

# A Corpus-Based Study on Chinese English Majors' Use of Discourse Markers

Yongping Ding<sup>1</sup>, Changyu Wang<sup>2</sup> & Yongjuan Ding<sup>3</sup>

<sup>1</sup> School of Foreign Languages, Jiangsu University, Zhenjiang, China

<sup>2</sup> School of Mechanical Engineering, Jiangsu University, Zhenjiang, China

<sup>3</sup> Jingcheng High School, Yangzhou, China

Correspondence: Yongping Ding, School of Foreign Languages, Jiangsu University, Zhenjiang, 301 Xuefu Road, Jingkou District, Jiangsu, China. E-mail: dyp19900603@163.com

Received: March 17, 2015      Accepted: April 2, 2015      Online Published: May 30, 2015

doi:10.5539/ijel.v5n3p56      URL: <http://dx.doi.org/10.5539/ijel.v5n3p56>

## Abstract

Discourse markers (DMs for short hereafter) are used widely by native speakers and L2 learners. Previous studies are mostly about the acquisition and application of DMs, but studies on the potential factors that might affect the use of DMs are rare. Under the Relevance Theory, the present corpus-based study aims to reveal the influence of gender and oral proficiency on the use of DMs by Chinese English majors. It found that: (1) generally speaking, the variety of DMs employed by Chinese learners of English is rather limited; (2) male Chinese learners of English use more DMs than female learners; (3) high proficiency Chinese learners of English employ more DMs than the low proficiency group; moreover, the variety of DMs used by the former group is much larger than that of the latter. These findings reveal that the learner factors of gender and oral proficiency do influence the use of DMs in L2 learners' conversations and they should be taken into consideration when we are learning or researching.

**Keywords:** Chinese English major, discourse marker, gender, oral proficiency

## 1. Introduction

Discourse markers (hereafter DMs) are words or structures of no propositional content but rich of communicative functions. DMs, generally speaking, include some connectives (e.g., *and*, *therefore*, *because*), adverbials (e.g., *actually*, *incidentally*), interjections (e.g., *well*, *oh*), some phrases (e.g., *as a consequence of*, *in conclusion*) and minor sentences (e.g., *you know*, *I mean*, *if I'm not wrong*).

In the past years, Chaudron and Richards (1986), Levinson (1983) and Schiffrin (1987) make theoretical studies on English DMs which are mainly focused on their definitions; Fraser (1999), Levinson (1983), Schiffrin (1987) and Sperber and Wilson (2001) do researches on functions of DMs; Fraser (1999) also makes some investigations on their classifications. Empirical studies are mostly concerned with their applications in language learning and how they are used differently by native and non-native speakers of English in verbal communication. However, there are few studies discussing the influence of learner factors on the use of DMs. As revealed in Second Language Acquisition, such learner factors as oral proficiency, gender, personality do influence the process of L2 learning. We have no reason to reject the hypothesis that proficiency and gender also influence the acquisition of DMs.

Taking the limitations of previous studies, the present research, by adopting a corpus-based method, is designed to disclose the influence of oral proficiency and gender on the use of DMs in conversations. To be specific, the present study aims to find answers to the following three questions:

- 1) How do Chinese learners use English DMs in conversations?
- 2) How do Chinese Learners' use of English DMs relate to their genders?
- 3) How do Chinese learners' use of English DMs relate to their oral proficiencies?

## 2. Method

### 2.1 Corpus

The corpus we referred to is the Spoken English Corpus of Chinese Learners (abbreviated as SECCL hereafter) constructed by Nanjing University, published by Foreign Language Teaching and Research Press in 2005 and well recognized in the field of Second Language Acquisition. As is summarized by Wen (2005), “it is featured by six characteristics”. The first one is representative of spoken English by Chinese Learners because all of its materials were randomly selected. It, therefore, provides an access for us to be familiar with oral English levels of Chinese students.

The corpus employed in this paper involves 68 English sophomores, of which 62 are females and the rest are males (The percentage reflects the actual situation in China). In terms of oral proficiency, 34 speakers are graded as high proficiency learners by experts chosen by Nanjing University and the others are graded as low proficiency learners. The number of words of each category (i.e., gender and proficiency) is listed. (See Table 1)

Table 1. Size of the corpus employed

SECCL	Gender		Proficiency	
	Male	Female	High	Low
	4 327	15 527	10 409	9 445
Total	19 854		19 854	

### 2.2 Theoretical Framework

According to the Relevance Theory proposed by Sperber & Wilson (2001), every ostensive utterance communicates a presumption of optimal relevance. In the theory, relevance is defined in terms of cognitive effects and processing effort. Cognitive effects are achieved when the newly-presented information interacts with a context of existing assumptions by strengthening an existing assumption, by contradicting and eliminating an existing assumption, or by combining with an existing assumption to yield a contextual implication. The greater the cognitive effects, the greater the relevance will be.

Cognitive effects, however, do not come free: it costs some mental efforts to derive them, and the greater the effort needed to derive them, the lower the relevance of an utterance will be. In order to achieve the greatest cognitive effects, the hearer (reader) must process the utterance in the right, i.e. the intended context. The selection or construction of context is governed by the search for optimal relevance. As far as the communicator is concerned, she may have reason to believe that the hearer will choose the appropriate contextual assumptions and draw the appropriate conclusions without any extra help from her, or she may decide to direct the hearer towards the intended interpretation by making a certain set of assumptions more easily accessible. DMs fulfill just this role.

Blakemore (1987, 1992, 2002) and Wilson and Sperber (1993) approach DMs within the relevance-theoretic framework. In particular, Blakemore (1987) reanalyzes Grice’s DMs, using a distinction between conceptual and procedural encoding. She proposes that DMs do not have a conceptual meaning, but have only a procedural meaning, which consists of instructions about how to manipulate the conceptual representation of the utterance. Wilson and Sperber also argue that discourse connectives are procedural and non-truth-conditional: they encode procedural constraints on implicatures. They help to provide utterance of optimal relevance, by guiding the search for intended contexts and cognitive effects, which saves a lot of processing efforts; consequently, the intended interpretation can be achieved much more efficiently.

#### 2.2.1 Choice of the Five Discourse Markers in the Study

A lot of researches made studies on the DMs which appeared most frequently. (See Table 2)

Table 2. The most frequently-used DMs revealed in the literature

Scholars	The most frequently-used DMs
Chaudron and Richards (1986)	<i>well, right, now, so</i>
Fuller (2003)	<i>you know, like, oh, well, yeah, I mean</i>
He and Xu (2003)	<i>oh, just, you know, right, well, I mean, I think, okay</i>
Holmes (1986)	<i>you know</i>
Holmes (1990)	<i>you know, sort of, I think</i>
Kyrtzias and Ervin-Tripp (1999)	<i>because, but, so, well, okay</i>
Wang and Wang (2008)	<i>well, I mean, you know, actually</i>

From Table 2, some researchers compare functions of DMs in women's and men's speech, some explain developmental changes in the use of DMs, some study the effects of DMs on the comprehension of lectures, some investigate the influence of speaker roles on the use of DMs, and some make a comparison of the use of DMs by L2 learners with that by native speakers.

According to previous studies, the DMs selected in the present study are *well, I mean, you know, just* and *I think*. To examine the use of DMs in the corpus, we only focus on the use of them as DMs and the use with other functions is excluded.

### 2.2.2 Exclusion of Irrelevant Instances

To achieve the authentic data, we weeded out the irrelevant situations in which the above words (*well, I mean, you know, just* and *I think*) are not used as DMs. For example, *wells* used in the flowing situations were excluded:

#### i) Adverb, e.g.

I'm not sure whether I can get on (with) it *well* or not. (Here *well* is used as an adverb to modify the verb phrase "get on with")

#### ii) Formulaic Expression, e.g.

Oh, then you will get a lot of social experience as *well* as pocket money. (Here *well* in the phrase "as well as" is to indicate the meaning of getting social experience and pocket money at the same time.)

With 71 irrelevant instances eliminated, the final statistics suggest that the raw frequency of the five DMs is 541, including 35 *wells*, 4 *I means*, 128 *you knows*, 67 *justs*, and 307 *I thinks*.

### 2.3 Data Analysis

Antconc 3.2.2 was employed in the analysis of the corpus to see how Chinese English majors use DMs in their conversations.

First, the corpus was processed in Wordlist Tool of Antconc 3.2.2 before tagging.

Second, Concordance Tool of Antconc 3.2.2 was employed to find out all instances of the five DMs communicated in the whole corpus.

Third, the irrelevant instances of each DM were weeded out by the researcher and the raw frequency of each DM is worked out.

Fourth, each DM is tagged according to different variables. For example, *well* in the sub-corpus of male learners is tagged as <DM-W-M>.

Fifth, all the raw frequencies were converted into normalized frequencies, because the four sub-corpora are different in size.

Sixth, the proportions of each DM in their overall DM usages were worked out.

Last, the various correlations of DMs between the two sub-corpora with different genders and those between the two sub-corpora with different oral proficiencies were carried out by implementing Chi-square analysis to see whether the differences were significant.

## 3. Results and Discussion

### 3.1 General Use of Discourse Markers by English Majors

Figure 1 reports the frequencies of the five DMs in the spoken corpus, showing how Chinese English majors employ DMs in general.

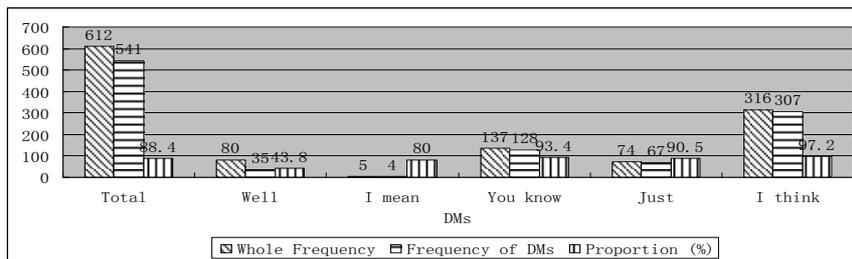


Figure 1. Overall pragmatic uses in SECCL

These five words appeared 612 times in the whole corpus among which they were used 541 times as DMs, accounting for 88.4%. It also shows that *I mean*, *you know*, *just* and *I think* were all used as DMs for over 80%, with the proportion of *I think* ranking the highest (97.2%) and that of *I mean* the lowest (80.0%). However, the proportion of *well* used as DMs was the lowest of all, only 43.8%. In terms of the number of each DM, the number of the use of DM *I mean* is the smallest, only 4 times.

Figure 2 presents the respective frequencies of the five DMs in their overall DM usages in SECCL.

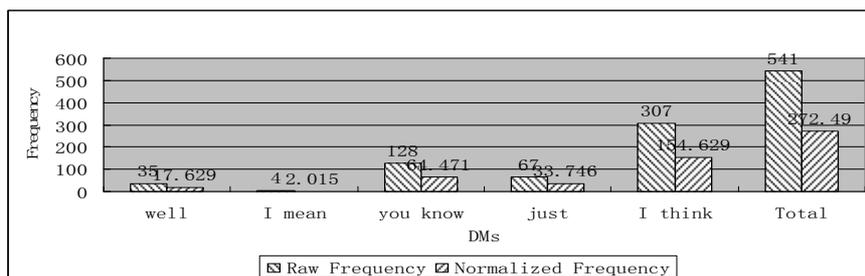


Figure 2. Respective frequencies of DMs used in SECCCL

Among the five DMs, the use of the DM *I think* is the most frequent, with the normalized frequency ranking the highest, that is 154.629, accounting for more than half of the total normalized frequency (272.490). We can also see that DMs *I mean* and *well* are the least used in this corpus, and the sum of their normalized frequencies is much smaller than that of DM *just* (17.629+2.015<33.746), accounting for about 7.2% of the total normalized frequency.

### 3.2 The Influence of Gender on the Use of Discourse Markers

Since the two sub-corpora are different in size (see Table 1), the data presented in this part are all normalized frequencies (the appearing times per 10, 000 words) in order to make a comparable statistical comparison.

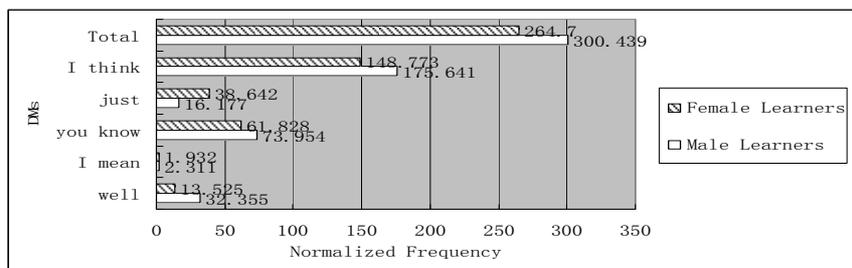


Figure 3. Normalized frequency of DMs used by male and female learners

Figure 3 shows that the use of all the five DMs by male learners is more frequently than that of females, since the former total normalized frequency is higher than the latter (300.439>264.700). Although there is a gap between the normalized frequency of each DM by different sex learners, the top one is the same DM *I think*, which accounts for more than half of each of their total normalized frequency (175.641 and 148.773 respectively). The second frequently-used DM is *you know*, with the normalized frequency of 73.954 and 61.828

respectively.

However, the uses of last three DMs are a little different. According to their normalized frequencies, when ordered from the highest to the lowest, the three DMs used by male learners were *well*, *just* and *I mean* but those used by female ones were *just*, *well* and *I mean*, but both of the rarely-used DM is *I mean*.

In order to see how significant these differences are, here we implement Chi-square analysis in Table 3 to compare the data fully.

Table 3. Differences of the use of DMs by male and female learners

DMs	Normalized Frequency		X <sup>2</sup>	P
	Male Learners	Female Learners		
well	32.355	13.525	6.818 4	.009**
I mean	2.311	1.932	0.024 1	.877
you know	73.954	61.828	0.776 9	.378
just	16.177	38.642	5.077 8	.024*
I think	175.641	148.773	1.604 7	.205
Total	300.439	264.700	1.630 6	.202

Note. "\*\*" here means  $P < .05$ ; "\*\*\*" here means  $P < .01$ .

We can see that the difference in the use of DM *well* is the most significant since the value of chi-square is the biggest (6.818 4) and P here is the smallest (.009\*\*), smaller than .01. DM *just* is also used differently by male and female learners, with the value of chi-square (5.077 8) ranking as the second and P (.024\*) is much smaller than .05. While the other three DMs *I think*, *you know* and *I mean* were used without significant differences especially the latter two DMs.

### 3.3 The Influence of Oral Proficiency on the Use of Discourse Markers

Since the two sub-corpora are different in size (see Table 1), the data presented in this part are all normalized frequencies (the appearing times per 10 000 words) in order to make a comparable statistical comparison.

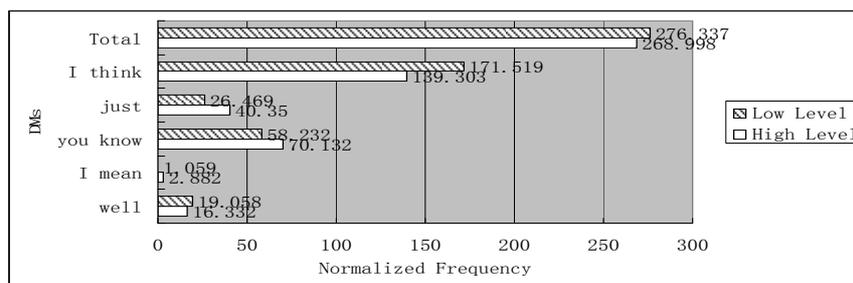


Figure 4. Normalized frequency of DMs used by learners with different oral proficiencies

Figure 4 shows that the use of all the five DMs by more advanced learners is less frequently than that of less advanced ones, since the former total normalized frequency is a little smaller than the latter ( $268.998 < 276.337$ ). Although there is a gap between the normalized frequency of each DM by learners with different oral proficiencies, the top one is the same DM *I think*, which accounts for more than half of each of their total normalized frequency (139.303 and 171.519 respectively). The second frequently-used DM is *you know*, with the normalized frequency of 70.132 and 58.232 respectively.

Meanwhile, the uses of last three DMs are a little similar. According to their normalized frequencies, when ordered from the highest to the lowest, the three DMs used by both groups are the same: *just*, *well* and *I mean*, and the rarely-used DM in both sub-corpora is *I mean*.

However, from Figure 4, we also know that comparing the use of DMs in general, learners with high proficiency used DMs more frequently than those with low proficiency (except for DMs *I think* and *well*). It must be noticed that the use of DM *I think* by less advanced learners is much more frequent than that of more advanced ones ( $171.519 > 139.303$ ). But for the other four DMs, the normalized frequency of each one is higher in the corpus of learners with high proficiency than that in the other corpus, except for the DM *well*, but the gap is not very big

(16.332 and 19.058).

In order to see how significant these differences are, here we implement Chi-square analysis to compare the data fully.

Table 4. Differences of the use of DMs by learners with different oral proficiencies

DMs	Normalized Frequency		X <sup>2</sup>	P
	High Level	Low Level		
well	16.332	19.058	0.209 1	.648
I mean	2.882	1.059	0.817 3	.366
you know	70.132	58.232	1.094 7	.295
just	40.350	26.469	2.836 8	.092
I think	139.303	171.519	3.376 0	.066
Total	268.998	276.337	0.100 6	.751

From Table 4, we can see that among the five DMs, the difference in the use of DM *I think* is the most significant since the value of chi-square is the biggest (3.3760) and P here is the smallest (.066), and learners with low level use DM *I think* more frequently than learners with high level. DM *just* is also used differently by learners with different oral proficiencies, with the value of chi-square (2.8368) ranking as the second and P (.092) is the second smaller. While the other three DMs *you know*, *I mean* and *well* were used without obviously significant differences.

#### 4. Conclusion

The aim of the present paper is to find the differences in the use of DMs by different sex and proficiency learners. On the basis of the corpus selected for analysis, the study yields the following major findings: (1) The use of DMs by Chinese English majors are not well balanced. That is to say, they prefer to use certain markers like *you know*, *I think*, and *just*, but some markers like *I mean* and *well* are of rather low occurrence frequency; (2) Male learners use DMs more frequently than female ones; (3) High proficiency learners use DMs more frequently than those low proficiency learners, and the variety of DMs employed by high proficiency learners are larger than the low proficiency group.

The present study has important theoretical and practical implications. Theoretically speaking, the study can further our understanding on the use of DMs by L2 learners, which will help enrich the theory of DMs. Practically, it will provide instructive suggestions to editors of oral English teaching material, teachers of English and all English learners, helping direct them to speak in a more native-like way.

#### References

- Blakemore, D. (1986). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Blakemore, D. (1987). *Semantics constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (1992). *Understanding utterances: An introduction to Pragmatics*. Oxford: Blackwell.
- Chaudron, C., & Richards, C. J. (1986). The effects of DMs on the comprehension of lectures. *Applied Linguistics*, 7, 113. <http://dx.doi.org/10.1093/applin/7.2.113>
- Fraser, B. (1999). Types of English discourse markers. *Acta Linguistica Hungarica*, 38, 19-33.
- Fuller, J. M. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, 35, 23-45. [http://dx.doi.org/10.1016/S0378-2166\(02\)00065-6](http://dx.doi.org/10.1016/S0378-2166(02)00065-6)
- He, A. P., & Xu, M. F. (2003). 中国大学生英语口语 Small Words 的研究. *Foreign Language Teaching and Research*, 6, 446-452.
- Holmes, J. (1986). Functions of You Know in Women's and Men's Speech. *Language in Society*, 15, 1-21. <http://dx.doi.org/10.1017/S0047404500011623>
- Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language and Communication*, 10, 185-205. [http://dx.doi.org/10.1016/0271-5309\(90\)90002-S](http://dx.doi.org/10.1016/0271-5309(90)90002-S)
- Kyratzis, A., & Ervin-Tripp, S. (1999). The development of discourse markers in peer interaction. *Journal of Pragmatics*, 31, 1321-1338. [http://dx.doi.org/10.1016/S0378-2166\(98\)00107-6](http://dx.doi.org/10.1016/S0378-2166(98)00107-6)

- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.  
<http://dx.doi.org/10.1017/CBO9780511611841>
- Sperber, D., & Wilson, D. (1993). Linguistic form and relevance. *Lingua*, 90, 1-25.  
[http://dx.doi.org/10.1016/0024-3841\(93\)90058-5](http://dx.doi.org/10.1016/0024-3841(93)90058-5)
- Sperber, D., & Wilson, D. (2001). *Relevance: Communication and Cognition*. Beijing: Foreign Language Teaching and Research Press.
- Wang, L., & Wang, T. S. (2008). 中国英语学习者语用标记语习得研究——一项基于 SECCL 和 BNC 的实证研究. *Modern Foreign Languages*, 3, 291-300.
- Wen, Q. F., Wang, L. F., & Liang M. C. (2005). *中国学生口语语料库*. Beijing: Foreign Language Teaching and Research Press.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).