreted as the language transfer from Chinese to English. In Chinese, people tend to adopt numbers, especially ordinal numbers, to highlight the structure of the writing, which is coincident with the way English native speakers construct articles in English.

However, despite the fact that the most and the least frequently used enumerative connectives coincide a lot, the overall normalized frequency of them in WECCL far outweighs that in NESSIE, which is also illustrated in Table 1, suggesting Chinese non-English majors use more enumerative connectives than native speakers. The finding is similar to that of studies on logical connectives (Zhao, 2003) and additive connectives (Chen & Jiang, 2015). Furthermore, some particular connectives including “last”, “on the other hand” rank largely higher in WECCL than those in NESSIE while “finally” and “next” in the former corpus lower than those in the latter corpus.

Additionally, several enumerative connectives appeared in the learner corpus manifest a zero-occurrence in the reference corpus, which is believed to result from a limited amount of discourses and tokens in NESSIE.

**Over-use and Under-use of Enumerative Connectives in WECCL and NESSIE**

Over-use and under-use (avoidance) of enumerative connectives by learners were identified though a comparison of the normalized frequencies of enumerative connectives in WECCL and NESSIE as well as the Chi-Square Tests of each type of these connectives. And the results are listed in Table 3:

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency (n/1,000,000)</th>
<th>Chi-Square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>WECCL</td>
<td>NESSI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first</td>
<td>663.53</td>
<td>144.45</td>
<td>60.964</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Standard Deviation</td>
<td>t-value</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td>then</td>
<td>566.76</td>
<td>113.49</td>
<td>55.422</td>
</tr>
<tr>
<td>second</td>
<td>421.62</td>
<td>67</td>
<td>47.653</td>
</tr>
<tr>
<td>secondly</td>
<td>393.97</td>
<td>30.95</td>
<td>58.676</td>
</tr>
<tr>
<td>last</td>
<td>359.41</td>
<td>5.16</td>
<td>66.417</td>
</tr>
<tr>
<td>firstly</td>
<td>304.12</td>
<td>82.54</td>
<td>22.953</td>
</tr>
<tr>
<td>on the other hand</td>
<td>290.29</td>
<td>5.16</td>
<td>53.034</td>
</tr>
<tr>
<td>on (the) one hand</td>
<td>103.68</td>
<td>15.48</td>
<td>12.121</td>
</tr>
<tr>
<td>first of all</td>
<td>96.76</td>
<td>5.16</td>
<td>15.691</td>
</tr>
<tr>
<td>finally</td>
<td>76.03</td>
<td>108.33</td>
<td>0.915</td>
</tr>
<tr>
<td>above all</td>
<td>69.12</td>
<td>5.16</td>
<td>10.430</td>
</tr>
<tr>
<td>next</td>
<td>62.21</td>
<td>30.95</td>
<td>1.826</td>
</tr>
<tr>
<td>for one thing</td>
<td>27.65</td>
<td>0</td>
<td>5.359</td>
</tr>
<tr>
<td>in the first place</td>
<td>20.74</td>
<td>5.16</td>
<td>1.701</td>
</tr>
<tr>
<td>in the second place</td>
<td>20.74</td>
<td>0</td>
<td>4.019</td>
</tr>
<tr>
<td>to begin with</td>
<td>20.74</td>
<td>0</td>
<td>4.019</td>
</tr>
<tr>
<td>for another</td>
<td>13.82</td>
<td>0</td>
<td>2.680</td>
</tr>
<tr>
<td>to start with</td>
<td>6.91</td>
<td>0</td>
<td>1.340</td>
</tr>
</tbody>
</table>
Table 1 in the last section has already confirmed that there is a significant difference in the use of enumerative connectives between the learner corpus and the native corpus, which suggests that in a general sense, Chinese non-English major learners are prone to overuse enumerative connectives compared to native speakers. Then Table 3 further shows that 13 out of 18 types of enumerative connectives are overused by Chinese learners since their Chi-Square Tests’ significance is less than 0.05.

Among them, “first”, “then”, “second”, “secondly”, “on the other hand”, “last” are of relatively high Chi-Square value, which indicates the degree of their over-use is much higher than the others.

The above results are highly consistent with many precious comparative studies on connectives used by Chinese EFL learners and English native speakers (see, e.g., Zhao, 2003; Pan & Feng, 2004; Mo, 2005; Chen & Jiang, 2015), which can be interpreted by multiple factors from the perspective of language transfer. To be concrete, initially, influenced by the implicit feature of grammar of Chinese, Chinese students are prone to attach more importance to parataxis and less to logic. Therefore, they are prone to neglect connectives in English writing (Zhao, 2003). Given this phenomenon, teachers consciously place more emphasis on this section and sometimes too much on the significance of connectives. What is more, to avoid more difficult and error-prone expressions as alternatives, students are likely to apply the simple forms of enumerative connectives to achieve communicative purpose in their writings. Therefore, a combination of the three cognitive processes mentioned, language transfer, transfer of training, and strategies of communication, results in an overuse of “first (ly)” and “second (ly)” in the learner corpus.

Next, “last” is an enumerative connective with the highest Chi-Square value, which suggests that Chinese non-English majors overuse the related expressions to a great extent. On the contrary, the token “finally”, which possesses a similar literal meaning to “last”, has shown a lower normalized frequency in WECCCL than in NESSIE. However, the difference is not significant (>0.05). The overuse of “last” rather than “finally” is enormously due to the impact on learners’ native language. In Chinese, expressions referring to a final point are of neutral meaning and a great multitude of these terms are able to be used in common, which are entirely opposite to those in English. In English, “finally” is not feeling-attached, while “last”, particularly when appearing in “at last”, has a negative connotation. As for native speakers, “last” or “at last” in most cases are adopted to declare impatience due to long delays. With regard to Ellis (1999), the difficulty of acquiring a second language can be directly influenced by the difference between the native language and the second language. In this study, likewise, the differences between the native language and the target language can lead to a negative language transfer, the over-use of “last” by Chinese learners.
Additionally, the cultural thought patterns aroused from mother tongue will affect the way the second language learners organize their writing discourses (Kaplan, 1966). From observation of the two corpora, the over-use of “on the other hand” can be mostly attributed to the language transfer and transfer of training. Chinese speakers, influenced by their native language, are likely to merely regard “on the other hand” as an enumerative connective and default that it should be used in a context in which there is “on (the) one hand” ahead. However, in the native corpus, 12 out of 13 are used to indicate a contrastive meaning. What is more, teachers’ introduction and explanation of “on (the) one hand” and “on the other hand” may not be comprehensive since only one of the functions and implications is covered, which can probably aggravate the fossilization of students’ interlanguage (Fang, 2014).

Conclusion

This study concentrated on the topic of interlanguage by investigating the similarities and differences in the use of enumerative connectives in the WECCL and NESSIE corpus as well as whether Chinese English learners tend to overuse or underuse such cohesive devices compared with native speakers. It was observed that there exist both similarities and differences in two corpora. And Chinese English learners generally manifested a tendency to overuse enumerative connectives, which turns out to be influenced by the learners’ cognitive processes: language transfer, transfer of training, and strategies of communication.

Based on the research findings, several implications were also provided for English education in China. Above all, teachers ought to properly emphasize the differences between native language and target language and instruct students to complete positive language transfer. In spite of the necessity of enumerative connectives to the organization of English discourses, these cohesive devices should be as concise as possible without superseding the focus. Otherwise, the excessive appearance of explicit logical connectives will distract readers from the major meaning of the whole article.

Secondly, teachers must improve the accuracy and richness of students’ use of enumerative connectives. When teaching a certain connective, they should be aware of comprehensively introducing its functions and usage in different discourses and contexts to prevent a rigid structure. In addition, it is vital to assist students in learning to insert more advanced enumerative connectives and cut down the over-use of lower-level ones.

Ultimately, in class, teachers can take a large number of discourses in the native corpus such as NESSIE and BROWN as evidence, guide students to imitate the application of various conjunctions and cohesion, help them transform their native language thinking, and thus promote their idiomatic use of logical connectives.
Limitations and Suggestions for Further Research

The present paper is far from perfect. Limitations and future research directions thus deserve comment.

Firstly, to explore the features of interlanguage, selecting the learner corpus in which learners are from a single native language context is far from enough. More corpora of language output from authors with different language backgrounds should be considered in the following studies.

Secondly, apart from comparative analysis of interlanguage, error analysis is also a crucial means of investigating language transfer. While the paper collected and analyzed the frequency of enumerative connectives in two corpora, it failed to examine them individually to see if they are used appropriately and the reasons and ways of the misuse, which are no doubt critical data to further identify the characteristics of interlanguage.

In addition, in terms of the native corpus, this study only focuses on the non-English majors’ English writing. However, the use of enumerative connectives of learners at different levels may possibly reveal the dynamic development of interlanguage. Therefore, it is essential to take English at different levels into consideration in future research.

References


Yang, L. (2019). A corpus-based descriptive study of the use of disjunctive connectives in argumentative writing by English Majors. Education Modernization, 6(17), 118-120.


