



# A Corpus-based Study of Connectors in Student Writing: A Comparison between a Native Speaker (NS) Corpus and a Non-native Speaker (NNS) Learner Corpus

Yuanyi Ma

School of Foreign Languages, Guangdong Institute of Science and Technology  
Jinwan District, Zhuhai, Guangdong Province, China  
Faculty of Humanities, The Hong Kong Polytechnic University  
E-mail: [yuanyi.ma@connect.polyu.hk](mailto:yuanyi.ma@connect.polyu.hk)

Bo Wang (Corresponding author)

Centre for Translation Studies, The Hong Kong Polytechnic University  
Hung Hom, Kowloon, Hong Kong  
E-mail: [bo.wang@connect.polyu.hk](mailto:bo.wang@connect.polyu.hk)

Received: 08-06-2015

Accepted: 21-09-2015

Advance Access Published: October 2015

Published: 01-01-2016

doi:10.7575/aiac.ijalel.v.5n.1p.113

URL: <http://dx.doi.org/10.7575/aiac.ijalel.v.5n.1p.113>

## Abstract

The present study offers a corpus-based analysis of two written corpora, i.e. a native speaker corpus (part of the Louvain Corpus of English Essays) and a non-native speaker corpus, which is composed of several essays written by Hong Kong first year university students. It aims to find out the differences in the use of connectors. Altogether twenty-five connectors used in the two corpora have been examined, with the frequency of their usage being calculated. The reasons that have led to the variations in using these connectors are further discussed. This corpus-based study on connectors would have some pedagogic implications on writing teaching. As connectors have played a significant role in both sentence construction and coherence, based on our analysis, we hope that they should be taught explicitly, practiced extensively, and illustrated with learning materials given to students, so as to avoid the overuse and misuse in writing.

**Keywords:** Connector, Native speaker corpus, Non-native speaker corpus, Pedagogic implication

## 1. Introduction

Born in the 1960s, corpus linguistics is defined as “a linguistic methodology, which is founded on the use of electronic collections of naturally occurring text, viz. corpora” (Granger, 1998; Granger, Hung, & Petch-Tyson, 2002:4). For the past 30 years after its development, studies on corpus have focused on issues like register, dialect analysis, etc. As a subcategory of corpora, learner corpora refer to the corpora of non-native English, which applies “the methods and tools of corpus linguistics to gain better insights into authentic learner language” (Granger, 1998: xxi). Despite their significance in application, studies on learner corpora have started at a much later time. Therefore, the present study is expected to further exploit the implications of learner corpora in teaching.

In various previous studies, connectors have been explored under different names, including discourse connectors, linking words, connectives, and discourse operators (see e.g. Cowan, 2008; Rezvani Kalajahi, Abdullah, Mukundan and Tannacito, 2012 for distinctions between these terms). Discourse markers, which bind the whole text together and indicate strategies of coherence, are indispensable to written discourse. It has been shown that discourse connectors should be used carefully in writing; both lack of use and overuse should be avoided (Rezvani Kalajahi et al., 2012).

## 2. Literature Review

### 2.1 Learner Corpora

The development of science and technology has allowed researchers to collect learner data in large quantities electronically, as well as to analyze the data with linguistic software. Granger (1998: 7) has given a detailed definition of computer learner corpora: “Computer learner corpora are electronic collections of authentic FL (foreign language) / SL (second language) textual data assembled according to explicit design criteria for a particular SLA (second language acquisition) / FLT (foreign language teaching) purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance”.

For computer learner corpora, several criteria are set in terms of corpus design. From the aspect of language, these features include medium, genre, topic, technicality, and task settings. For medium, formal and informal corpora are distinguished. For genre, argumentative, narrative and spontaneous writings are discerned. In addition, topical features are related to learners’ lexical choice. Technicality would affect both lexical choice and grammar. And for task setting, degree of preparedness and whether the data are part of an exam would be taken into consideration (Granger, 1998).

Meanwhile, as to the profiles of learners, aspects such as age, sex, mother tongue, region, other foreign languages learned, language proficiency, learning context, and practical experience should all be considered as important factors in corpora design.

There are two ways of analyzing learner corpora, namely, contrastive interlanguage analysis (CIA) and computer-aided error analysis (CEA). Firstly, CIA involves two types of comparisons, NS (native speaker) / NNS (non-native speaker) comparison, and NNS/NNS comparison. On one hand, after a detailed comparison of NS/NNS corpora, e.g. under and overrepresentation of words, phrases and structures, we could gain a deeper and better understanding of the features of learners' writing. On the other hand, an NNS/NNS comparison would help us to find out the specific features pertaining to learners of a certain nationality, and may further improve our knowledge of interlanguage (Granger, 1998). Secondly, CEA, the other way of analysis, could provide data from linguistic perspectives (e.g. word, phrase, word category, syntactic structure) for understanding interlanguage development, which would eventually shed some light on pedagogical framework (Granger, Hung & Petch-Tyson, 2002). The present study would adopt the CIA method, as it is suitable for studying features of interlanguage.

## 2.2 Conjunctions in English

Using conjunctions or connectors is a way of achieving cohesion, and it is a linking device across sentences (Field & Oi, 1992). As to the function of conjunctions, Halliday and Hasan (1976) have the following description: "conjunctive elements are cohesive not in themselves but indirectly, by virtue of their specific meanings which presuppose the presence of other components in the discourse" (Halliday & Hasan, 1976: 226).

Four categories of conjunctive relations are introduced in Halliday and Hasan (1976), i.e. the additive, adversative, causal and temporal. The additive "signals the 'and' link between sentences", and includes conjunctions such as "also", "furthermore", "in addition", "besides", "similarly", "in the same way", "on the other hand", "by contrast", etc. The adversative "signals that what follows is contrary to what has been stated", such as "yet", "though", "but", "however", "nevertheless", "actually", "on the other hand", etc. The causal conjunction "presupposes a reason or argument", such as "so", "then", "hence", "therefore", "because", etc. The last type of conjunction, the temporal refers to the time sequence of events, e.g. "then", "next", "at once", "there upon", etc. Despite using different terminology from the present study, Halliday and Hasan (1976) have clearly shown the effectiveness of connectors as cohesive devices in the semiotic processes.

Rezvani Kalajahi et al. (2012) have identified the significance of connectors in the writing of second language users. To some extent, the inappropriate use of connectors could hinder successful communication or even lead to misunderstandings. It is therefore significant for second language learner to acquire the proper way of using connectors in the target language.

## 2.3 Some Previous Studies on the Chinese Students' Use of Conjunctions

Field and Oi (1992) have carried out a comparative study on conjunctive cohesion used in written essays by Cantonese speakers and native speakers. The analysis is based on the four categories of conjunctions mentioned by Halliday and Hasan (1976). A complete range of conjunctions is found in the data, which include "149 adversatives, 140 additives, 97 causals and 49 temporals" (Field & Oi, 1992:2). It is concluded that Cantonese students use more conjunctive devices than their native speaker counterparts; the Cantonese students are more inclined to use conjunctions at the beginning of a paragraph or a sentence rather than in the middle; and Cantonese speakers choose conjunctions in a wider range than L2 speakers.

The use of cohesive features in argumentative writings by Chinese college students were studied by Liu and Brain (2005), who examined 50 argumentative essays of Chinese undergraduate non-English majors to investigate their use of cohesive conjunctions. Both timed and untimed essays were chosen as data. Analysis of the data shows that students can use lexical, reference and conjunction devices in their writing. However, some of the conjunction words such as "and", "but", "or" and "so" are used more frequently than other conjunction words or phrases like "furthermore", "on the contrary", "moreover", "in addition", etc.

Similar to Field and Oi's (1992) study, another two researches in Hong Kong are both concerned with the use of connectors. One study (Milton & Tsang, 1993) compares the use of 25 logical connectors of students from a Hong Kong tertiary institute. Three corpora are compared, namely the American Brown Corpus, the British LOB Corpus, and the HKUST corpus, which consists of excerpts from first year Computer Science textbooks. After comparisons of raw frequency, it is concluded that half of the connectors in the list were overused in the NNS corpus, which includes items such as "also", "moreover", "furthermore", "regarding", "namely", "nevertheless", "although", "because", "therefore", "first", "secondly", and "lastly". The study also relates the misuse or overuse of some items such as "moreover", and "therefore".

Another study conducted among Hong Kong students is by Bolton, Nelson and Hung (2002), which is about the underuse and overuse of connectors by comparing essays written by Hong Kong students with those by British students. In this study, the comparison is based on data from the Hong Kong component and British component of the International Corpus of English (ICE). Moreover, "10 untimed essays and 10 timed examination scripts written by undergraduate Hong Kong students" (Bolton, Nelson and Hung, 2002: 173) are also included in the data. By calculating the raw frequency and frequency per sentence of a list of self-designed connectors in both NS and NNS corpora and comparing the numbers with that of professional academic writing, they have arrived at the conclusion that the use of connectors by Hong Kong students are much more different from the academic norm. For Hong Kong students, the most overused connector is "so", while the most overused connector in the British data is "however".

### 3. The Study

The present study aims to find out the differences in the use of conjunctions as well as variations between the NS and the NNS written corpora. We have retracted part of the LOCNESS (The Louvain Corpus of English Essays), which is a corpus of native English essays as the NS corpus in our study. LOCNESS is a corpus of native English essays made up of British and American students' written essays, which includes 324,304 words totally. The present NS corpus consists of 46 essays (48,281 words) written by the first year and the fourth year US native students from Marquette University. These are untimed essays of around 500 words each, focusing on a wide range of social issues. Moreover, the students are allowed to use reference tools while writing.

On the other hand, our NNS corpus (48,721 words in total) is composed of 45 essays written by Hong Kong first year university students of different majors as one of the assignment for a selective English written course. The students' mother tongue is Cantonese, and their essays are argumentative in genre, and have centered on the topic of "New media, new meanings". The students are allowed to complete the essays in one week, and are encouraged to cite references from their reading materials.

As previously mentioned, the study aims to compare the use of connectors between NS (featured by US Marquette University students) speakers corpus and NNS (featured by Hong Kong college students) learner corpus, to find out their different use patterns and to discuss the possible reasons behind these patterns, so as to provide future pedagogical implications.

The present study has first chosen 25 connectors from Milton and Tsang's (1993) research on Hong Kong students' use of connectors. The reason lies in that these connectors have frequently been used in the essays of Hong Kong students (cf. Milton and Tsang, 1993), and they are also the frequent choices of connectors found in our data. For this reason, they are selected for the analysis in the present study. The raw frequency of these connectors and their percentage of use in the corpora are calculated and illustrated in Table 1.

Table 1. Raw frequency and percentage of connectors used in the writings by Hong Kong students

connectors	NNS corpus		NS corpus	
	rf	%	rf	%
also	180	0.37	122	0.25
moreover	23	0.05	1	0
furthermore	8	0.02	7	0.01
and	1486	3.03	1091	2.28
besides	9	0.02	3	0
actually	10	0.02	18	0.04
alternatively	0	0	0	0
including	10	0.02	6	0.01
regarding	10	0.02	6	0.01
similarly	1	0	2	0
likewise	0	0	0	0
namely	2	0	0	0
nevertheless	6	0.01	1	0
although	20	0.04	25	0.05
because	52	0.11	161	0.34
therefore	40	0.08	25	0.05
consequently	1	0	1	0
firstly	6	0.01	0	0
secondly	6	0.01	0	0
previously	7	0.01	6	0.01
afterwards	1	0	0	0
eventually	4	0	4	0
finally	4	0	7	0.01
lastly	1	0	0	0
anyhow	0	0	0	0
anyway	0	0	3	0

As it is seen from Table 1, some of the most common connectors have not appeared in both NS and NNS corpora, e.g. “alternatively”, “likewise”, and “anyhow”, which is reflected in the low frequency of use of these words in NS and NNS written discourse. Some words have more occurrences in the corpora, such as “similarly” (NNS corpus: 1 occurrence; NS corpus: 2 occurrences), “namely” (NNS corpus: 2 occurrences; NS corpus: 0 occurrence), “afterwards” (NNS corpus: 1 occurrence; NS corpus: 4 occurrences), “lastly”, (NNS corpus: 1 occurrence; NS corpus: 0 occurrence), “any” (NNS corpus: 0 occurrence; NS corpus: 3 occurrences), “consequently” (NNS corpus: 1 occurrence, NS corpus: 1 occurrence), etc. The low frequency of occurrences may partly due to the rare usage of these connectors in NS/NNS college students’ written discourse, or because of the limited size of the present corpora.

Some other connectors are used quite frequently, such as “although” (NNS corpus: 20 occurrences, NS corpus: 25 occurrences), “because” (NNS corpus: 52 occurrences; NS corpus: 161 occurrences), “therefore” (NNS corpus: 40 occurrences; NS corpus: 25 occurrences), “moreover” (NNS corpus: 23 occurrences; NS corpus: 1 occurrence), “also” (NNS corpus: 180 occurrences; NS corpus: 122 occurrences). Meanwhile, the most frequently used connector in both NNS and NS corpora is “and” (NNS corpus: 1486 occurrences; NS corpus: 1091). In respect of the different uses between the NNS and NS corpora, some connectors are used more frequently by native speakers, such as “because” (109 more occurrences), while three connectors are used more frequently by the NNSs, i.e. “moreover” (22 more occurrences), “also” (58 more occurrences), “and” (395 more occurrences). As teachers of a second language, we should notice these differences in terms of percentages, help students to avoid the overused and misused connectors, and encourage them to use proper connectors where necessary so as to achieve naturalness in writing.

A detailed concordance list is generated with a purpose of examining the patterns of using “because” and “moreover” between non-native speakers and native speakers (see Table 2).

Table 2. Patterns of using “because”: A comparison between NNS and NS corpora

Different patterns of “because”	NS (161 in total)	NNS (52 in total)
Sentence initial	15	6
“argument because”	0	10
Connect parallel sentence	Most of the cases	Most of the cases
be because	10	6
because of	39	22
because the	24	7
because it	21	3
because they	17	4

In the concordance list of “because”, several patterns are found in the NS corpus. Firstly, “because” is mostly used as a connector between two parallel sentences. Out of the 161 occurrences, 15 of them are sentence initials, 10 mark the collocation “argument because”, and 10 follow the “be because” structure. Secondly, we have found 39 occurrences of “because of”, 24 of “because the”, 21 of “because it”, and 17 of “because they” (see Figure 1 for examples of concordance). Comparatively, in the NNS corpus, 6 of them are used as sentence initials, 6 are used in the “be because” pattern, 22 “because of”, 7 “because the”, 3 “because it”, and 4 “because they”. Hence, it seems that “because” has been mostly used to link parallel sentences. Native speakers tend to use “because” more frequently than the non-native speakers. There are 39 (24%) occurrences of “because of” in the NS corpus, 22 (42%) occurrences in the NNS corpus, from which we have observed a percentage of 18% higher of using “because of” in the NNS corpus. Moreover, the fewer occurrences in the NNS corpus may also because the NNS speakers have chosen other conjunctions or phrases to express the causative relationships, such as “thus”, “therefore”, “hence”, “so”, “for this reason”, etc. The construal of logical meaning (cf. Matthiessen, 1995; Halliday and Matthiessen, 2014) in second language writing deserves our further attention. We should not only consider the raw frequency or distribution, but should also support the analysis of large data with detailed grammatical analysis.

1 to women's rights advocates. *Because* the men's teams serve  
 2 n schools is the following: *Because* the prayer is non sect  
 3 pressionable part of society. *Because* children may model wha

1 city would be effective in an *argument because* virtually all  
 2 s a very strong aspect to his *argument because* it is hard to  
 3 f. This statement damages his *argument because* the compariso

1 uch a huge capital investment is *because* it will then run sa  
 2 likely that he will think it is *because* of something he did  
 3 a person is on death row, it is *because* they are a known mau

1 to its different values, and *because it* is not yet establis  
 2 strong aspect to his argument *because it* is hard to argue ag  
 3 its strength in this argument *because it* makes people think

1 tire basis of their argument. *Because of* the large amount of  
 2 ause of its inefficiency, but *because of* its brutality and l  
 3 rtile couples to have a child *because of* the emotional strai

1 to women's rights advocates. *Because the* men's teams serve  
 2 tatement damages his argument *because the* comparison of musi  
 3 nite weakness in the argument *because the* reader is not full

Figure 1. Some examples of concordances

We have also searched for “since”, which functions as an equivalent of “because” in the two corpora. We have found 55 and 29 occurrences respectively in the NNS and the NS corpus, which may due to the fact that in second language classroom, “since” is a more formal word to use in essay than “because”. Therefore, Hong Kong college students tend to use “since” more often than “because” while stating the reason. Similarly, Hong Kong students tend to use “due to” more often (58 occurrences) than their NNS counterparts (14 occurrences). This may indicate a preference of Hong Kong college students to use more noun phrases than sentences.

In the concordance list, 26 occurrences of “moreover” are found in the NNS corpus, compared with the 1 occurrence in the NS corpus. A detailed look at the instance would give us more hints on this issue (see Table 3, with “moreover” being italicized). Among these examples, the first one is a redundant use of “moreover”, which has the similar meaning to “not only... but also...” In instance 2, “moreover” is not used to link an additional statement, but a different statement. So it is misused here. In instance 3, the writer uses “moreover” to cite a researcher’s critic without giving related statements previously. Therefore, the use of “moreover” here is an example of being overused. From these instances, we may suggest that Hong Kong students prefer to use “moreover” in their essay writing, but some of the usages are problematic. Our finding coincides with that of Milton & Tsang (1993) on the overuse of “moreover” in Hong Kong students’ writings.

Table 3. Three examples of “moreover” in the corpus

No.	Instance	Description
1	The messages of the new media do not flow in one way but in both directions. It enhances the degree of participation of the general public. <i>Moreover</i> , they not only act as consumers, but also as the producers.	used redundantly
2	Leaving a voice message is possible, yet it still cannot contact the person and convey impotent message immediately. <i>Moreover</i> , wired telephone usually installed per family unit, company or organization as a common property mainly nowadays.	misused
3	Indeed, in reality, the media tasks would never stand individually and the communicative purposes of contexts of use overlaps. (Bhatia, 2004). The most typical example would be news report on television which conveys information about current events and circumstances in society and the world as well as makes audience amused and relaxed. <i>Moreover</i> , McQuail (1994) criticised the Functionalist theory could not explain power and conflicts.	overused

#### 4. Conclusion

The present study has compared 25 connectors from the NNS/NS corpora, and has arrived at the following conclusions: firstly, both the Cantonese and the US students use some of the connectors rarely, such as “likewise”, “anyhow” and “alternatively”; secondly, the Cantonese and the US students use some connectors occasionally, e.g. “besides”, “actually”, “previously”, “eventually”, “namely”, “similarly”; thirdly, some connectors are used more often by the Cantonese and the US students, such as “therefore”, “because”, and “although”. Specifically, the US students use “because” more frequently than the Cantonese students, while the Cantonese student use “since” and “due to” more often than their counterparts in the Western Hemisphere.

The reason for the Cantonese students to use “because” less frequently may be based on the established notion that “since” is a more formal word than “because”. Obviously, the US students do not have this notion, so they have used “because” quite extensively in their written essays. Also, the reason for the more frequent use of “due to” may be because that the Hong Kong students take phrasal nouns as more controllable expressions than clauses. Besides, we have also found that the Cantonese students use “moreover” more often than their US counterparts, which could signal the overuse and sometimes misuse of this word.

Due to the significance of connectors in sentence construction and coherence, we suggest that they should be taught explicitly and practiced extensively, and standard examples should be given as learning materials so as to avoid further overuse and misuse. We believe that this corpus-based study on connectors would have some pedagogic implications on writing teaching.

#### References

- Bolton, K., Nelson, G., & Hung, J. (2003). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong ICEHK. *International Journal of Corpus Linguistics*, 7(2), 165-182. doi: <http://dx.doi.org/10.1075/ijcl.7.2.02bol>.
- Cowan, R. (2008). *The teacher's grammar of English*. Cambridge: Cambridge University Press.
- Field, Y., & Oi, Y. L. M. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC journal*, 23(1), 15-28. doi: 10.1177/003368829202300102, <http://dx.doi.org/10.1177/003368829202300102>.
- Granger, S. (Ed.). (1998). *Learner English on computer*. London and New York: Longman.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition, and foreign language teaching (Vol. 6)*. Amsterdam: John Benjamins.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*, (4th ed). London and New York: Routledge.
- Rezvani Kalajahi, S. A., Abdullah, A. N., Mukundan, J., & Tannacito, D. J. (2012). Discourse connectors: An overview of the history, definition and classification of the term. *World Applied Sciences Journal*, 19(11), 1659-1673. doi: 10.5829/idosi.wasj.2012.19.11.1990.
- Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33(4), 623-636. doi: 10.1177/2158244013506715, <http://dx.doi.org/10.1177/2158244013506715>.
- Matthiessen, C. M. I. M. (1995). *Lexicogrammatical cartography: English systems*. Tokyo: International Language Sciences Publishers.
- Milton, J. C., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: directions for future research. In Pemberton, R., & Tsang, E. S. C. (Eds.) *Studies in Lexis*. Hong Kong: The Hong Kong University of Science and Technology Language Center.
- Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). London and New York: Longman.