

Inferential Discourse Markers in Contrast: A Corpus-based Comparative Study

Lijiao Chen¹, Ling Zhao²

¹ Weiqiaoguo Education Group, Binzhou 256600, China

² School of Foreign Languages and Literature, Wuhan University, Wuhan 256600, China

Abstract

Discourse markers are an indispensable part of any coherent discourse. The present research seeks to investigate the difference in the use of inferential discourse markers in English writing between Chinese EFL learners and native speakers based on corpus data. The materials used are taken from TECCLv1.1 and NESSIEv2. Quantitative analysis suggests that Chinese English learners use slightly more inferential discourse markers than native English speakers in general in written English and tend to overuse certain contrast words while undersuse some others. Possible causes are discussed and some suggestions are offered for the teaching and learning of English in China.

Keywords

Inferential discourse marker; Contrast; Pragmatic competence; Corpus; Chinese EFL learners.

1. Introduction

The diverse names of discourse markers have been confusing—the term discourse markers coined by Deborah Schiffrin has been used interchangeably with *prefaces* [1], *pragmatic formatives/markers* [2], *discourse connectives* [3], *discourse operators* [4], *cue pragmatic expressions* [5], *cue phrases* [6], *cue words* [7], *discourse particles* [8], and many others. Subtle differences do exist, for example, the term discourse connectives is usually used when attention is paid to the role discourse markers play in enhancing the overall coherency of the whole discourse, whereas *pragmatic markers* is often employed when discourse markers organize the discourse and mark the turns [9]. Nowadays, most linguists adopt discourse markers as a neutral, all-inclusive term. According to Schiffrin, discourse markers are sequentially dependent elements which bracket units of talk [10]. Therefore, they can be words such as adverbs (actually), conjunctions (and), exclamations (well), or phrases (you know) as long as they fulfill procedural functions without conveying any actual propositional meaning like change the true-false value.

There are many ways to classify discourse markers. Since this study intends to focus on the cause-effect relationship between utterances, we adopt Fraser's theory which defines a discourse marker as a lexical expression (LE) that signals a semantic relationship of elaboration, contrast, inference or temporality [11]. Correspondingly, discourse markers fall into four categories, i.e. elaborative discourse markers, contrastive discourse markers, inferential discourse markers, as well as temporal discourse markers. This study will focus on one of the subcategories, that is, inferential discourse markers. As the name suggests, they are the words or expressions that signal a causal relationship between utterances. The common inferential discourse markers in modern English include *so*, *therefore*, *then*, *thus*, *as a result*, *consequently*, etc.

Discourse markers are frequently used in our daily communication, in fact they are so common that they are often taken for granted. However, the littlest things give us away when we try to

speak or write in a foreign language. According to Aijmer, a major difference between second language learners and native speakers lies in the way they use discourse markers [12]. In light of this, a series of studies on second language acquisition have been conducted to compare native and non-native speakers' usage of discourse markers. However, most previous studies have focused on the discourse markers as a whole while few have made the inferential discourse markers in contrast, a specific subcategory, as their main focus. To narrow the research gap, this paper intends to compare the usage of inferential discourse marker between Chinese EFL learners and native English speakers based on corpus data.

2. Research Questions

Based on previous studies, the present research presupposes that Chinese English learners and native speakers differ in the use of inferential discourse markers. This study seeks to investigate the differences between the use of these inferential discourse markers by Chinese EFL learners and English native speakers, and specifically, the following questions were addressed:

Do Chinese EFL learners use more or less inferential discourse markers than native speakers in English writing tasks?

Do Chinese EFL learners overuse or underuse certain inferential ?

Do Chinese EFL learners put inferential discourse markers in the same syntactic positions as native speakers?

3. Method

3.1. Data

Data were taken from TECCLv1.1 and NESSIEv2. TECCL (Ten-thousand English Compositions of Chinese Learners) was developed by Prof. Xu at Beijing Foreign Studies University in 2015[13]. It is a collection of ten-thousand compositions to represent a wide cross-section of Chinese learners' written English. There are altogether 1,990,258 words in 9,864 texts that come from a range of sources, including 4 levels—primary school, middle school, university and other. The compositions are completed during 2010 and 2015 in 32 provinces, municipalities and special administrative regions (except for Tibet and Macao). NESSIEv2 (Native English Speakers Similarly or Identically-prompted Essays Corpus 2nd release) was also developed by Prof. Xu in 2013[14]. It is a 321,768 word collection of essays written by native English speakers. It recognizes 5 sources, namely BAWE, BNC ESSAYS, CET topics by NS, MICUSP and TEM topics by NS.

The two corpora were selected over many others because they both cover a large amount of English compositions from a wide range of sources, and compared to many well-used corpora, they are relatively new and more representative of Chinese language learners. To make the two corpora comparable, a subcorpora was built and this mini-corpora had 314,280 words, close to NESSIE's 321,768 words.

3.2. Materials

Fraser divided inferential discourse markers in English into 4 groups according to their occurrence, including a) so; b) then, in that case; c) thus, therefore, hence, consequently; d) as a result of, for that reason, as a consequence, that's why, as a result[11]. The twelve inferential discourse markers mentioned above are examined in this study.

Since some of the inferential discourse markers are polysemous, they sometimes do not function as discourse markers, such as *so* in

(1) Besides many are tended to be addicted to computer games and *so*(TECCL00007).

(2)I think the death penalty is really a step back in such a world which weighs so much on human rights (TECCL00015).

(3)I believe that existence is truth, so is the death penalty (TECCL00023).

(4)What's more, people's life were so worthless in some places (TECCL 00031). and then in

(1)Since then, general contend about this penalty has never concluded (TECCL00015).

(2)From then on , the puppy named Duck lived with Bob for a long time (TECCL07878).

For this reason, the data were screened manually to get the net frequency.

3.3. Instrumentation

BFSU CQPweb is a web-based fourth generation corpus analysis tool [15] while the mainstream software like WordSmith and AntConc are third generation tools. The major advantage of a web-based analysis tool is that it saves the trouble of downloading corpus data and everything can be done simply with an Internet access.

SPSS 22.0 was used to run the Chi-square test to determine whether there is a significant difference between the expected frequencies and the observed frequencies, in this case, whether the differences between the two group's use of inferential discourse markers are significant or not.

4. Results

Results of word frequency were displayed in Table 1. Results suggest that Chinese English learners use an average of 2240 inferential discourse markers per million words while native English speakers use 2027 of them. The conclusion is that Chinese English learners actually use slightly more inferential discourse markers in English composition in general, which is quite the contrary to our assumption based on previous research and commonsense. The reason may be that Chinese English learners tend to overuse use so and consequently, which compensate for the underuse of the others.

Table 1. Frequency of inferential discourse markers in both corpora

Rank	IDM	Raw Freq.	Raw Freq.	St. Freq.	St. Freq.
		TECCL	NESSIEv2	TECCL	NESSIEv2
1	<i>so</i>	1325	463	754.1	189.6
2	<i>then</i>	256	286	792.3	864.0
3	<i>therefore</i>	109	202	346.8	627.8
4	<i>thus</i>	44	136	140.0	422.7
5	<i>as a result</i>	25	35	79.5	108.8
6	<i>as a result of</i>	11	24	35.0	74.6
7	<i>consequently</i>	14	7	44.5	21.8
8	<i>hence</i>	9	23	28.6	71.5
9	<i>as a consequence</i>	2	2	6.4	6.2
10	<i>in that case</i>	2	0	6.4	0
11	<i>for that reason</i>	0	0	0	0
12	<i>that's why</i>	0	0	0	0
	Total	1797	1178	2240	2027

With regard to the preference, both groups use then the most frequently and use that's why, for that reason, in that case the least often. Chinese English learners use then 792.3 times per million words, followed by so (754.1 occurrences per million words) and therefore (346.8

occurrences per million words). Native English speakers use then 864.0 times per million words, followed by therefore (627.8 occurrences per million words) and thus (422.7 occurrences per million words). Chinese English learners may have overused so and consequently and underused the others; they use so almost four times as frequently as native speakers do. The results need to be validated by the Chi-square test.

Result of the Chi-square test of independence was shown in Table 2. There are significant differences between two groups' use of so, therefore, thus, as a result, as a result of, consequently, and hence. The chi-square value of in that case, for that reason and that's why cannot be calculated. The difference between two groups' use of then and as a consequence are not significant. According to the data, Chinese English learners overuse so, and consequently and underuse therefore, thus, as a result, as a result of and hence. To sum up, it is apparently that the use of inferential discourse markers by Chinese English learners are different from native speakers. Compared to native speakers, Chinese learners depend too much on so, and lack diversity and balance in choosing inferential discourse markers.

Table 2. Chi-Square Test of Independence (n=1,000,000)

Rank	IDM	St. Freq. TECCL	St. Freq. CROWN	Chi-square	Significance (p)	
1	SO	754.1	189.6	337.125	.000***	+
2	THEN	792.3	864.0	3.133	.077	-
3	THEREFORE	346.8	627.8	81.025	.000***	+
4	THUS	140.0	422.7	142.294	.000***	+
5	AS A RESULT	79.5	108.8	4.450	.035*	+
6	AS A RESULT OF	35.0	74.6	14.546	.000***	+
7	CONSEQUENTLY	44.5	21.8	7.896	.005**	+
8	HENCE	28.6	71.5	18.308	.000***	+
9	AS A CONSEQUENCE	6.4	6.2	0.000	1.000	-
10	IN THAT CASE	6.4	0	#NUM!	#NUM!	#
11	FOR THAT REASON	0	0	#NUM!	#NUM!	#
12	THAT'S WHY	0	0	#NUM!	#NUM!	#

*p<0.05, **p<0.01, ***p<0.001, # null or meaningless

Table 3. Distribution of Syntactic Positions

IDM	TECCL		NESSIE	
	INITIAL	MEDIAL	INITIAL	MEDIAL
SO	0.4%	99.6%	0	100%
THEN	27.3%	72.7%	8.3%	91.7%
THEREFORE	86.2%	13.8%	46.0%	54.0%
THUS	65.9%	34.1%	40.4%	59.6%
AS A RESULT	92%	8%	71.4%	28.6%
AS A RESULT OF	54.5%	45.5%	25%	75%
CONSEQUENTLY	64.3%	35.7%	42.9%	57.1%
HENCE	66.7%	33.3%	21.7%	78.3%
AS A CONSEQUENCE	100%	0	50%	50%
IN THAT CASE	100%	0	/	/
FOR THAT REASON	/	/	/	/
THAT'S WHY	/	/	/	/

The distribution of syntactic positions of inferential discourse markers in two corpora is shown in Table 3. There are both similarities and divergences. Both groups tend to put *so* and *then* in the middle of a sentence and *as a result* in the beginning of a sentence. While Chinese English learners prefer to start a sentence with *therefore*, *thus*, *consequently*, *hence*, *as a consequence*, native speakers often put them in the middle more often than not. In general, Chinese learners are inclined to use inferential discourse markers as a sentence starter while native speakers use them more flexibly.

To sum up, Chinese English learners use more inferential discourse markers than native English speakers in general in written English. Chinese English learners overuse *so*, and consequently and underuse *therefore*, *thus*, *as a result*, *as a result of* and *hence*. Compared with native speakers, Chinese learners prefer to put inferential discourse markers at the beginning of a sentence.

5. Discussion

5.1. Potential Causes

The gap between native speakers and ELF learners is often explained as a result of incompetence, despite the fact that even advanced learners could display similar tendencies as other learners. This section intends to discuss why Chinese EFL learners of different competence levels tend to develop a similar pattern from three perspectives, namely transfer of training, interlinguistic transfer and intralinguistic transfer.

Transfer of training could readily explain the fact that Chinese English learners use an excessive amount of inferential discourse markers. It is widely accepted that among the basic language skills Chinese English learners are better at reading and writing compared to listening and speaking. Although the recent decades have witnessed a growing concern about oral English, teachers and students themselves still attach great importance to writing. For one thing, Chinese teachers are pretty strict about the way students write. When marking a composition many would search for connectives on purpose as for many teachers coherence is even more important than the content itself. Therefore, teachers often overemphasize the importance of inferential discourse markers and require students to make use of them whenever possible, which results in the overuse. For another, many students rely on the so-called writing templates, which seem to have made writing effortless. However, this comes with a price. The templates are designed to make the whole writing seem logical no matter what the topic is, so to reach the flexibility they are crowded with connective and transitional expressions.

Transfer of training also attributes to Chinese English learners' preference for specific inferential discourse markers and their syntactic positions. Chen (2014) examined the inferential discourse markers in maritime English textbooks and the results suggested that the distribution and collocation of the inferential discourse markers were different from those of native speakers [16]. In addition, English teachers themselves usually cannot reach native-like level, which means their use of inferential discourse markers may be deficit in variety as well as quality. In a word, Chinese English learners are likely to be exposed to inadequate and imperfect input, which hinders their acquisition of inferential discourse markers.

Another major influencing factor is L1 transfer. Although positive transfer could originate from the similarities between the Chinese and English inferential discourse markers, it did not counteract but could deteriorate the negative transfer effect brought by subtle differences. Among all the inferential discourse markers Chinese learners have an apparent strong preference for *so*, whose Chinese counterpart is one of the most frequently used logical connector in both oral and written Chinese. Coupled with the transfer of training such as translation drills, learners would easily associate *so* with *suoyi* in their mental lexicon. It takes a lot of input and practice to overcome such inertia.

Overgeneralization of rules in the target language can also lead the Chinese learners to use inferential discourse markers in a different way. Intralingual transfer usually occurs when the learner has learned some rules of the target language but has not learned the restrictions of their application. Chinese learners tend to put inferential discourse markers in fixed position rather than use them flexibly as the native speakers do, which could be due to their assumption that some inferential discourse makers can only be used in that specific position based on scanty input.

5.2. Pedagogical Implications

Based on the problems discussed above, three tentative suggestions are provided to ameliorate the situation.

Firstly, explicit teaching of discourse markers is necessary. Inferential discourse markers could be taught consequently. Teachers should inform students when and where to use them, which one to use in detail rather than counting on unconscious acquisition. Not only should teachers raise students' awareness of inferential discourse markers, they'd better warn students to use them consciously and rationally rather than put them everywhere as if it is compulsory for them to use a whole bunch of connectives in every paragraph.

Secondly, learners should be exposed to authentic materials and native speakers. Since the differences are partly due to the inadequate and imperfect input, it would be better if the students could be exposed to a great amount of correct input. Authentic materials should be employed more often and native speaker teachers should be hired instead of relying solely on Chinese English teachers and textbooks written by Chinese. Exposure to authentic material and native speakers could not only increase the correct input but also increase the chance of being corrected. Reading authentic material can lead to self-correction if the students find the usage different from his. Communication with native speakers can lead to self-correction and also provide corrective feedback.

Thirdly, Teachers need to reflect on current teaching and reduce the use of drilling and templates. Teachers should reflect on their previous teaching and try to find the source of the problem in the classroom. For example, they can start with reducing the drilling of some inferential discourse markers and the use of writing templates, which would lead to dependence, stabilization and fossilization.

6. Conclusion

Although the concept of a single Standard English is sometimes rejected, as well as the idea of taking native speakers' language as the sole criterion, English teaching and learning in most contexts still take native-like usage as the ultimate goal. Contrastive study between interlanguage and target language is an important research method in Second Language Acquisition, which helped us find out how well Chinese EFL learners acquire discourse markers, and unveil the gap between Chinese EFL learners and native speakers of English. Then we took a step further to explore the potential causes and offered some implications on how to bridge the gap in English classrooms. The findings also shed some light on the nature of Second Language Acquisition process, exposing the gap between interlanguage and the target language.

References

- [1] M.Stubbs: *Discourse Analysis: The Sociolinguistic Analysis of Natural Language* (The University of Chicago Press, USA 1983).
- [2] B.Fraser: *Pragmatic Formatives*. In J. Verschueren, & M. Bertuccelli-Papi (Eds.), *The Pragmatic Perspective* (John Benjamins, The Netherlands 1985), p.179-194.

- [3] D.Blakemore: Semantic constraints on relevance (Blackwell, UK 1987).
- [4] L.Polanyi: A formal model of the structure of discourse, *Journal of Pragmatics*, Vol.12 (1988) No. 5-6, p.601-638.
- [5] B.Erman: Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction, *Language Variation & Change*, Vol.4 (1992), No.2, p.217-234.
- [6] A.Knott, R.Dale: Using Linguistic Phenomena to Motivate A Set of Coherence Relations, *Discourse Processes*, Vol.1 (1994), p.35-62.
- [7] V.Rouchota: Discourse Connectives: What Do They Link? *UCL Working Papers in Linguistics*, Vol. 8 (1996), p.1-15
- [8] K.Aijmer: *English Discourse Particles: Evidence from Corpus* (John Benjamins, The Netherlands 2002).
- [9] X.R.Chen: *Pragmatics and Foreign Language Teaching* (Foreign Language Teaching and Research Press, China, 2013).
- [10] D.Schiffrin: *Discourse Markers* (Cambridge University Press, USA, 1987).
- [11] B.Fraser: Toward a Theory of Discourse Markers. In F. Kerstin (Ed.), *Approaches to Discourse Particles* (Elsevier, The Netherlands 2006), p. 189-204.
- [12] K.Aijmer: Pragmatic Markers in Spoken Interlanguage, *Nordic Journal of English Studies*, Vol.3 (2004), p.173-190.
- [13] Information on: <http://114.251.154.212/cqp/teccl/>
- [14] Information on: <http://114.251.154.212/cqp/nessie2/>
- [15] J.J.Xu, L.P.Wu: Web-based fourth generation corpus analysis tools and the BFSU CQPweb case, *Technology Enhanced Foreign Languages*, Vol.5 (2014), p.10-15, 56.
- [16] M.M.Chen. A Study on Inferential Discourse Markers in CELWS Texts.