

A Corpus-based Analysis on Learners' Use of Inferential Discourse Markers in Their Argumentative Written English: A Case of Some Chinese Mainland University Students

Huai-kui Li

College of Foreign Studies
Guangxi Normal University
Guilin, China

Abstract:

*This paper, based on a corpus data analysis, examines the range, and extent of inferential discourse markers (IDMs), and their composite forms used by non-English major undergraduates (nEMUGs) and English major postgraduates (EMPGs) in their argumentative writing and problems remaining in the use of IDMs. The data first get processed in AntConc 3.5.9 and then analyzed manually. Conclusions are that, first, both nEMUGs and EMPGs, like NESs use most of the thirteen IDMs, nEMUGs using twelve, EMPGs, ten, and NESs, eleven. In the use of composite IDMs, except **and then, and then, and accordingly** employed by learners with some percentage, few composite IDMs are employed and the frequency is quite low. Second, both groups use **so** and **therefore** more often than other IDMs. As opposed to EMPGs, nEMUG group uses **so** much more frequently than **therefore**. Third, the problems are that nEMUGs make some semantic mistakes while EMPGs and NESs do not. And the range and the extent of composite IDMs used and the distribution and the degree an IDM used in the positions of S2 by both nEMUGs and EMPGs are different from those used by NESs. Therefore, both groups of learners are far from acquiring the native-like competence of English IDMs.*

Key words: Chinese EFL learners, corpus-based study, English IDMs, composite IDMs, problems in the IDMs use

1. Introduction

1.1 Background of the study

This paper aims to analyze inferential or implicative discourse markers (hereafter, IDMs) used by university learners of the Chinese Mainland in their written argumentative English as a foreign language. IDMs “signal that the force of the utterance is a conclusion which follows from the preceding discourse” (Fraser, 1996, p.188). As a part of discourse marks which are included in pragmatic markers, most IDMs may occur in sentence-initial position in the typical form of <discourse segment 1 (S1). IDM+ discourse segment 2(S2)> (Fraser, 1999, p. 938). Despite the importance of IDMs for effective argumentative writings, surprisingly little research has been conducted on what IDMs are used, and how they are distributed in learner English. The only ones available to pay attention to how learners use individual IDMs are House (2013) and Buysse (2012) on *so*.

It is an agreement that EFL learners tend to overuse or misuse discourse markers in conversation, thus leading to misunderstandings (Polat, 2011). Few studies, however, based on corpora of learner language, set out to analyze which IDM tends to be misused in what context or co-text. To the knowledge of the present author, the only research is the one done by Buysse (2012), and he finds that Belgium Dutch-speaking EFL learners overuse the IDM *so*. Do Chinese Mandarin-speaking EFL learners overuse or under-use IDMs? According to common sense, they may under-use them because Chinese Mandarin is a kind of paratactic language, in which sentences or clauses can be put together without using any connective like conjunction or adverb. Different from hypotactic languages like English, in which a conjunction or an adverb is used to link sentences or clauses to reveal their relation, Mandarin speakers may use fewer conjunctions or adverbs in utterances than English native speakers do. Whether this common sense is tenable and what misuse of them, including semantic errors and non-natural use which remain in learner English are issues deserving our concern.

Therefore, the present paper, based on Chinese EFL learners' data on argumentative written English, studies what IDMs are deployed, how they are distributed, and what problems remain in their use of these IDMs.

1.2 Studies of English IDMs: a review

1.2.1 Theoretical studies

There are four directions in the theoretical studies of English IDMs: what counts and does not count as an IDM, how they should be classified, the functions of individual IDMs, and the co-occurrence, the ordering, and the functions of composite IDM.

In one of his studies, Fraser (1996, p. 188) includes such items as IDMs: *accordingly, after all, all things considered, as a consequence, as a logical conclusion, as a result, because of this/that, consequently, for this/that reason, hence, in this/that case, it can be concluded that, it stands to reason that, of course, on this/that condition, so, then, therefore, thus*. In another study, however, Fraser's (1999, p. 948) counterpart is a bit different: *accordingly, all things considered, as a (logical) consequence/conclusion, as a result, because of this/that, consequently, hence, in any case, in this/that case, it can be concluded that, of course, on that condition, so, then, therefore, thus*. The markers *after all, for this/that reason, and it stands to reason that* in the previous class are omitted in his 1999 counterpart and *in any case* is added as an IDM. The most prominent change is to put *for this/that reason* into another minor nameless class. And even more radical is one of his studies in 2015, in which he gave another name, implicative discourse markers, to expressions most of which are similar to the previous group of IDMs: *so, therefore, thus, then, given that, as a result, as a consequence, consequently, as a conclusion, all in all, accordingly, hence, for that reason* (Fraser, 2015, p.49). The present paper is in line with Fraser's 2015 class of IDM, checking which of them are used and how often a certain IDM is used in the learners' written argumentative English.

In the issue of how IDMs should be organized into classes, Fraser (1999, p. 949) points out that there can be five types: (1) *so*, (2) *of course*, (3) *accordingly, as a consequence, as a logical conclusion, as a result, because of this/that, consequently, hence, it can be concluded that, therefore, thus*¹, (4) *in any case, in this/that case, under these/those condition, then*², (5) *all things considered*. The basis of the division is the subtlety of the S2 conclusion, but he does not make clear what such subtlety is. In one of his later studies, Fraser (2015, p. 51) holds that there are two major classes of IDMs, primary and secondary markers. The primary markers include *so*, and the secondary ones involve further three sub-classes: (1) *thus, hence, therefore, consequently*, (2) *then, given that*, (3) *as a result, as a consequence*. The rationale of the classification is the co-occurrence and ordering of composite IDMs, which will be reviewed as follows.

In the studies of the semantic functions of IDMs, the one most attracting attention is *so*. Schiffrin (1987) holds that it signals S2 being a result. Blakemore (1988) and Fraser (1996, 1999, 2015) argue that *so* marks inference. Bolden (2006, 2009) describes *so*, except for its role of marking inference, as having the function of launching sequences of new actions. That is to say, on the one hand, its use conveys to the hearer that the upcoming action is "emerging from incipency" and has already been on the agenda of interaction for some time (Bolden, 2006, p. 663). On the other hand, *so* is often used when a speaker has to deal with a problem in interaction typically arising whenever the current discourse segment does not occur as a sequence to the immediately preceding utterance and helps answer the question "why that now", thus telling the hearer to comprehend the following move as one belonging to a pending one (Bolden, 2009, p. 996). Raymond (2004) draws similar findings. He demonstrates that, based on its ability to introduce upshots of prior talk, it can be deployed as a stand-alone to prompt the hearer to produce the next relevant action. Johnson (2002) concludes that *so* can also be used to preface a question functioning as a topic developer. There has also been some recognition that *so* may play other roles in discourse as well. Buysse (2012, p. 1767), based on data from native and learner English, documents ten functions of *so*: (1) indicating a result, (2) drawing a conclusion, (3) prompting, (4) holding the floor, (5) introducing a summary, (6) introducing a section of the discourse, (7) indicating a shift back to a higher unit of the discourse, (8) introducing a new sequence, (9) introducing elaboration, and (10) marking self-correction. The first function is included in ideational relation, and the following three (functions (2) to (4)) are included in interpersonal relation, and the last six, textual relation. These studies remind the present author of the various meaning of *so*, helping exclude those having non-inferential uses from the corpus.

Despite a somewhat large literature on semantic studies of *so*, the functions of other IDMs receive far less attention. The only study available to the present author is Fraser (2015, p. 51). The IDMs *thus, hence, therefore, consequently* signal that there are some logical or rational grounds to conclude that S2 should be taken as true or relevant. With the IDMs *then, given that*, there are often two speakers and interpretation is that the speaker of S2 is drawing a conclusion, implying that "If S1, then S2 follows", as what is shown in the conversation: A: I want you to help with the dishes. B: *Then* why don't you bring me the soap? The IDMs *as a result, as a consequence* signal that S2 is a fact but not a conclusion or an inference derived from S1. The present study tries to check to what extent learners make mistakes in using IDMs so that they fail to realize the semantic function of signaling S2 being the conclusion, reason, or fact drawn from S1.

¹ Fraser included *for this/that reason* in his original third class. It seems that this may be a slip because he points out that the IDM is a member of another class, specifying that S2 provides a reason for the content presented in S1.

² The present author adds *in any case* to this fourth group while it is not included in Fraser's original counterpart.

In the studies of composite IDMs, one has to answer such theoretical questions as to what extent they can co-occur, what order is when they co-occur, and what is their semantic role when they get combined. As to the issues of co-occurrence and ordering, it is an intuition that the co-occurring behavior of IDMs is highly selective and that there can be a strong ordering preference. Except for the study of Koops & Lohmann (2015), Lohmann & Koops (2016), Crible (2017), and Haselow (2019) on co-occurrence and ordering of non-IDMs, Fraser (2015) and Cuenca & Crible (2019) have set out to deal with such issues of IDMs. Cuenca & Crible (2019), based on English corpora, conclude that co-occurrence of a non-IDM with an IDM is common. Most of the time conjunction comes as a first item, and among such co-occurrence, *and so*, *and then* are quite often used. The study most prominent can be Fraser (2015, p. 52), who points out that, within the category of IDM, *so* can be put together with most other IDMs and usually it must come before them. *Thus*, a member in the first subclass of the secondary IDMs, can be put together with other secondary ones and come before them. *Then*, a member in the second subclass of the secondary IDMs, can only be put before the one in the third subclass of IDMs. As to the question of semantic functions of the composite IDMs, the only study available to the present author is Koops & Lohmann (2022). They find that *and so* more often marks the upcoming utterance as a result or conclusion, similar to the IDM *therefore*. *So and*, however, is more often used to change the discourse topic. This paper studies the range and the extent composite IDMs are used in learner English rather than follows Fraser, Cuenca & Crible, and Koops & Lohmann, who researched co-occurrence, ordering, and meaning of IDMs in native English.

1.2.2 Empirical studies

There are two branches of the empirical studies of English IDMs. The first is the research on the classification, the meaning, the co-occurrence, and the ordering of IDMs based on native English corpora. The literature has been reviewed in the above section, and there is no need to repeat it here. The second branch is on the research of IDMs used in learner English. Despite various studies on non-IDMs (e.g. *you know* in Buysse (2017), *like* in Magliacane & Howard (2019), *like, you know, well* in Hellermann & Vergun (2007) and Polat (2011), *yeah, oh, you know, like, well, I mean, ok, right, and actually* in Liao (2009), to name only a few), the study on the use of IDMs in learner English is in its infancy. The only one available is Buysse (2012), who finds that, similar to native English speakers' language, the ten functions of *so* are found in the Dutch-speaking EFL learners' speech, but these learners show a higher incidence for its use than their English peers do. The learners' major also accounts for the different extent of such overuse. Those participants majoring in English Linguistics, regarded as learners with a higher English level, turn to *so* even more than those majoring in Commercial Sciences, learners with a lower English level.

1.3 Statement of the problem

If Fraser's (2009), and Liao's (2009) study, which searches for all possible markers from the definite corpora, are regarded as the macro research of discourse markers, those document uses of individual markers can be thought of as a micro one. The above review shows that most research studies IDMs from a micro perspective, many of which are based on the data of native English and fewer on those of learner English. The background of the participants is also quite limited, and the only one research is Buysse (2012), his participants being English learners from native Belgium speakers of Dutch. There is not any study of IDMs having Chinese Mandarin native speakers as the learner participants. Therefore, a macro study in this field can shed light on the research of IDMs in learner English.

Most studies on discourse markers of learner English collect qualitative data from a wide variety of sources. Among them, Ament, Pares & Perez-Vidal (2020) collect data from an English proficiency test, a monologue, and an interaction task, Magliacane & Howard (2019), from interviews of the research participants, Polat (2011), from an observational method, and Liao (2009), from both observational and interviewing method. All these data are learners' spoken speech. After having the raw materials, the researchers transcribe them and do the analysis manually or with the help of various kinds of computer software. The present paper is in line with Babanoğlu (2013), using learners' written language from argumentative essays as the data, and also in line with Polat (2011) and Magliacane & Howard (2019), analyzing what IDMs, which are delimited by Fraser (2015, p.49) as including *so, therefore, thus, then, given that, as a result, as a consequence, consequently, as a conclusion, all in all, accordingly, hence, for that reason*, are used by Chinese EFL learners and to what extent an IDM is used with the help of the concordancer, Ant Conc 3.5.9.

Buysse (2012), based on the cases of the IDM *so* from the learner and native English, asserts that learners overuse it in their spoken speech. Following him, the present paper, also taking native English speakers' language as a reference, intends to check what problems exist in Chinese EFL learners' use of IDMs.

2. Research methodology

2.1 Research questions

This paper addresses three research questions about Chinese EFL learners' performance in the use of IDMs in their argumentative writing:

What IDMs and to what extent an IDM is used by learners of two proficiency English levels?

What range and to what degree an IDM has a composite form or forms in learner English?

What problems remain in the use of IDMs?

The IDMs here refer to the thirteen markers mentioned by Fraser (2015). A composite form of an IDM includes the co-occurring of an IDM with another IDM or non-IDM. The two parts of a composite IDM need not necessarily be put together. *And accordingly* is regarded as a composite IDM in the sentence *If teachers keep attendance then they know who these students are and will punish them accordingly* (andaccordingly(NES)4) even though it is separated by the underlined part. The problems here refer to the mistaken and non-natural use of IDMs. Mistakes refer to semantic errors, in which a marker fails to achieve its inferential function. That is, it does not signal that S2 is the conclusion, fact, or reason based on S1. The non-natural use of IDM is defined to be a problem in three aspects: the over- or under-use of a certain IDM, the nonnative-like distribution and extent that a certain IDM is used compositely, and the nonnative-like range and degree that a certain IDM is used in the initial, medial or final position of S2. The standard for evaluating such problems is native English speakers' performance in the use IDMs in their written English.

2.2 Research design

This is a qualitative study with data being students' essays of English argumentative writing. The analysis of the data mainly involves documenting the two-level learners' use of English IDMs and the extent each IDM is used and analyzing the problems remaining in such use.

2.3 Research setting and population of the study

There are two research settings in this study. The first one is Guangxi Normal University (GXNU). The University, which was established in the 1920s, has been a leading center in teaching-learning, and research of humanities and sciences since its founding in the region of Guangxi Province, the Chinese Mainland. The second setting is the universities having an English major with MA postgraduates around the Chinese Mainland. It is not the fact that the present author went to various universities to collect data from the potential participants, but that he collected their MA thesis which was put open access on the website of *China National Knowledge Infrastructure* (CNKI).

There are accordingly two groups of the population. The first one is non-English major undergraduates (nEMUGs), who are formally taught English four class hours a week, two periods of reading and writing, and another two, listening and speaking, in their first and second academic years. The nEMUG participants of this study are more than 600 non-English major undergraduates of GXNU in their first and second academic years of college study (ages ranging from 19 to 20). These students are from three English classes (with an average size of 40 students each class) in each grade of 2012, 2014, 2016, 2018, and 2020 majoring in chemistry, physics, and electronics, a total of 15 classes of students to whom the present author taught the course of English reading and writing. They have finished writing more than 6000 essays, among which 4240 are argumentative ones. The second group of population is the English major postgraduates (EMPGs), who are taught English full time in all of their courses in the first two academic years, the third year being assigned to write the MA thesis. The EMPG participants are those 53 English major postgraduates from the grades 2015 to 2019 in their third academic year (average age of 25), and they come from universities around the Chinese Mainland with the orientation of research being linguistics.

2.4 Sample size and sampling technique

As for nEMUGs, all 600 students are taken as the research participants, and all 4148 argumentative essays they finished are taken as the data. And as for EMPGs, a convenient sampling technique is employed, 53 of them are taken from the section of the Chinese MA Theses Full-text Database (CMFD) in CNKI. The conclusion chapter of the 53 MA theses is the data.

2.5 Data collection

First is the collection of data on learner English. There are two sets of data being collected: short argumentative essays by nEMUGs, and the conclusion chapter of MA theses by EMPGs. MA theses are formal argumentative writing because the writers have to base their opinions on some fact, reason, and observation. Second is the collection of data by native English speakers (NESs), essays written by British and USA English speakers.

The average word length of each text written by nEMUG, EMPG, and native English speakers is respectively 197, 1181, and 412. The topics of nEMUG essays and MA theses are shown in Table 1³.

Table 1 Topics in nEMUG and EMPG essays

	Topics
Argumentative essays by nEMUGs	Impact of: Internet on English learning, AI on daily life; importance of: knowledge, patience, learning English by heart, professional training in school teaching, helping those in need, new technologies on daily life, e.g. communication in social Apps, pros and cons of online shopping; advantages and disadvantages of: fame, working in big cities, majoring in science, banning firecrackers, free charge of freeway during holidays, the in-peak holidays, the self-employment, employment in new types of job, having a second child; opinion on different attitudes toward job hunting; the solution to: needs, to plastic pollution; the ways to judge a person
MA theses by EMPGs	Political discourse and discourse analysis; teaching and acquisition of EFL; English language and linguistics; Chinese minority language; translation; comparison between English and Chinese and the localization of linguistic theories; multi-modal linguistics

The reason argumentative essays and the conclusion chapter of MA theses are used as data is that in such type of writing, writers can have more chance than in narratives, expositives, and descriptives to draw some conclusion, fact, or reason from their observations, thus having more opportunities to use IDMs. The nEMUG data are collected from a writing app. *Juku* (sentence-smith) with its website being <https://www.pigai.org/>, the EMPG data, from Chinese MA Theses Full-text Database (CMFD) in the website of <https://kns.cnki.net/>, and the native English data, the purpose of which is to take them as a reference to track the problems in the use of learner English, are from the corpus of native English speakers similarly or identically-prompted essays (NESSIE v2) from the website of <http://114.251.154.212/cqp/>. The nEMUG data are the writing tasks the present author assigned to the students he taught from the year 2012 to 2022. Ninety-two essays are eliminated because they get deviated from the writing directions or they are found to be plagiarized from other sources. For example, in the task of writing a passage titled *The advantages and disadvantages of fame*, some students copied parts of the text they just learned in their reading class, *The tail of fame*. In another task *The benefits and drawbacks of free charge of expressway during long holidays*, some essays are copies from the Internet because they are found to have the same wording. Therefore these deviated and plagiarized essays are deleted from the data. The final nEMUG data are the full text of 4148 English argumentative essays, having more than 820900 tokens of words. The EMPG data are the full text of the conclusion chapter of 53 MA theses of linguistics, having more than 62600 tokens of words. The native English data are the full-text essays from the corpus of NESSIE v2 which includes part of a corpus of British Academic Written English (BAWE), British National Corpus (BNC), Michigan Corpus of Upper-level Student Papers (MICUSP), and essays by native speakers of English in the written tasks of model College English Test (CET) and Test for English Majors (TEM). The NESSIE v2 corpus has 321768 words in 781 essays. The information of the three corpora is shown in Table 2. The segments with IDM tokens are extracted with the help of the software AntConc 3.5.9, and the NES data are directly extracted from the online corpus of NESSIE v2. Then according to the operational definition of IDM mentioned in the following section 2.5, the actual IDM data get manually analyzed.

Table 2 The size of NEMUG, EMPG, and NES corpus

	NEMUG (mid-low level)	EMPG (mid-high level)	NES
Number of essays	4148	53	781
Number of words	820900	62600	321768

2.6 Methods of data analysis

³ The native English essays are collected from the existing online corpus, and the present author gets no access to the information about their topics.

Descriptive statistics are employed to analyze the qualitative data. Firstly, the occurrence of each token of the markers concerned is documented. Secondly, the inferential use and the non-inferential use of each marker have to be distinguished manually. According to Fraser (1999, 2015), the criteria that a marker is an IDM are as follows:

(1) It is generally put at the beginning of S2, and in some few cases it is put at the end of S2, or is inserted in S2, as Fraser (1999, p. 938) remarks “almost all discourse markers occur in initial position (of a sentence), fewer occur in medial position and still fewer in final position.”

(2) An IDM does not contribute to the propositional meaning of both S1 and S2 (Fraser, 1999, p. 940). That is, it is not a modifier to any specific part of S1 or S2.

(3) S2 must be the conclusion, the reason, or the fact inferred from S1 (Fraser, 2015, p. 51).

(4) S1 and S2 must be separate messages (Fraser, 1999, p. 940). Specifically, S1 or S2 should not be subordinated to each other.

Therefore, the following use of a marker is not regarded as an IDM:

(1) *so*, *accordingly*, *thus*, and *then* modifying a verb, adjective, or adverb (e.g. *so do I*, *thus do sth.*, *accordingly do sth.*, *so tall/fast*, *so-called*, *thus far*, *just then*, etc.),

(2) *so* signaling S2 to be the summary of the previous segments of discourse (e.g. *So that is my view*),

(3) the idiomatic expressions of *so* and *then* (e.g. *so...that*, *so long as*, *if so*, *so as to*, *so far*, *even so*, *and so on*, *so on and so forth*, *from then on*, *if... then*),

(4) the possessive use of *as a result*, *as a conclusion*, *as a consequence*, and *for that reason* (e.g. *as a result/conclusion of sth.*, *for that reason of sth.*) as a modifier to a specific part of S1 or S2,

(5) *then* signaling time sequence in such sentences as *John shaved*, *then he listened to the radio*.

(6) any marker introducing any subordinated sentence fragment liking V-ing, V-ed, or infinite form.⁴

There are marginal cases, though:

(1) In the segments <S1. *So* many/much S2>, it is occasionally confusing whether *so* functions as IDM or non-IDM.

(2) Sometimes it is not easy to make clear whether *then* signals time sequence or IDM function between S1 and S2.

Under such conditions, the discourses previous S1 and the relation of S1 and S2 are closely checked. In the first case, if it can be decided that the segments are in the form of <S1. *So* [many/much S2]>, then *so* is of IDM use (e.g. *Stars will have no time to care of their family. So many of them will get marry late.*), and if in the form of <S1. [So many/much] S2>, it is of non-IDM use (e.g. *Now some fireworks is of not good quality and so many accidents and environmental problems there are to cause people worry.*) As to the use of *then* in the second case, if after a close check of co-text decision is still hard to make, then it is regrettably treated as non-IDM use. Take the sentences for example. *He turned the key. Then the engine started.* It is difficult to decide whether *then* is of time sequence or inferential use. Therefore it is not regarded as IDM data.

Thirdly, after documenting each case of IDMs, their frequency is computed. Such frequency is measured by the proportion of the total cases of a certain IDM in the total cases of all IDMs in the corpora. The criterion of counting a marker as a case of IDM is its actual occurrence in the discourses, and so there are two cases in a composite IDM. For example, the segment <S1. *So* *therefore*, S2> is counted as two cases of IDM use, *so* being one, and *therefore*, another.

The final step is the analysis of the semantic errors and the non-naturalness in the IDM use. The former refers to the failure of S2 to signal the conclusion, the fact, or the reason derived from S1. The degree of non-naturalness of IDM use is determined by (1) to what degree an IDM is over- or under-used by learners compared with the NES data, (2) what range and to what extent that learners can use composite IDMs like those used by native English speakers, (3) what range and to what degree learners can use a certain IDM in the initial, medial, or final position of S2 like those in which NESs use it. The criteria of whether S2 signals the conclusion, the fact, or the reason based on S1 is about the semantic relation between S1 and S2. In the segments, *We must make people realize the significance of knowledge, consequently, the most important thing is making people realize how vital the knowledge is, consequently* is not an IDM because S2 is just the repetition of S1.

⁴ But a marker introducing an elliptical sentence S2, with its subject or both subject and predicate derived from that of S1 can be an IDM. For example, *We are sociable animals and therefore need to have interpersonal relationships.* (andtherefore(NES)1) *It was hoped that this would reduce independent expenditure and thus overall spending.* (andthus(NES)11)

By computing to what the extent an IDM is used, to what range and to what degree a composite IDM is used, and how an IDM is distributed in the initial, medial or final position in S2 in both the learner and NES corpus, the naturalness of learner English can be determined. The standard of an initial, medial and final IDM is that, first, an initial IDM is put at the beginning of S2, even if it is an elliptical sentence like *60% of the electorate are opposed to the party in government and therefore have no political voice.* (andtherefore(NES)20) *It was hoped that this would reduce independent expenditure and thus overall spending.* (andthus(NES)11) Second, a medial IDM is the one being inserted in S2 and it is not a modifier to any part of S2. Third, a final IDM is the one being put at the end of S2. A separated composite IDM, which is combined by an IDM with a non-IDM, is defined to be an IDM. *And accordingly,* therefore, is regarded as a final-position IDM in the sentence *If teachers keep attendance then they know who these students are and will punish them accordingly* (andaccordingly(NES)4).

3. Results

3.1 IDMs used and the extent they are used in learner English

Of the thirteen IDMs, *so, therefore, thus, then, given that, as a result, as a consequence, consequently, as a conclusion, all in all, accordingly, hence, for that reason,* twelve, except *given that*, are employed by nEMUGs to various extent, with *so, therefore, thus, then,* and *accordingly* having the composite form of *and so, and therefore, and thus, and then, and accordingly*. The specific information about their distribution is shown in Table 3. Take some typical cases for example.

So:

(1) I will surf the internet to find more helpful information. *So* I decide to make more foreign friends whose mother tongue is English. (so(EMUG)2)

(2) We can discover it requires a lot of time, *so* in my opinion, we should work hard to meet the highest standards set by the course and to complete assignments on time. (so(nEMUG)19)

(3) Online shopping has been more and more popular, *and so* many people like online shopping. (andso(nEMUG)4)

Therefore:

(1) We need to know how to get along well with someone, *therefore*, I support to develop the social ability. (therefore(nEMUG)36)

(2) Fireworks do lots of harm to characters *therefore* we should not light it. (therefore(nEMUG)153)

(3) And our teachers always said that she will sample the condition of recitation *and therefore* we had to make our best to recite texts fluently. (andtherefore(nEMUG)1)

Thus:

(1) China is a country which pays more attention to the traditional history. *Thus*, I think we should continue to keep the custom — setting off firecrackers. (thus(nEMUG)14)

(2) Most family now have only one child, *thus* parents could pay more attention to their children and spend more time and energy on them. (thus(nEMUG)17)

(3) They are overwhelmed and thus suffer from psychological diseases.(andthus(nEMUG)6)

Then:

(1) I will spend more time in learning online, *then* I will try my best to improve my English on the Internet. (then(nEMUG)5)

(2) Many big fires are caused by firecrackers. *And then* some children are seriously hurt because of it. (andthen(nEMUG)10)

As a result:

(1) What's more, they can't do anything they want to do. *As a result*, they are losing their freedom. (asareult(nEMUG)10)

(2) I really need a brother or sister, *as a result*, I hope our family will add a new member. (asareult(nEMUG)21)

As a consequence:

(1) The more famous you are the more focus on you. *As a consequence*, fame may not be an ideal thing. (asaconsequence(nEMUG)2)

(2) If we study online we have to stare at the electronic screen for a long time, *as a consequence* it will do harm to our eyes. (asaconsequence(nEMUG)17)

Consequently:

(1) The main cause of this phenomenon is due to people's ignorance. *Consequently*, it is important to improve people's point that knowledge is useful. (consequently(nEMUG)10)

(2) We should protect the environment *consequently* I think that firecrackers should be banned. (consequently(nEMUG)20)

As a conclusion:

(1) No one can be tolerant to these things. *As a conclusion*, fame is good, but a trouble one. (asaconclusion(nEMUG)1)

All in all:

(1) We must keep the faith. *All in all*, I try my best to help the strangers who are in great need of help. (allinall(nEMUG)6)

(2) And it requires you to recite English. That is all, *all in all*, I think it is necessary to ingest by heart in the English study. (allinall(nEMUG)136)

Accordingly:

(1) It is a big problem in the society. *Accordingly*, we must try hard to acquire as much as we can. (accordingly(nEMUG)2)

(2) We encourage manufacturers to produce large quantities of recyclable shopping bags *and* give them subsidies *accordingly*. (andaccordingly(nEMUG)1)

Hence:

(1) It will be an indifferent social relationship, which is not beneficial to the progress of the society. *Hence* people are involved in a heated discussion about whether they should help the old man or not and how to help the people who are truly in need. (hence(nEMUG)3)

(2) There is one more a point, it will attract many people coming for a visit, hence it will result to heavy work load for the university management. (hence(nEMUG)1)

For that reason:

(1) I was born in the generation which influenced by the one-child policy. *For that reason*, I have neither brother nor sister. (forthatreason(nEMUG)1)

(2) Nowadays, less and less people need a job to save their lives, *for that reason* their attitudes toward hunting for jobs changed negatively. (forthatreason(nEMUG)7)

For EMPGs, as is shown in Table 3, ten IDMs, *so, therefore, thus, then, as a result, as a consequence, consequently, all in all, accordingly, and hence* are used in the corpus, only *therefore, thus* and *then* having the composite form of *and therefore, and thus, and then*. Take some typical segments for example.

(1) The corpus is selected initially from students' oral English test, *so* they may feel anxious to speak English. (so(EMPG)2)

(2) The reading efficiency of the high level group was 49.4608 and another group was 27.9254 with a significance level of 0.001. *So* their reading efficiency was significantly higher than another. (so(EMPG)53)

(3) Characters in equal social status have comparable education, manners and upbringing; *therefore*, their speech act will be influenced by these factors and tend to be similar. (therefore(EMPG)44)

(4) Most studies are closely related to the times, *and therefore* as a product of the times and cognition, the research results are inevitably bound and influenced by the times and the subject's cognition. (andtherefore(EMPG)1)

(5) Because the theoretical framework chosen for this study is the theory of crisis management, the relevant study in this field is not thorough enough, *thus*, the research depth is insufficient. (thus(EMPG)4)

(6) This will make the findings of future research more convincing and thus the findings could be applied into practice to a large extent. (andthus(EMPG)2)

(7) Therefore the substitution is employed flexibly on the basis of maintaining the coherence of the text. *Then*, in terms of ellipsis, the thesis finds that ellipsis is used less frequently than in Chinese for the purpose of avoiding redundant and achieving conciseness of the text. (then(EMPG)3)

(8) It is simple and easy for learners to remember, *and then* it reduces output errors significantly. (andthen(EMPG)1)

(9) In this thesis, only 135 Chinese college students' English spoken texts are selected in the corpus, which may still be insufficient as a shred of evidence to reveal the relationship between Chinese students' spoken features and the uses of x. *As a result*, the total number of x and x is not large enough to reach a more comprehensive and complete conclusion. (asareult(EMPG)1)

(10) These findings are corroborated by some research of figurative language processing that the lower the degree of salience of the final statement, the larger the N400 amplitude. *As a consequence*, the graded salience hypothesis is supported by the present study. (asaconsequence(EMPG)1)

(11) Theoretically, the findings of this study are in favor of graded salience hypothesis. *Consequently*, to a certain extent, the current study corroborates the psychological reality of salience. (consequently(EMPG)2)

(12) For the US, its personal information laws followed the principle of privacy protection in its common law system and advocated free business growth driven by data. *All in all*, this thesis used the three-dimensional analysis framework under the theory of CDA to discuss legal texts. (allinall(EMPG)1)

(13) Next, the analysis of examples is inevitably influenced by the author's limited competence. *Accordingly*, some suggestions for future research are provided. (accordingly(EMPG)2)

(14) This study just analyzes the similarities and differences of the two constructions. *Hence*, the changes and evolution of the constructions on span of the times cannot be discussed. (hence(EMPG)3)

Table 3 The distribution of the IDMs and their composite form in nEMUG, EMPG and NES corpus

IDM	form	nEMUG	EMPG	NES
<i>so</i>	—	2301/99.8%	59/100%	141/90.4%
	<i>and—</i>	5/0.2%	0	15/9.6%
<i>therefore</i>	—	645/99.8%	89/98.9%	150/77.4%
	<i>and—</i>	1/0.2%	1/1.1%	44/22.6%
<i>thus</i>	—	89 /92.7%	16/84.1%	70/66.7%
	<i>and—</i>	7 /7.3%	3/15.9%	35/33.3%
<i>then</i>	—	124/81%	6/75 %	29/82.9%
	<i>and—</i>	29/19%	2/25%	2/5.7%
	<i>but—</i>	0	0	2/5.7%
	<i>well—</i>	0	0	2/5.7%
<i>given that</i>	—	0	0	9/90%
	<i>furthermore—</i>	0	0	1/10%
<i>as a result</i>	—	71/100%	4/100%	31/83.8%
	<i>and—</i>	0	0	6/16.2%
<i>as a consequence</i>	—	18/100%	2/100%	1/100%
<i>consequently</i>	—	62/100%	3/100%	2/28.6%
	<i>and—</i>	0	0	4/57.1 %
	<i>—though</i>	0	0	1/14.3%
<i>as a conclusion</i>	—	1/100%	0	0
<i>all in all</i>	—	173/100%	2/100%	2/100%
<i>accordingly</i>	—	5/83.3%	4/100%	3/37.5%
	<i>and—</i>	1/16.7%	0	5/62.5%
<i>hence</i>	—	4/100%	3/100%	15/68.2%
	<i>and—</i>	0	0	7/31.8%
<i>for that reason</i>	—	8/100%	0	0

Note: In the second column, the dash refers to the related IDM itself and the one behind or before the italicized word represents the related IDM itself. For instance, in the line of *so*, *and—* means *and so*.

Table 3 shows that the nEMUG's and EMPG's corpora share some similarities. Firstly, most IDMs are used by the two groups of participants to various extents. Secondly, most IDMs are used non-compositely. For the twelve IDMs used by nEMUGs, although five IDMs are used compositely, the composite form of *so* and *therefore* occurs in quite low frequency (both lower than 1%). Therefore, the composite use of the IDM in the nEMUG data accounts for 25%, and ten IDMs used EMPGs, three are used compositely, accounting for 30%. Such tendency is somewhat similar. In what follows let's present the extent of IDMs used by learners of both groups.

Table 4 shows that from the frequency of each marker there are three groups of IDMs in the nEMUGs corpus. The first one includes *so*, which is most frequently used, and the second includes *therefore*, which is much less often used than *so* (18.2% versus 65.1%). The third group includes the rest ten IDMs, defined to be the least used markers, or markers with a low probability of use (lower than 5%), in which there can be three sub-groups: *all in all* and *then* can be put together and regarded as IDMs with low frequency because their respective frequency ranges from 4.9% to 4.3%. The next sub-group includes *consequently*, *as a result*, and *thus*. They can be taken as IDMs with quite low frequency since their percentage ranges from 2.7% to 1.7%.

The third sub-group includes *as a consequence*, *hence*, *for that reason*, *accordingly*, and *as a conclusion*. They can be regarded as IDMs with extremely low frequency of use because their percentage is all far lower than 1%, nearly negligible. For the EMPGs data, there are also three groups of IDMs: the most frequently used *therefore* (46.4%), the less frequently used *so* (30.4%), and the other eight IDMs, the least used markers. These least frequently used markers can be further divided into three sub-groups. *Thus* is in the first sub-group, and its frequency is 9.8%. And the second sub-group includes *then*, with its frequency being 4.1%. The rest six markers, *as a result*, *as a consequence*, *consequently*, *all in all*, *accordingly*, and *hence*, can be put in one group because the frequency of the cases in which they are used ranges from 1% to 2.1%. The difference between the two groups' use of IDMs is that, first, the distribution of IDMs in the nEMUG data tends to be polarized and the first two most frequently used markers, *so* and *therefore* account for respectively 65.1%, and 18.2% of the total markers, and their discrepancy is quite obvious. The IDMs used in the EMPG data, however, are relatively more evenly distributed, and the first two most frequently used markers, *therefore* and *so* account for 46.4% and 30.4% of the total markers, and the difference is not that prominent. Second, as the use of IDMs is quite unevenly distributed in the nEMUG data, there are five IDMs the frequency of which is far lower than 1%, while the frequency of those used by EMPGs is all at or above the level of 1%. If the borderline of actual use of IDMs is defined to be 1% in the frequency of use, then there are only seven IDMs used by nEMUGs, fewer than the ten ones used by EMPGs. Third, to make the IDM use in the nEMUG and EMPG data comparable, the frequency of each IDM is transformed into the expected frequency per million-word tokens. Table 5 shows that the two groups are quite significantly different in the extent of using *so*, *therefore*, *thus*, *then*, *consequently*, *all in all*, *accordingly*, *hence* and *for that reason* because their probability values from the chi-square test are all lower than the significant level of 0.05 or 0.01 ($df=1$). And nEMUGs use *so*, *then*, *consequently*, *all in all*, and *for that reason* to a much greater extent than EMPGs. The latter group, however, uses *therefore*, *thus*, *accordingly*, and *hence* to a higher degree.

Table 4 The distribution of the thirteen IDMs in nEMUG, EMPG and NES corpus

IDM	nEMUG		EMPG		NES	
	tokens	frequency	tokens	frequency	tokens	frequency
<i>so</i>	2306	65.1%	59	30.4%	156	27%
<i>therefore</i>	646	18.2%	90	46.4%	194	33.6%
<i>thus</i>	96	2.7%	19	9.8%	105	18.2%
<i>then</i>	153	4.3%	8	4.1%	35	6.1%
<i>given that</i>	0	0	0	0	10	1.7%
<i>as a result</i>	71	2%	4	2.1%	37	6.4%
<i>as a consequence</i>	18	0.5%	2	1%	1	0.2%
<i>consequently</i>	62	1.7%	3	1.5%	7	1.2%
<i>as a conclusion</i>	1	0.02%	0	0	0	0
<i>all in all</i>	173	4.9%	2	1%	2	0.3%
<i>accordingly</i>	6	0.2%	4	2.1%	8	1.4%
<i>hence</i>	4	0.1%	3	1.5%	22	3.8%
<i>for that reason</i>	8	0.2%	0	0	0	0
sum	3544	100%	194	100%	577	100%

Table 5 The distribution of the thirteen IDMs in per million word tokens in the three corpora

IDM	nEMUG	EMPG	NES	χ^2
<i>so</i>	2809	942	485	928.3**
<i>therefore</i>	787	1438	603	189.9**
<i>thus</i>	117	304	326	82.2**
<i>then</i>	186	127	109	10.7**
<i>given that</i>	0	0	31	0
<i>as a result</i>	86	64	115	2.9
<i>as a consequence</i>	22	32	3	1.5
<i>consequently</i>	76	48	22	5.9*
<i>as a conclusion</i>	1	0	0	1
<i>all in all</i>	211	32	6	130.4**
<i>accordingly</i>	7	64	25	44.2**
<i>hence</i>	5	48	68	33.3**
<i>for that reason</i>	10	0	0	10**

Note: The chi-square value in the last column is computed from the nEMUG and EMPG data. The value with single asterisks means $p < 0.05$, with double asterisks, $p < 0.01$. The result of the expected frequency of each IDM is computed from the formula: $EF_{IDM} = \frac{OF_{IDM} \cdot OS_{cor}}{N}$, in which OF_{IDM} refers to the observed frequency of a certain IDM, and OS_{cor} to the observed size of the corpus. For example, the observed frequency of *so* is 2306, and the observed size of nEMUG corpus is 820900. Therefore, the $EF_{so} = \frac{1000000 \cdot 2306}{820900} = 2809$.

3.2 Range and degree of composite IDMs used in learner English

Table 3 shows that both nEMUGs and EMPGs use composite IDMs in a quite limited range. In the twelve IDMs deployed by nEMUGs, only five have composite form, *and so*, *and therefore*, *and thus*, *and then*, and *accordingly*. And in the ten IDMs used by EMPGs, only three are used compositely, *and therefore*, *and then*, and *and thus*. As to the extent that composite IDMs are used, nEMUGs are more liable to use *and then*, and *accordingly* (19% versus 16.7%), then is *and thus*, accounting for 7.3% in the cases of *thus*. The percentage of *and so*, and *and therefore* is so low that they can be negligible (both 0.2%). For EMPGs, except obvious liable used ones *and then*, and *and thus* (25% versus 15.9%), there is only one case in *and therefore*, accounting for 1.1% in the cases of *therefore*. The only common shared by both groups of nEMUGs and EMPGs is that they use *and then* to a similar extent (19% versus 25%). The other composite ones are different both in the range and the extent that they are used.

Therefore the above results reveal that the mid-low level and the mid-high level learners are different in their competence in using some IDMs of English.

3.3 Problems in participants' use of IDMs

In what follows, semantic mistakes and to what extent the learner's use of IDM is native-like are presented. Firstly, from a semantic perspective, an IDM signals that S2 is the conclusion, the reason, or the fact inferred from S1, and it "does not contribute to the propositional meaning of either segment" (Fraser, 1999, p.944). Some nEMUGs, however, make semantic mistakes, as is shown in the following examples (to name only a few):

(1) I hold the view that the Internet has created some serious problems in my English study. *Therefore*, I rely more on the Internet in my English study.

(2) I can participate in conversation which is made in the classroom though I still make mistakes in speaking English. *Then* there are many ways to learn English.

(3) People not only attach importance to knowledge but also to ability because of the fierce competition. *As a result of that*, there are 33% people who don't think knowledge means power at all.

(4) We must make people realize the significance of knowledge, *consequently*, the most important thing is making people realize how vital the knowledge is.

There is a contradiction between S1 and S2 in example (1). If *therefore* still has to have the function of an IDM here, S2 should be reworded as *online learning of English is not beneficial for me*, or, *I will not learn English on the Internet*. S2 in example (2) is not the ground from S1. Therefore, it can be reformed as *I prefer learning English face to face with my teacher and classmates*. S1 and S2 are contradictory in example (3). S2 should be reworded as *they tend to acquire knowledge, say, by doing some reading in the newspaper, or on their cell phone in their daily life*. S2 in

example (4) to some extent is just a repetition of S1. Its reformulation can be: *they can show more respect for school education and invest more energy to learn something in their spare time*. EMPGs, however, like NESs, make no such semantics mistakes. In this aspect, it seems that great length and more exposure to English learning make a difference in the development of IDM competence.

Secondly, let's analyze the native-like use of IDMs in learner English from three perspectives: over- or under-use of IDMs, how learners, like what NESs usually do, can vary their use of composite IDMs, and vary their use of an IDM in different positions of S2 like what NESs can do. Table 5 shows that both groups of learners use more *so*, *therefore*, *then*, *as a consequence*, *consequently*, and *all in all* than NESs do. On the contrary, learners use fewer *thus*, *given that*, *as a result*, and *hence* than NESs do. And nEMUGs use more *then* and *for that reason* than both EMPGs and NESs do, but EMPGs use *accordingly* more frequently than both nEMUGs and NESs do.

Therefore, one should not simply claim that Chinese EFL learners do under- or over-use English IDMs. It is better to assert that in some IDMs learners overuse them while in some others they under-use them. If the factors like English proficiency and the genre of the essays are taken into account, the situation can be more complicated.

Table 3 shows that of the eleven IDMs, the NESs deploy composite markers in nine. Except for the use of *and* + IDM, they also have *but*, *well*, *furthermore* + IDM, or IDM + *though* (e.g. *but then*, *well then*, *furthermore given that*, and *consequently though*). Except for the relatively low frequency of the composite use of *so* (9.6%), the degree in the rest of the composite forms is all above 15%⁵, and the composite use in *consequently*, *accordingly* even more prevails in their respective non-composite one (71.4% versus 28.6% in *consequently*, and 62.5% versus 37.5% in *accordingly*). NEMUGs and EMPGs, however, deploy fewer cases of composite IDMs, with the former using five, *and so*, *and therefore*, *and thus*, *and then*, and *and accordingly*, and the latter, three, *and therefore*, *and thus*, and *and then*. Therefore, the range is far less than that of NESs, and except for *then*, *accordingly* and *thus* with the frequency ranging from 7.3% to 25%, the degree of composite use in learner English is so low that it can be overlooked.

As is shown in Table 6, nEMUGs, EMPGs and NESs are different in six IDMs in the positional use in S2⁶. Nearly all the learners use the IDMs in the initial position of S2 while native English speakers show a variety of using these IDMs in different positions of S2. Specifically, there are initial and medial positional uses in *therefore*, *thus*, and *then*, and initial and final positional uses in *as a result*. *Accordingly* and *hence* have all initial, medial, and final uses.

Therefore learners, both mid-low and mid-high level ones, have not acquired the native-like range of composite IDMs, and they also can not vary the IDMs in different positions of S2. From this respect, one can tentatively assert that EFL learners have a long way to go to acquire the native-like competence of English IDMs.

Table 6 The information of position where six IDMs are put in S2 in nEMUG, EMPG, and NES corpus

IDM	nEMUGs			EMPGs			NESs		
	I	M	F	I	M	F	I	M	F
therefore	646/100%	0	0	90/100%	0	0	140/72.2%	54/27.8%	0
thus	71/100%	0	0	19/100%	0	0	96/91.4%	9/8.6%	0
then	125/100%	0	0	4/100%	0	0	20/57.1%	15/42.9%	0
as a result	68/100%	0	0	4/100%	0	0	34/91.9%	0	3/8.1%
accordingly	5/83.3%	0	1/16.7%	4/100%	0	0	4/50%	3/37.5%	1/12.5%
hence	4/100%	0	0	3/100%	0	0	19/86.4%	1/4.5%	2/9.1%

Note: The sign I in the second line refers to the initial position, M to the medial position, and F to the final position of S2.

4. Discussions

This section covers on the one hand the theoretical implication for the studies of IDMs and the practical implication for the EFL teaching of IDMs on the other hand. NEMUGs use *so* quite more frequently than EMPGs. Such a finding is contradictory to that of Buysse (2012), who documents that learners with higher English level more tend to use *so*. The discrepancy may be due to the different sources from which the data are collected. Compared with the data of written

⁵ The frequency of the composite use of *then* is computed from the total frequency of its three composite forms, *and then*, *but then* and *well then*. Thus it is 17.1% in the use of *then*. So is the case in *consequently*, its frequency of composite form is contributed by *and consequently* and *consequently though*, with the total percentage being 71.4%.

⁶ Some of the other seven IDMs are either all put in the initial position of S2 in the three sets of corpora or some of them are not used by the participants in their writing.

language in the present research, Buysse's corpora are spoken English from informal interviews. In spontaneous speech, more proficient learners may use particles such as *so* to enable themselves to keep the flow of the speech. The higher English level participants' corpora in this research are serious and formal written language. It is natural for them to use the informal IDM *so* a much less extent. That is, difference in the language style can explain the participants' various extents in the use of IDMs. Such generalization seems to account for the use of some other IDMs. As is shown in Table 5, *therefore*, *thus*, *accordingly* and *hence* are significantly more used in the EMPG corpus because they are of formal use. *So*, *then*, *consequently*, *all in all*, and *for that reason* are somewhat informal when used as an IDM, and therefore they are more used in nEMUGs' argumentative essays. Future researchers in this field, when comparing discourse markers used by different levels of learners, should pay attention to the comparability of data, including that of genre, topic, and formality of the writing tasks.

Next is the discussion on the practical implications of the present research. There has been an agreement in the study of EFL teaching and learning that the accuracy of the L2 itself does not fulfill the native-like competence of English. One has to embrace the ability of appropriateness. But researchers are divided as to what constitutes the contents of appropriateness in the actual practice of EFL instruction. Liao (2009, p. 1313) holds that to sound more like a native speaker of English, a learner in one way should adopt the "conventional expressions" (e.g. the discourse markers) used by the English native speakers. But in the EFL environment, learners do not acquire conventional knowledge "for free" (Kasper & Rose, 2006, p. 2). As is shown in the present research, some nEMUGs make mistakes in the use of many IDMs in their argumentative essays, and both nEMUGs and EMPGs do not acquire the native-like competence in the range and extent of using English IDMs. Therefore, the intervention of IDMs in EFL teaching seems to be necessary. If this is right, then the pedagogical issues such as what to teach and how to teach IDMs have to be considered. In the question of what to teach IDMs, the pedagogy researchers first have to delimit classes of English IDMs, and then based on the empirical findings of the acquisition sequence of individual IDMs, the relation of their acquisition with the grammatical and semantic development of other language elements, one can decide such issues as (1) whether to teach them specifically in some periods of class hours or merge them into the teaching of other language elements, (2) the orders of presenting them in the actual teaching practice, (3) the length of class time spent on the teaching of each item of IDM. People are also divided on how to teach IDMs. Ellis (2002), as opposed to the approach of focus on forms (FonFs), puts forward focus-on-form (FonF), in which there are implicit and explicit FonF (also see Tateyama, 2006, Yoshimi, 2006, and inductive and deductive technique of language instruction by Rose & Kwai-fun, 2006). The sense of implicitness and explicitness is attached to the feature of feedback. In the implicit FonF, the instructor can give feedback by a recast. The feedback in explicit FonF can be performed in these ways, (1) by directly signaling that the student has made an error, (2) by using meta-language to indicate what wrong the student has made, (3) by providing a correction and a practice on the use of the correct form. However, the explicitness and implicitness in feedback is more a question of degree than a dichotomy of yes or no. In the EFL teaching of IDMs, to what extent corrective feedback should be made explicit needs further studies.

5. Conclusions and suggestions

Based on the corpora constructed by AntConc 3.5.9 from argumentative writing of nEMUG and EMPG and the online corpus of NES data, this research has shown that first, like the NES participants, most of the thirteen IDMs delimited by Fraser (2015) are used by both groups of participants to various extent, the former using twelve of them, the latter, ten. In the use of composite IDMs, except *and then*, *and then*, and *accordingly* employed by learners in some percentage, few composite IDMs are used and the extent of their use is quite low. Second, the most frequently used IDM by nEMUGs is *so* (2306/3544), and the one that comes next is *therefore* (646/3544). The most frequently used IDM by EMPGs, however, is *therefore* (90/194), and the one that ranks next is *so* (59/194). The other IDMs can be regarded as low-frequency markers because the percentage of each marker being used is not more than 10%. The occurrence of *as a consequence*, *as a conclusion*, *accordingly*, *hence*, and *for that reason* in the nEMUG corpus is all lower than 1%, nearly negligible. Compared with EMPGs, nEMUGs use *so*, *then*, *consequently*, *all in all*, and *for that reason* to a greater extent. EMPGs, however, use *therefore*, *thus*, *accordingly*, and *hence* to a higher degree. Although the participants employ a wide range of IDMs, the mid-low level participants make some semantic mistakes in such use. Specifically, in some segments with *therefore*, *then*, *as a result*, and *consequently*, S2 does not signal the conclusion, the fact, or the reason drawn from S1. However, compared with NES data, from the perspectives of the over- or under-use of IDMs, the range and the extent of composite IDMs used, and the distribution and the degree of an IDM used in different positions of S2, both groups of nEMUGs and EMPGs are far from acquiring native-like competence of English IDMs.

There are some potential problems in this study that require caution in making any claims of any kind. The validity of the results depends on the fact that, first, the corpora are actual data from the participants. Second, the data analysis is valid, and third, the data from both learners and native English speakers are comparable. As to the first issue, although

the app. *Zuku* has the function of checking plagiarism, the present author does find that about 90 essays of the nEMUG data are the copy from other sources, thus deleted in the final corpus. However, the present author himself is not fully confident about the reliability of the data. The same is true for the EMPG data. It is impossible to know to what extent the authors' supervisor plays a part in writing the conclusion chapter of the MA theses. If it should be the supervisor's partial or total wording, then the validity of the data is heavily contaminated. In collecting and analyzing the data, the present author is cautious and repeatedly research reads the materials so that all the data can be extracted based on the operational definition of IDMs. But given the size of the corpora, especially the nEMUG and the NES data, there may be an overlook of any kind. Finally, the nEMUG data are from informal argumentative writing on various aspects of campus life and those social issues that university students should be concerned with. However, the EMPG data are from formal and serious argumentative writing on issues related to linguistics and the English language. The two sets of corpora are sufficient for this study, but it is clear that they are not so much comparable. Therefore, future studies should be cautious about the conclusion drawn from this study.

Due to the limitations mentioned above, future studies can, first, collect parallel data (that is, spoken speech or written language of English from similar speech events or topics) from both Chinese English learners and native English speakers to check what IDMs and to what extent the individual IDMs are used. Second, researchers can design a longitudinal study and do classroom observation to see the developmental features of IDMs. Third, they may develop a certain instrument to collect data in a survey to check what factors can be correlated with the competence of the IDM use. Finally, one can design an experimental study to check to what extent an instruction can affect learners' use of one or more IDMs, or to what degree such instruction can enhance their awareness of the semantic function of individual IDMs.

References

- Ament, J., Pares, J. B. & Perez-Vidal, C. 2020. A study on the functional uses of textual pragmatic markers by native speakers and English-medium instruction learners. *Journal of Pragmatics*, 156, 41-53.
- Babanoğlu, M. P. 2013. A corpus-based study on the use of pragmatic markers as speech like features in Turkish EFL learners' argumentative essays. *Procedia - Social and Behavioral Sciences*, 136, 186-193.
- Blakemore, D. 1988. 'So' as a constraint on relevance. In: Kempson, R. M. (Ed.) *Mental Representations: The Interface Between Language and Reality*. New York: Cambridge University Press, 183-195.
- Bolden, G. B. 2006. Little words that matter: discourse markers "so" and "oh" and the doing of other-attentiveness in social interaction. *Journal of Communication*, 56, 661-688.
- Bolden, G. B. 2009. Implementing incipient actions: the discourse marker "so" in English conversation. *Journal of Pragmatics*, 41, 974-998.
- Buysse, L. 2012. *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44, 1764-1782.
- Buysse, L. 2017. The pragmatic marker *you know* in learner Englishes. *Journal of Pragmatics*, 121, 40-57.
- Crible, L. 2017. Discourse markers and (dis)fluencies in English and French: variation and combination in the DisFrEn Corpus. *Int. J. Corpus Linguist*, 22 (2), 242-269.
- Cuenca, M. J. & Crible, L. 2019. Co-occurrence of discourse markers in English: from juxtaposition to composition. *Journal of Pragmatics*, 140, 171-184.
- Ellis, R. 2002. Doing focus-on-form. *System*, 30, 419-432.
- Fraser, B. 1996. Pragmatic marker. *Pragmatics*, 6(2), 167-190.
- Fraser, B. 1999. What are discourse markers. *Journal of Pragmatics*, 31, 931-952.
- Fraser, B. 2009. Topic orientation markers. *Journal of Pragmatics*, 41, 892-898.
- Fraser, B. 2015. The combining of discourse markers — a beginning. *Journal of Pragmatics*, 86, 48-53.
- Haselow, A. 2019. Discourse marker sequences: insights into the serial order of communicative tasks in real-time turn production. *J. Pragmat*, 146, 1-18.
- Hellermann, J. & Vergun, A. 2007. Language which is not taught: the discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39, 157-179.
- House, J. 2013. Developing pragmatic competence in English as a lingua franca: using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics*, 59, 57-67.
- Kasper, G & Rose, K. R. 2006. Pragmatics in language teaching. In: Rose, K. R. & Kasper, G.(Eds.) *Pragmatics in Language Teaching*. Beijing: World Publishing Corporation, 1-9.

- Koops, C. & Lohmann, A. 2022. Explaining reversible discourse marker sequences: a case study of *and* and *so*. *Journal of Pragmatics*, 191, 156-171.
- Koops, C. & Lohmann, A. 2015. A quantitative approach to the grammaticalization of discourse markers: evidence from their sequencing behavior. *Int. J. Corpus Linguist*, 20 (2), 232-259.
- Liao, S. 2009. Variation in the use of discourse markers by Chinese teaching assistants in the US. *Journal of Pragmatics*, 41, 1313-1328.
- Lohmann, A. & Koops, C. 2016. Aspects of discourse marker sequencing: empirical challenges and theoretical implications. In: Kaltenbock, G., Keizer, E. & Lohmann, A. (Eds.) *Outside the Clause: Form and Function of Extra-Clausal Constituents*. Amsterdam: Benjamins, 417-445.
- Magliacane, A. & Howard, M. 2019. The role of learner status in the acquisition of pragmatic markers during study abroad: the use of 'like' in L2 English. *Journal of Pragmatics*, 146, 72-86.
- Polat, B. 2011. Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics*, 43, 3745-3756.
- Raymond, G. 2004. Prompting action: the stand-alone "so" in ordinary conversation. *Research on Language and Social Interaction*, 37, 185-218.
- Rose, K. R. & Kwai-fun, C. N. 2006. Inductive and deductive teaching of compliments and compliment responses. In: Rose, K. R. & Kasper, G.(Eds.) *Pragmatics in Language Teaching*. Beijing: World Publishing Corporation, 145-170.
- Schiffrin, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Tateyama, Y. 2006. Explicit and implicit teaching of pragmatic routines: Japanese *sumimasen*. In: Rose, K. R. & Kasper, G.(Eds.) *Pragmatics in Language Teaching*. Beijing: World Publishing Corporation, 200-222.
- Yoshimi, D. R. 2006. Explicit instruction and JFL learner's use of interactional discourse markers. In: Rose, K. R. & Kasper, G. (Eds.) *Pragmatics in Language Teaching*. Beijing: World Publishing Corporation, 223 -244.

Huai-Kui Li is professor of Linguistics and Applied Linguistics at Guangxi Normal University. His research interests include studies of pragmatics and second language pragmatics.