The University of Nottingham Ningbo, China

English Discourse Markers in Native and Non-native Academic Discourse

Lin Xiuli

MA Applied Linguistics and ELT

English Discourse Markers in Native and Non-native Academic Discourse

by

Lin Xiuli

2018

A Dissertation presented in part consideration for the degree of
MA Applied Linguistics and ELT

Abstract

This study aims to examine and compare the differences in discourse marker uses by native speaker and non-native (L1 Chinese) speakers based on the Corpus of Chinese Academic Written and Spoken English (CAWSE), an on-going corpus project at The University of Nottingham Ningbo China, and the Michigan Corpus of Academic Spoken English (MICASE), an academic spoken corpus at the University of Michigan in the U.S. The results shows significant differences in two cases, namely *you know* and *I think*. It also indicates that discourse markers are important in organizing, understanding and communicating at the textual and interpersonal levels. The two groups of speakers share a partially overlapping pattern in using discourse markers. The native speakers prefer to use discourse markers with more pragmatic functions, whereas the L1 Chinese speakers may overuse discourse marker functions mainly at the textual level. Finally, possible implications and suggestions are discussed so as to involving discourse markers in English language teaching and learning.

Keywords: Discourse markers, Spoken grammar, Oral English, Second language acquisition, Native-like speech, English language teaching

Acknowledgments

To begin with, I would like to thank my beloved supervisor, Dr. YuHua Chen, without whom I could not have finished this dissertation. She was patient and helpful when answering my questions and offering suggestions. I would like to acknowledge her insights in Corpus Linguistics, which enlightened my understanding towards this study. I would also like to thank my classmates Wu Fan and Xue Yu-Tong, the peer discussion with whom always sparked a new idea and dispelled my doubts. At last, I am also grateful for the University of Nottingham Ningbo China, in which the one-year experience of the MA program will absolutely make difference in my life.

Contents

Chapter 1 Introduction

Speaking is an important skill in second language learning. Although the necessity of producing native-like speech has been argued by researchers (Götz, 2013), imitating a native speaker's speech, namely conventional expressions, is one way to help learners to comprehend and interact naturally with others in the target language. According to my own experience, of all the "conventional expressions", discourse markers are highlighted, because the grammar-translation teaching method and exam-oriented learning prevent learners from practicing speaking skills inside and outside the English as foreign language (EFL) classrooms in China (Fung, 2003; Fung & Carter, 2007; Liu, 2013). Discourse markers are often neglected in such a teaching style, however, which can "facilitate listener comprehension and help smooth spontaneous interaction between speakers" (Liao, 2009). This study aims to bring more attention on implementing DMs in EFL classrooms by investigating the patterns of EFL learners' DM use and exploring pedagogical implication in spoken English teaching.

The current study will exemplify a combination of qualitative and quantitative analyses based on the approach, namely the corpus-driven approach or "bottom-up" approach (McEnery & Hardie, 2012; Mu¨ller, 2005). The justification for applying the corpus-driven approach will be presented in the literature review part. In the methodology chapter, an overview of data source, data selection, the design of the study, and a preliminary finding in the pilot study will be presented. Chapter 4 consists of two case studies of selected discourse markers and comparison of the differences in their discourse marker functions used by native speakers and non-native Chinese speakers of English (L1 Chinese speakers). *You know* in discourse marker function is preferred by native speakers; while *I think* is used with a significant high frequency by L1 Chinese speakers (p<0.0001). In the last chapter, suggestions and limitations in regard to this study will be discussed.

Chapter 2 Literature Review

2.1 Introduction

    (01)   s075: **I think you know** the mm the company break up is the is the main

               punishment

                                                                      (C075yk80)

Discourse should not only be seen "as a unit of language, but as a process of social interaction (Maschler & Schiffrin, 2015, p. 190). In example (01), the speaker is definitely not uttering that "I believe you know the fact that I am saying". Rather, the speaker is trying to maintain the floor while thinking for what is going to say and draw the attention from the hearer. In this chapter, the previous studies of discourse markers will be reviewed in four parts, namely definition, previous study areas in regard to discourse markers, features of discourse marker as a framework, and a review of approach and related theories in the current study.

2.2 Definition of Discourse Markers

2.2.1 Terminologies

Many terminologies have been used to describe the lexical items in bold in example (01), for instance, pragmatic particles (Östman, 1995, p. 31), discourse particles (Aijmer, 2002), pragmatic markers (Aijmer, 2004; Brinton, 1996; Buysse, 2017; Erman, 2001), discourse markers (Brinton, 2010; Carter & McCarthy, 2006; Fraser, 1990; Jucker & Ziv, 1998; Redeker, 1993; Schiffrin, 1987), adverbials, conjunctions, etc. (Dér, 2010; Heine, 2013). Researchers are found to change their terms after several years or simply lack of distinction between different terms, for example, Aijmer (2004) had transformed her term from discourse particles to pragmatic markers, which was discussed in Aijmer, Foolen, and Simon-Vandenbergen (2006) by distinguishing the scope of two terms, in which her former term "discourse particles belong […] to the more general category of pragmatic markers". Similarly, Carter and McCarthy (2006) included discourse markers as a part of pragmatic markers that are "a class of items which operate outside the structural limits of the clause and which encode speakers' intention and interpersonal meanings" (p. 208). In this study, I choose to follow the term used by Schiffrin (1987), Carter and McCarthy (2006) as Discourse Markers (henceforth as DMs).

2.2.2 Definition

Discourse makers are lexical items which are outside the structural limits and serve as the connection and cohesion between the segments of the discourse "in ways which reflects the choices of monitoring" (Carter & McCarthy, 2006, p. 208). Schiffrin (1987) defined discourse markers operationally as "sequentially dependent elements which bracket units of talk" (p. 31). In Maschler and Schiffrin (2015), discourse markers are defined as "expressions such as *oh, well, y'know*, and *but*" which "are one set of linguistic items that function in cognitive, expressive, social, and textual domains." Opinions vary on the definition of DMs. One common point that is supported by many researchers is where discourse markers are syntactically independent from the sentences or clauses (Brinton, 1996; Jucker & Ziv, 1998; Lutzky, 2012; Mu¨ller, 2005). In other word, words or phrases that construct as a part of a sentence or clause structure should not be considered as a discourse marker. For example, *you know* in example (01) serves as a DM function, because it connect the segments of utterances and does not appear within the structure; while *you know* in "You know her?" is a part of the question of "Do you know her?" where "do" is omitted (Mu¨ller, 2005, p. 153). Consequently, the first definition is considered as a better description in this study. More importantly, the features that make a lexical item as a discourse marker will be discussed in Section 2.4.

2.3 Previous Areas of Discourse Markers Studies

The studies of discourse markers raises the interest of research from around the age of 1990. As one of the pioneers in this field, Schiffrin (1987) attempted to "describe in a systematic way the discourse in which markers occurred" (p. 312; Cf. Aijmer, 2013, p. 10)
Brinton (1996) made great efforts to identify the features for defining DMs. Some researchers focused on the DMs used in early modern English (Jucker, 2002; Lutzky, 2012). In this section, two areas will be reviewed, namely DMs studies in corpus linguistics and second language acquisition.

2.3.1 Discourse Markers, Corpora and Corpus Linguistics

Discourse markers have been studied by analysing the texts from large corpora since 1980 (Mu¨ller, 2005, p. 10). Erman (1987, 2001), for instance, took corpora as her main resource for analysing *you know* and *I mean*. Aijmer (2002, 2004, 2013) also studied pragmatic markers on the basis of the texts from corpora applying corpora-based approach. She also

mentioned that the increasing corpora "[…] make it possible to compare pragmatic markers both across text types and across regional and social varieties" (Aijmer, 2013, p.8). Corpus linguistics, on the other hand, "is an area which focuses upon a set of procedures, or methods, for studying language" (McEnery & Hardie, 2012, p. 1). The procedures refer to a set of corpus tools, such as concordancing, which are central to the approach. The concordancer allows users to search a specific word, part of a word or a phrase in a corpus and present the results in the context, viz. concordance lines, which are displayed in "one-example-per-lime format" (McEnery & Hardie, 2012, p.35) . These procedures enable researchers to explore in linguistic studies. In this study, the concordancer will applied as a major tool for searching, sequencing and evaluating the discourse markers in two different contexts, i.e. native and non-native speakers in academic discourse.

With the development of the computer technology, corpus study is no longer an exclusive area for specialists but is open to the public (Hunston, 2002), which also makes the corpora as a resource of materials for language learning and teaching. In the next section,

2.3.2 Discourse Markers and Second Language Acquisition

Exploring the pedagogical implication of DMs in EFL classrooms is one of the motivation for me to conduct this study. According to my experience, the main source of knowledge for learning English in China is from the classroom where written discourse markers such as *firstly, secondly, because, but, so, however*, etc. are focused in the EFL classrooms (Fung & Carter, 2007); whereas, knowledge of spoken discourse markers is often neglected in English language teaching (Mu¨ller, 2005). Trillo (2002) conducted a corpus-driven comparison in DMs uses by native and non-native Children and concluded that non-native children might have the capability to pick up the pragmatic uses of English in communication in the same way as native children does, if those pragmatic functions were introduced in EFL classrooms. (p. 783). Therefore, EFL teachers should take a great proportion of responsibility for this consequence, i.e. pragmatic fossilization as suggested by Trillo (2002) that is "the phenomenon by which a non-native speaker systematically uses certain forms inappropriately at the pragmatic level of communication" (p. 770). My assumption is that if there is a practical way for teachers to acquire the knowledge of DM functions and evaluate the accuracy of students' uses of those DMs, teachers may be motivated to devote their energy in teaching such kind of knowledge.

2.4 Features of Discourse Markers

Many researchers have tried to distinguish the functions depending on the formal features, for instance, the position of DMs in utterances, stress and intonation and collocations; and they also find the multifunctionality (Aijmer, 2013, pp. 16-17; Erman, 1987, pp. 182-185; Fung & Carter, 2007; O¨stman, 1981, pp. 21-23; Schiffrin, 1987, pp. 291-294; Tree & Schrock, 2002). Brinton (1996, pp. 33-35) in *Pragmatic Markers in English* concluded twelve broad characteristics of DMs by examining previous studies on general and individual DMs. Jucker and Ziv (1998, p. 3) reorganized and abbreviated the features of spoken discourse markers based on Brinton's (1996) list. Those features are summarised according to the levels of linguistic descriptions (see Table 2.1). This list will be regarded as a scale for measuring the degree of "discourse markerness" (Jucker & Ziv, 1998; see also Brinton, 2008, p. 15), since DMs are found to have no necessity to comply with all the features (Brinton, 2010, p. 29) . "Prototypical discourse markers" (Jucker & Ziv, 1998) such as *well* satisfy more features on the list, while less prototypical markers (*you know* and *like*) tend to exhibit fewer qualities, instead, they have semantic meanings (Jucker, 2002, p. 213).

Table 2.1 List of basic features of discourse markers (Jucker and Ziv, 1998, p. 3; based on Brinton, 1996, pp. 33-35)

- Phonological and lexical features:
  a. They are short and phonologically reduced.
  b. They form a separate tone group.
  c. They are marginal forms and hence difficult to place within a traditional word class.
- Syntactic features:
  d. They are restricted to sentence-initial position.
  e. They occur outside the syntactic structure or they are only loosely attached to it.
  f. They are optional.
- Semantic feature:
  g. They have little or no propositional meaning.
- Functional feature:
  h. They are multifunctional, operating on several linguistic levels simultaneously.
- Sociolinguistic and stylistic features:
  i. They are a feature of oral rather than written discourse and are associated with informality.
  j. They appear with high frequency.
  k. They are stylistically stigmatised.
  l. They are gender specific and more typical of women's speech.

2.4.1 Phonological and lexical features

Schiffrin (1987, p. 328) notes that a discourse particle 'has to have a range of prosodic contours e.g. tonic stress and followed by a pause, phonological reduction " (see also Fraser, 1996). The phonological features of discourse markers describe them as being phonologically reduced and (often) forming a separate tone group (Dér, 2010, p. 15).

> "Unlike clitics, which are prosodically dependent, discourse markers and their parenthetical kin are prosodically independent. Typically, they are both accented and prosodically seperated from their surrounding context, by pauses or intonation breaks or both" (Zwicky, 1985, p. 303).

As corpus displays spoken data in transcriptions (some of them may not provide audio data), the evidence of phonological reduction may not be directly available in every corpus. Indirect evidence like the abbreviated spelling of *you know* as *y'know*, *because* as *cuz*, etc. However, lack of checking the actual pronunciation by listening to the audio recordings will make it difficult to identify this feature, for "the orthographic representation […] may or may not reflect the phonological reduction" (Jucker, 2002, p. 212).

In the corpora of this study, both linguistic and non-linguistic features, such as hesitation markers (*en, er, erm*), pauses, filled pauses and breathing sound (*hh*), are transcribed (CAWSE convention, see Appendix). Thus, although phonological reduction could not be easily identified without examining the audio, indirect clues for distinguishing a DM from other linguistic items can be drew from the transcriptions (Fung, 2003). For example, *well* in example 01 exhibits its independent phonological feature by the following vocalization *um*, serving as a filled pause.

> (02)   S3: **well**, um i'm not really sure like she of course is like you know, fourth
>   fifth […] generation but her husband's first generation like he lived in China.
>
> <div align="right">(MSGR)</div>

Lutzky (2012) suggests that DMs that form a separate tone group can also be identified from written data, for spoken discourse markers may "have functions that come close to e.g. those of punctuation or paragraphing in written texts" (Erman, 2001, p. 1339). Fraser (1990) illustrates in examples (03a) and (03b), where the functions of *now* are distinguished by the comma: in (03a) it functions as a discourse marker, and a time adverbial in (03b).

(03a)   A: John left. **Now,** Mary was really frightened.

(03b)   A: John left. **Now** Mary was really frightened.

<div align="right">(Cf. Fraser, 1990, p. 388)</div>

However, punctuations in the transcriptions are not explicit in utterances by speakers, but added by transcribers manually. Commas and full stops in MICASE corpus represent pauses between 1-2 seconds (see MICASE convention in Appendix). For instance, in example (04), *yeah* in Line (a) was separated from the following utterances by the comma and marks a separate tone unit. It is thus distinguished prosodically. On the contrary, *well* in b reflects a different manifestation (we can only tell from the given transcription), because the orthographic representation seems prosodically dependent on the utterances that follows. This is supported by one empirical study in phonetics that the DM function well only appears to be an independent tone unit at 50% rate (Dér, 2010). On the other hand, the definition of a pause by its timing has been discussed in studies of fluency, and pauses ranging from 0.2 to 0.5 seconds are considered as a cut-off (Lennon, 1990; Riggenbach, 1991). It is possible for the speaker s3 to leave a pause under 1 second following *well*. Consequently, the current study will merely take the prosodic feature as a reference in identifying position.

Example (04)

(04)   S2: **yeah**, well didn't Carol say some stuff about being able to hook you up
             with someone from Dinersty? or what did she say about that?
        S3: **well** she um the guy Brian who Kelly knows…

<div align="right">(MSGR)</div>

Finally, discourse markers are "marginal forms and hence difficult to place within a traditional word class" (Table 2.1; see also Brinton, 1996, p.34), indicating that DMs are "a heterogeneous set of forms"(Muller, Brinton). DMs consist of various kinds of expressions which are derived from the traditional grammatical categories such as adverbials (*now, then*), interjections (*well*), verbs (*look, see*), coordinators (*and, or, but*), subordinators (*so, however*), phrases (*sort of*) (Fraser, 1990, p.388; Cf. Lutzky, 2012, p.14) and even clauses (*you know, I think*) (Carter & McCarthy, 2006). Many attempts have been made to add DMs as members of the word classes as interjections (e.g. Fraser, 1990, p.391), adverbials (Fraser, 1990, p.388), particles, and function words (Cf. Brinton, 1996, p.34-35). The functions of

DMs share "various degrees of functional similarities and partially overlapping distributions" (Schiffrin, 1987, p.65) with other classes of pragmatic markers.

2.4.2 Syntactic Features

In this study, discourse markers are discussed on the utterance level rather than the sentence level, whose pragmatic functions are examined on the level of discourse, for the difference between spoken and written texts (Lutzky, 2012). The first description in Jucker and Ziv's (1998) syntactic features list is arguable. Brinton (1996) has explicated that discourse markers not only occurs sentence initially but also occurs medially and finally (p.33). Likewise, many other studies have shown that discourse markers are not restricted to sentence-initial position (Aijmer, 2013; Fung & Carter, 2007; Lutzky, 2012; Schiffrin, 2006), as well as in my data. Different positions may associate with different pragmatic functions. DMs that occur in the initial position usually function to initiate the discourse (e.g. *well, so, and*), including drawing attention from the hearer (e.g. *you know*) (Brinton, 1996, pp. 37-38); to mark boundaries of discourse, "that is, to indicate a new topic, a partial shift in topic (correction, elaboration, specification, expansion), or the resumption of an earlier topic (after an interruption)"; to close discourse or a topic (e.g. *you know, so*) (Brinton, 1996; Fung & Carter, 2007).

DMs that are inserted in the medial utterance serve functions such as holding floor or clarifying meaning (Fung & Carter, 2007). Some are regarded as filled pauses without meaning, but just a delaying tactic used to sustain discourse. Others are observed at the interpersonal level, that is, to appeal for understanding or mark shared knowledge (Fung & Carter, 2007). DMs that appear at the end of an utterance are the least frequent. They are usually considered as "comments, classification or as an afterthought" (Fung & Carter, 2007).

The second description is supposed to be a reliable criterion to distinguish an expression of discourse marker from those of non-discourse marker (Lutzky, 2012, p.16). DMs should be excluded from a clause structure and are not essential for the semantic or grammatical meaning of an utterance (Carter & McCarthy, 2006). It is closely related to or may be mixed up with the third description that discourse markers are optional, "which means that their omission usually does not render the discourse either ungrammatical, unintelligible or incomplete" (Lutzky, 2012, p.17). Disagreements have been voice by Lutzky (2012, p.18; also see Dér, 2010, p.14; Redeker, 1993) that the optionality does not appear to be a

unique characteristic to DMs (e.g. the adverbial *now* in Example 02-b can be left out without influencing the propositional content), and "where the DMs are not present, the hearer is left without a lexical clue as to the relationship intended between the two segments" (Fraser, 1999, p.944). Brinton (1996) also states that the omission of a DMs will cause the discourse to be "unnatural" or "awkward", even though it is grammatically acceptable (p.35). From the observation in my samples, though discourse markers are syntactically independent, they function as important cohesive devices connecting utterances. At least, DMs are not optional in all instances. While the syntactic independence is found to be the major criterion to distinguish their function from non-DM use. Likewise, Fung and Carter's (2007) conceptual framework for defining a "linguistic item or expression" as a discourse marker consists of five key criteria, namely, position, prosody, multigrammaticality, indexicality and optionality. For instance, *you know* in "Do *you know* what it is*?"* serves as a part of the question leading a noun clause, which differs from its DM use in "Math is difficult, *you know*?"

2.4.3 Semantic Features

Discourse markers have little or no propositional meaning, which also raises a debate between supporters and opponents. It is usually difficult for a translator to find equivalence for a DM in the target language (Lutzky, 2012, p.18; Jucker, 2002, p.210). Although the meaninglessness is widely accepted as a characteristic of DMs (Brinton, 1996; Erman, 2001; Schiffrin, 1987), their opponents think otherwise and argue that some DMs do convey meaning, e.g. truth-conditional implications such as *you know* (Lutzky, 2012). Fraser (2009) also criticizes the idea that a DM should be defined as an expression that "does not contribute to the truth conditions" (p.10). In the above example, *you know* with interrogative intonation in "Math is difficult, *you know*?" function as a question tag indicating a mutual knowledge or "you know what I mean?" Consequently, the semantic feature considered in this study is not to determine a universal criterion for all DMs. Rather, it is to provide a clue for defining the function by its semantic or pragmatic meaning. The meaning of DMs and their functions are "determined by the context" (Lakoff, 1973, p.466; Cf. Lutzky, 2012, p. 20; see also Aijmer, 2013, p.13).

Based on this notion, the idea of indexicality is proposed to associate discourse markers with the context (Fung & Carter, 2007 based on Aijmer, 2002; see also Aijmer, 2013, p.13-14). Indexicality of DMs refers to both textual and interpersonal dimensions. Deictic words such as adverbs only focus on structures or cohesion, while discourse markers (or pragmatic

markers) also index the socio-cultural elements and stance. It signposts "the relation of an utterance to the preceding context and to assign the discourse units a coherent link" (Fung & Carter, 2007). Further discussion of how to distinguish functions by identifying formal patterns will be presented in Section 2.5.

2.4.4 Functional Features

Discourse markers may be multifunctional, "operating on the local (i.e., morphophonemic, syntactic, and semantic) and global (i.e., pragmatic) levels simultaneously, as well as on different planes (textual and interpersonal) within the pragmatic component" (Brinton, 1999, p.35). Throughout the literature, researchers have tried to distinguish the functions of DMs from different perspectives with different categorizations of parameters, domains or planes. As a pioneer in the DMs study, Schiffrin (1987) analysed five different planes in her discourse model, including participation framework, information state, ideational structure, action structure, exchange structure (Cf. Maschler & Schiffrin, 2015, p.191). However, others are found to endorse fewer levels. For instance, M. A. Halliday (1979) and O¨stman (1981) both analyse three dimensions: the former simultaneously analyses the textual, interpersonal and experiential level; while the latter suggests "level analysis" including also three levels of coherence, politeness-modality and utterance structure (Cf. O¨stman, 1981, p.38-40; see also O¨stman, 1995, p.104; Aijmer, 2013). Brinton (1996, 2008, 2010) distinguishes only two levels, namely textual and interpersonal level. On the textual level, the DMs are used by speakers to "structure meaning as text, creating cohesive passages of discourse; it is 'language as relevance', using language in a way that is relevant to context"; while on the interpersonal level, they help to express stances and organize the social exchange (Brinton, 1996, p.38). Different though the descriptions are, their categorizations all cover textual and interpersonal levels. The next chapters will apply Brinton's mode to classify the functions because the functional patterns observed in the current study comply with the two levels.

2.4.5 Sociolinguistic and Stylistic Features

Lutzky (2012) points out that the sociolinguistic and stylistic features of discourse markers can only be referred to when analysing established texts. These features, however, are "controversial" (Brinton, 1996, p.35) and do not contribute to class-membership for a DM (Jucker & Ziv, 1998, p.4; Cf. Lutzky, 2012, p.23). In this study, the issue of gender will not

be discussed, since the profiles are not given and gender is an irrelevant feature in comparison between two language groups. This section will describe the features from the perspective of the effect of first language on the use of DMs by learners of English.

DMs are often considered in oral rather than written discourse, thus they associate with informality (see Table 2.1). Although text types in academic settings such as classroom discourse, oral tests, interviews, etc. are less informal than casual conversation, the nature of spontaneous speech links to the informality and speakers tend to use more DMs in informal contexts (Brinton, 1996, p.33). Additionally, DMs are stylistically stigmatised, which is interpreted as another aspect of informality in oral discourse. Nevertheless, taking such feature as "a sign of disfluency and carelessness" (Brinton, 1996, p.33) is inappropriate and overstated. It is a ubiquitous phenomenon that an expression is acceptable in oral discourse but not necessarily correct in written grammar (Brown, 2007).

Liu (2013) compares different uses of DMs between native speakers and Chinese speakers of English and finds that Chinese DMs have a certain influence on their English DMs use. Three Chinese DMs *wo je de*, *dui* and *a* are found equivalent to English DMs *I think*, *yeah/yes*, *ah*, which indicates the problems learners have in communicating in English. It is due to the different psychological status of speakers. As pointed out by Östman (1982; Cf. Aijmer, 2004, p.188), learners are expected to "express uncertainty or hesitation" by using DMs or clustering of markers, or intend to yield the floor to allow the hearer to finish the sentence (only backchannels are received in the current study); while native speakers tend to use DMs for "face-saving or to signal politeness."

Consequently, a revised list of basic features of DMs on different linguistic levels is presented in Table 2.2.

Table 2.2 Revised list of basic features of discourse markers based on Jucker and Ziv (1998, p.3) and Brinton (1996, p.33-35)

Linguistic Levels	Description
Phonological and lexical feature	a. They are short and phonologically reduced;
	b. They form a separate tone group;
	c. They are marginal forms and hence difficult to place within a traditional word class;
Syntactic features	d. They frequently occurred in utterance-initial and utterance-medial position, and have less frequency in final position;
	e. They occur outside the syntactic structure or they are only loosely attached to it;
	f. They connect the fragments of utterance as a cohesive whole;
Semantic features	g. They are indexical;
	h. Prototypical discourse markers have little or no propositional meaning; but less prototypical discourse markers do;
Functional feature	i. They are multifunctional, operating on several linguistic levels simultaneously;
Sociolinguistic and stylistic features	j. They are a feature of oral rather than written discourse and are associate with informality;
	k. They appear with high frequency;
	l. They are ubiquitous in oral discourse, however are stylistically stigmatised in written or formal discourses;
	m. They are used diversely by different groups of speaker for different reasons.

2.5 From Form to Function

2.5.1 Basic Approach

This section is going to distinguish two approaches related to corpus linguistics, namely the corpus-based and corpus-driven approach, and explain how the latter is applied in the current study. Tognini-Bonelli (2001) compares the characterizations of both approaches and

distinguishes them as follows: corpus-based approach is to "apply a scheme based upon a pre-existing theory", whereas corpus-driven approach is activated when the "scheme" applied to data is found to be deficient and requires an improvement or refinement (Cf. McEnery & Hardie, 2012, p. 150). Discourse markers are multifunctional (see Table 2.2, i), which may function different levels (on the local level, i.e. morphophonemic, syntactic, and semantic; and global, i.e. pragmatic) simultaneously, or one maker may function on different planes (textual and interpersonal) (Brinton, 2008, p. 35). In the current study, the multifunctionality is discovered in DM function of *you know* and *I think*. In example (01), the initial *I think* will likely be distinguished as an interpersonal function by the corpus-based approach, for studies (Baumgarten & House, 2010; Liu, 2013) have shown that I think uttered in the initial position tend to convey a deliberative meaning which is categorized as an interpersonal function. However, this case implies a more textual function, for it is followed by *you know* which serves at the textual level to indicate on-line planning.

(01)   s075: **I think you know** the mm the company break up is the is the main punishment

(C075yk80)

2.5.2 Related Theories

The current study applies two basic theoretical frameworks, namely Coherence and Relevance Theory. As is pointed out by Mu¨ller (2005), Coherence deals with textual functions, whereas Relevance Theory targets on cognitive processes (p. 8).

2.5.2.1 Coherence

Based on coherence, Schiffrin (1987) claims that "discourse markers provide contextual coordinates for utterances: they index an utterance to the local contexts in which utterances are produced and in which they are to be interpreted" (Cf. Mu¨ller, 2005, p. 8). In O¨stman (1981, pp. 37-41), she examines the functions of discourse markers at three levels based on Halliday's theory, including coherence, politeness-modality, and utterance structure. In her approach, coherence is "all kinds of 'semantico-functional' (ibid., p. 39) phenomena which collaborate to give as output a functionally acceptable and adequate text", which does not only refer to cohesive markers such as conjunctions but also includes markers that serve this function.

2.5.2.2 Relevance Theory

Relevance theory plays an important part in studying pragmatics which has the "potential to provide a theory of utterance interpretation which is consistent with generative grammar" (Blakemore, 2002, p. 7; Cf. Aijmer, 2013, p.11). Such a role can be illustrated by the process of inferences for interpreting utterances "while generative grammar provides an account of the coding-decoding mechanism in communication (Aijmer, 2013, p. 11). Based on the relevance-theoretic framework, Smith and Jucker (1998) find that discourse markers are "negotiating strategies of the common ground between speaker and hearer" (Cf. Müller, 2005, p. 9). Take *you know* as an example, Smith and Jucker (1998, p. 194) state that "you know invites the addressee to recognize both the relevance and the implications of the utterance marked with you know" (Cf. Müller, 2005, p. 155). Although their claim is not further expressed, it is made based on the relevance-theoretic framework. Müller (2005) compares Schiffrin's (1987, p. 318) notion with Relevance Theory that the general idea in the interpretation process of the hearer it to "find the most relevant interpretation" among all the potential meanings in the given context" (pp. 8-9).

2.5.3 Analytical Framework

Jucker and Ziv (1998, p. 4) argue that the linguistic levels in Brinton's list (see Table 2.1) are not weighted equally as crucial tests. Only the first three levels are diagnostic and the left two are predominantly descriptive. Discourse markers and their functions are defined regardless of the multifunctionality or polyfunctionality. On the other hand, multifunctionality only adds to the difficulties in distinguishing different functions (Aijmer, 2013, p.16). An utterance-initial discourse marker may function as a turn-taking device to introduce a new topic; however, other initial markers may also have the same function (ibid.).

The current study mainly examines the functions of discourse markers at the textual level and interpersonal level. At the textual level, discourse markers can serve as cohesive devices in discourse (Aijmer, 2013, p.32). Such pragmatic markers are called "coherent markers" which are the main organizers of discourse (Ostman 1995, p. 104; Cf. Aijmer, 2013, p. 32). "They are 'the grease' between the propositional parts of discourse making it work as discourse and they can create coherent 'locally' within the speaker's turn" (Aijmer, 2013, p. 32). At the interpersonal level, discourse markers are considered to mark "the relationship between speaker and hearer" (Müller, 2005, p. 31) and indicate "the attitude and

expectations of the speaker" (Ö̈stman, 1981, p. 39). Again in example (1), both I think and you know are examined at the textual level (on-line planning) and interpersonal level. The differences underline two points. In the first place, *I think* is used at the interpersonal level to mark a deliberative function as to express the speaker's opinion with confidence and certainty; whereas, *you know* is used as a cohesive device while preparing for the content utterance. Secondly, *I think* is followed by *you know* which somehow are combined as a cohesive device in searching for an answer of the question; whereas, *you know* also indicates the speaker's intention to involve hearer's inferences of the following utterance.

(01)   s075: **I think you know** the mm the company break up is the is the main
punishment

(C075yk80)

2.6 Summary

This chapter provides a definition that is suitable for the current study, as well as a discussion of the approach, theories and framework. The definition of a discourse marker is chosen by comparing proposals suggested by different researchers. In terms of the current study, the definition in Carter and McCarthy (2006) is selected for it involves both textual and interpersonal levels. The reviews of the previous studies shed light on the value of corpus in studying discourse markers and providing abundant resources to teachers and learners, which however may require high proficiency in actual application. The next two sections focus on the framework for distinguishing DM functions, which is exemplified by analysing one controversial instance.

Chapter 3 Methodology and Data

The current study will apply the corpus-driven approach with quantitative and qualitative analytical methods. The two sources of data are from the Corpus of Academic Writing and Spoken English (CAWSE) project (Chen et al., 2018) and Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, & Swales., 1999).

3.1 Data Overview

3.1.1 CAWSE

The CAWSE corpus is an on-going project, aiming to collect student's academic written and spoken English samples from the University of Nottingham Ningbo China (UNNC). In the current study, all the spoken data in CAWSE come from the Year One Programme in Centre for English Language Education (CELE) at UNNC, which contain 119 transcriptions of individual interviews with total tokens of 129,908 words. The individual interview is an oral test, consisting of three parts in which the student should answer questions from the teacher. The three parts are part 1 (4-5-minute general conversation), part 2 (1-minute preparation time and 2-minute speech) and Part 3 (further discussion on part 2). The total length of test time is 10 minutes. In the first part, the student has a general conversation with the teacher, including greetings, Q&A regarding daily life, hobbies, etc. The second part is a short speech in which the student shares ideas toward a given topic within two minutes. According to the record, the types of topic vary from common sense (e.g. medicine, advertising, media) to business knowledge (e.g. market research, business crime) and challenging topics (e.g. placebo effect, artificial intelligence). Various topics allow students to share their opinions and express the ideas that may require the mutual background knowledge between them and the teachers. All the data of interviews were selected in CAWSE corpus.

3.1.2 MICASE

The MICASE contains 1.85-million-word American spoken English data in 152 transcripts, within the academic context of Michigan University (Simpson et al., 1999). The corpus is available online and includes 15 different speech types, including advising, colloquiums, lectures, defences, discussions, interviews, presentations, meetings, seminars, and so on, which are uttered by various different speakers, from native speakers/near-native speakers

to non-native speakers. As a reference corpus, the MICASE corpus has the advantage of being able to provide sorting options of various types of speakers and plenty types of speeches. Therefore, it will be suitable to provide a comparable size of sample data to sample in CAWSE corpus.

3.2 Data Selection

The current study attempts to select the texts from the MICASE corpus as the reference sample to compare with the spoken data from CAWSE project. This section will present the process of data selection. Discourse should not only be seen "as a unit of language, but as a process of social interaction (Schiffrin, 2015, p.56). For this reason, the study of discourse markers is better to choose highly interactional data, such as interviews, discussions, study groups and so on. Less interactive or monologue data such as presentation and lecture will not be discussed in this study. Three criteria for data selection process are set up:

(1) All the data should be conversational or highly interactive data in the academic setting. Of all the 15 speech types, less interactive contexts such as small lectures, large lectures and student presentations are excluded, in which monologues take up a great proportion of the speeches. The left 12 types are listed in Table 3.1.
(2) Data with only native speakers of English in the MICASE corpus are selected. This is for the comparison between native and non-native DM uses. The data in MICASE may also include the speaker category of near-native speaker, which is acceptable in the current study, for the multi-cultural environment in the U.S.
(3) The total tokens of selected data in the MICASE corpus should be similar to that of CAWSE.

According to these criteria, the ideal speech type is supposed to be interviews that is closer to the text type in CAWSE corpus. However, only three interviews with total tokens of 13,015 words are sorted out, which seems not sufficient for comparison. Therefore, I decided to randomly select one transcript from each twelve speech types (excluding students presentations and lectures as noted before). Such selection is considered to be a representation of the academic activities in the University of Michigan (U-M) in Ann Arbor. The comparison within current study will be designed to provide a rough distinguish of the differences in DMs use between native speakers and Chinese speakers of English in the academic setting. Table 3.1 presents the basic information about sample data selected from two corpora:

Table 3.1 Overview of data selected from CAWSE and MICASE in regard to current study

	CAWSE	MICASE
Selected Speech Type	Individual interviews	Advising, Colloquia, Discussions, Defences, Interviews, Labs, Lectures, Meetings, Office Hours, Seminars, Study Groups, Service Encounters
Language Status	Non-native Speakers (L1 Chinese)	Native and Near-Native Speakers
Selected Transcriptions	119	12 (randomly chosen and one for each speech type)
Tokens	129,908	143,592

3.3 Methodology

As is reviewed in Section 2.4, the formal features are considered as the basic framework for identifying a discourse marker from the non-discourse marker uses. Although all the data should be examined line by line, features such as syntactical independent are regarded as a standard for identifying discourse markers in the preliminary study, in which the corpus tool, AntConc (Anthony, 2018), is applied to generate concordance lines with customized KIWIC sorting. Based on the preliminary findings, including instances of non-DM function, patterns of positions, clusters, etc., a more detailed qualitative analysis followed to further distinguish the discourse marker function for each lexical item. The detailed analyses were achieved by recognizing the patterns and analysing the function of DMs within the contexts.

3.4 Preliminary Study

In the first step of the preliminary study, five transcriptions of individual interviews in CAWSE were randomly selected and imported into the AntConc tools so as to generate a wordlist. By referring to the list of common DMs in Fung and Carter (2007), 16 words or phrases were found among the top 100 words presented in the wordlist. I then checked the concordance

lines for each word and phrase to identify their discourse marker uses. Phrases like *you know* and *I think* were checked separately by searching directly in the concordancer. Apart from the most frequent marker *and*, *you know* and *I think* were noticed with extremely low and high frequency as shown in Table 3.2. Thereafter, the preliminary study moved to the next step.

Table 3.2 Discourse markers in CAWSE identified in preliminary study

Discourse Markers	Frequency	Per 1000 words	Rank1
Total words count	3525		
and	79	22.4	1
I think	43	12.2	2
so	24	6.8	3
because/cos	23	6.5	4
but	17	4.8	5
yeah	17	4.8	6
just	15	4.3	7
ok/okay	14	4.0	8
yes	9	2.6	9
like	5	1.4	10
or	4	1.1	11
really	4	1.1	12
now	2	0.6	13
you know	1	0.3	14
actually	1	0.3	15
oh	1	0.3	16

The second step was to search the concordance lines for the two lexical items, *you know* and *I think* within the whole selected sample data in both corpora. Again the concordance lines of each phrase was checked. The frequencies of *you know* and *I think* in the DM function are presented in Table 3.3. The frequencies of these two lexical items between the two corpora samples were tested for statistically significant difference using Paul Rayson's online log-likelihood calculator (http://ucrel.lancs.ac.uk/llwizard.html). The results shows that the discourse marker function of *you know* is used significantly less often by non-native speakers than by native speakers; while *I think* is used significantly more often by non-native speakers.

The preliminary study stopped here since I have discovered a starting point for the current study, which is to identify the discourse marker function of *you know* and *I think* and compare their uses in the two corpora.

Table 3.3 Occurrence of *you know* and *I think* in CAWSE and MICASE respectively

		CAWSE		MICASE	
		N	%	N	%
You know	Occurrence of *you know*	95	-	676	-
	Occurrence of *you know* in DM Function	67	70.53%	594	87.87%
I think	Occurrence of *I think*	1179	-	339	-
	Occurrence of *I think* in DM Function	1148	97.37%	301	88.79%

Chapter 4 Analyses and Findings

4.1 Introduction

Studies have shown that both learners and native speakers use discourse markers, the major difference underlines the frequency of individual DMs (Fung, 2003; Fung & Carter, 2007; Liu, 2013). After a rough pilot test in the previous chapter, the frequencies of two discourse markers *you know* and *I think* were calculated by times of occurrence (excluding non-DM use) per 1000 words. The frequency of *you know* used by L1 Chinese speaker in CAWSE corpus is 0.5 per 1000 words, which is much lower than the native speakers in MICASE corpus, with a frequency of 4.1 per 1000 words. Interestingly, *I think* is uttered by L1 Chinese speaker with a frequency of 8.8 per 1000 words, which is more than fourfold the frequency of 2.1 per 1000 words by native speakers. The results show that both DMs are used with significant differences between the two corpora (both p < 0.0001). In comparison to the native speakers, *I think* is used significantly more often, whereas the native speakers tend to use *you know* more often.

4.2 You know

4.2.1 Previous studies of *you know*

Since the early-years studies of *you know* (see Ö̈stman, 1981; Schiffrin, 1987), it has been considered one of most typical and versatile items in the heterogeneous set of forms (Buysse, 2017; M. A. Halliday, 1979; Mü̈ller, 2005; Ö̈stman, 1981). The main challenge for study *you know* or other multifunctional DMs is to identify the "disparate functions" (Tree & Schrock, 2002, p. 736; Cf. Buysse, 2017). In this section, the previous studies will be reviewed from the basic meaning of *you know,* categories of *function* to its formal features.

4.2.1.1 Basic meaning of you know

You know as a DM "occurs almost exclusively in more informal" face-to-face spoken contexts (Ö̈stman, 1981, p.19). A speaker tends to use it, particularly in narration in the conversation, which makes no difference to the basic propositional meaning (p.16). And the literal meaning of *you know* will directly influence the functional use in discourse (Schiffrin, 1987, p. 26). The general meaning of you know, as is suggested by Ö̈stman (1981, p. 17), is:

The speaker strives towards getting the addressee to cooperate and/or to accept the propositional content of his utterance as mutual background knowledge.

He then further explicated that the speaker will (1) use *you know*, regardless of the fact that the addressee may have no idea about the content after *you know*; (2) attract the addressee's attention by involving 'you' in the conversation, cueing the addressee's intention of "a direct plea for cooperation"; (3) and imply the addressee to presuppose the plausibility of his/her utterances and appeal for understanding (pp. 17-19).

Jucker and Ziv (1998) argues that *you know* is used to invite addressees inferences, which at the same time requires shared knowledge as stated in Östman's proposal (Tree & Schrock, 2002, pp. 736-737). Tree and Schrock also point out that the latter proposal is more precise that it seems that the addressees are taking charge of making the "right inference", which emphasizes more on the interpersonal role of *you know* in discourse marker function.

4.2.1.2 Categories of functions

Linguists have been trying to fit *you know* into word classes of "interjections (James, 1973, p.18), hedges (G. Lakoff, 1975, p. 53), turn-taking devices (Hinton & R. Lakoff, 1979-80)" (Cf. Östman, 1981, p.19). Östman (1981, p. 16) argued that the functions of you know should be discussed with different situation and contexts. She distinguished the interpersonal functions of you know based on its position and prosody, namely, general meaning, stylistic marker, turn-switching marker and a marker of Politeness and hedge, and labelled them as the "Coherence Level" and the "Politeness-Modality Level" (Östman, 1981, pp. 39-40). The speaker tends to have less frequent use of *you know* when talking to an addressee who shares a closer relationship with him/her, which was observed by Östman (1981) in her dinner-table conversations. It is implicated that the power of speech and politeness strategies are embedded in the use of *you know*. This function suggests that the speaker want to draw closer to the addressee by using *you know as a stylistic marker* (p.19-20). *You know* can function as a lexical hedge that refers to the implicit transmission of information, as well as signpost the speech act of hedging in the utterances (Erman, 1987). The sense of hedge in *you know* is not at the same level of "fuzziness" (R. T. Lakoff, 1972; Cf. Östman, 1981) as that of *sort of* or *kind of* which modifies the following content to certain extent. It is the "conversational indirectness" (Östman, 1981, p. 9) that a speaker uses you know to create when lacking in planning, but without altering the propositional content.

Erman (1987) distinguishes the functions of *you know* at two levels, namely, the micro-level and the macro-level, claiming that the frequencies of the functions are equal at both levels. At the micro-level, *you know* appears within the sentence; while at the macro-level, it is examined within the context (Erman 1987, p.114; Cf. Mu¨ller, 2005, p. 151). However, some of her findings are lacking in support. Mu¨ller (2005) argues that only four functions, such as "introduce an exemplification", "introduce a clarification of some part of a previous statement", "introduce background information" and "mark the boundary between" direct and quotative speech, are supported by other studies (Ostman; Erman 1992, 2001; Shiffrin, 1987; Muller). In her later studies, Erman (1992) distinguished between textual and interpersonal level based on Halliday's theory. *You know* as one of the heterogeneous linguistic items "conveys meaning at a number of different levels simultaneously or expresses several functions at once (Halliday 1973; O¨stman, 1981; Cf. Erman, 1992). At the textual level, Erman mentions turn-taking, floor-yielding and planning; at the interactional level, she distinguishes the following functions between two broad categories, in consideration of the degree of certainty that the speaker expresses in utterances.

Partially based on Erman's categories, Mu¨ller (2005) also examined the discourse marker functions of *you know* at two levels, namely the textual level and the interactional level (see Table 4.1.1). Five functions each at the two levels have been discovered in her data. In terms of this study, all five functions at the textual level and two from the interactional level (as the interpersonal level in this study) have been found useful to classify my own samples. According to Tree and Schrock (2002), *you know* plays an important role in achieving both positive and negative politeness. The negative politeness is achieved when the speaker insert *you know* before introducing shared knowledge purposely to "invite addressees' interpretations" (Tree & Schrock, 2002, p. 737). While the positive politeness is also achieved by such shared knowledge, drawing the addressees closer to the topic (ibid., p. 738). Tree and Schrock (2002) state that it is the context that produces a consequence of politeness, not merely by using a DM. Consequently, the politeness is not taken as a category of DM function in the current study. Additionally, another two categories at the textual level from the literature are added so as to achieve a complementary distinction (see Table 4.1.2). The description of each function will be presented in the following sections.

Table 4.1.1 Discourse marker functions of *you know* in Mu¨ller (2005, p. 157)

Textual Level:

 – marking lexical or content search

 – marking false start and repair

 – marking approximation

 – introducing an explanation

 – quotative you know

Interactional Level:

 – "imagine the scene"

 – "see the implication"

 – reference to shared knowledge

 – appeal for understanding

 – acknowledge that the speaker is right

Table 4.1.2 Adjusted list of discourse marker function of *you know* according to sample data from CAWSE and MICASE (Based on Mu¨ller, 2005, p. 157)

Textual Level:

  ➢ Initiating a new topic
  ➢ Marking lexical or content search
  ➢ Marking false start and repair
  ➢ Marking approximation
  ➢ Introducing an explanation
  ➢ Transition to reported speech
  ➢ Floor-yielding

Interpersonal Level:
  ➢ Securing comprehension
  ➢ Referring to shared knowledge

4.2.2 Formal Features of *you know*

4.2.2.1 *You know* and positions

Fung and Carter (2007) conclude that most of the DMs start in the utterance-initial position; while some DMs are more flexible, which can "be inserted in utterance medial position" or in the utterance-final position. In her investigation of you know, Erman (1987, 2001) finds that you know occurs mostly in the medial position of an utterance in her data, thus she concludes that the turn-switching device is not the main function. As is reviewed in Section 1.1, the DMs will be discussed at the utterance level (Lutzky, 2012). Due to the ambiguity of marking the boundaries of utterances, an utterance in this study refers to the lexical items or chunks uttered between two pauses which include filled pauses, such as *ah, en, er, erm, huh, mhm, mm, oh, uh,* and *uhu (*no other fillers are transcribed in CAWSE project, see Appendix-convention for details). DMs in the utterance-initial position often function as "marking boundaries of talk", while the medial position where DMs are inserted to hold the floor or to clarify meaning. DMs appear at least frequently in the utterance-final position and "are understood as comments […], clarification […] or as an afterthought […]" (Aijmer, 2013; Fung & Carter, 2007, p. 28). However, the boundaries of utterances in conversations or informal speeches are sometimes complicated and not clear enough to figure out. Thus, the integrity of the utterance's meaning should be considered. For instance, initial positions such as *you know* in Example (00), where it serves as a connection to the preceding utterance, although it is preceded by a pause (marked by a comma, indicating a 1- 2 seconds unfilled pause, see Appendix convention). In this case, the position where *you know* occurs will be categorized as a medial position.

> Example (00)
> so we don't go off talking about, **you know** the hockey game or something for forty minutes […]
>
> (MINTyk011)

According to Table 4.2, the total occurrence of *you know* as discourse marker function in CAWSE corpus is 67 times, as 594 times in MICASE. Three types of position are examined, namely utterance-initial (including utterance-initial *you know* after a preceding filler), utterance-medial, utterance-final position. It shows that the frequency of *you know* in medial position uttered by Chinese speakers is much higher than by native English speakers. While the native speakers tend to use you know in both initial and medial positions.

Table 4.2 Distribution of *I think* in three positions by L1 Chinese speakers in CAWSE and native speakers in MICASE

Utterance Position	CAWSE		MICASE	
	N	%	N	%
Initial	18	26.87%	247	41.58%
Medial	47	70.15%	274	46.13%
Final	2	2.99%	73	12.29%
TOTAL	67	100%	594	100%

Note: Initial incudes *you know* that is preceded by filled pauses (ah, er, erm, etc.) and markers (yeah, but, so)

4.2.2.2 *You know* and pauses

O¨stman (1981) analysed *you know* that preceded and followed by pauses. The pauses in O¨stman's refer to both filled and unfilled pauses, including hesitation markers (p.28) such as *erm, er*, etc. If *you know* is followed by a pause, the speaker is trying to search for a "particular lexical item"; however, if *you know* is preceded by a pause, the speaker is uncertain about the propositional content that he/she is going to utter (p.29-30). Table 4.3 shows that the L1 Chinese speakers in CAWSE used too many filled pauses, which may indicate their lack of planning and lower level of fluency than the native speakers in MICASE. Interestingly, the ratios of *you know* accompanied by pauses per total occurrences of DM function *you know* are close to each other.

Table 4.3 *You know* with pauses before or after in CAWSE and MICASE

Types of Pause	CAWSE		MICASE	
	N	%	N	%
Unfilled Pauses	3	4.48%	279	46.97%
Filled Pauses	33	49.25%	50	8.42%
No pause	31	46.27%	265	44.61%
TOTAL	67	100%	594	100%

4.2.2.3 Cluster and Collocation of *you know*

The Cluster of DMs shows a similarity in functions of two or more DMs (Aijmer, 2004, p. 185). It is different from collocations (which forms a single marker, e.g. *as you know*) and has no fixed sequences among the internal words (ibid., p. 185). From the sample in my data, I have discovered occurrences of clusters such as *like you know* and *you know like* in both corpora (see Table 4.x for other examples). It is inferred that native and non-native speakers make use of DMs clusters to buy time for lexical or content search, to start a new topic or to repair the previous utterance (ibid., p. 186). On the other hand, Aijmer (2004) also points out that the DMs cluster only functions as textual cohesion prevailingly in learner speech, while native speakers tend to reinforce the interpersonal functions. In this study, *you know* is find frequently co-occurring with *and, but, because, like, er, uh, um*, etc.

4.2.3 Non-DM function of *you know*

Identifying the non-DM use is the first step in my observation of the concordance lines. As argued by Mu¨ller (2005), Schiffrin's (1987) analysis did not make a clear distinction between non-discourse marker uses and discourse marker uses, such as "You know why?" (Cf. Mu¨ller, 2005, p. 153). In her study, Mu¨ller (2005) further exemplified with examples of her own data that some instances of *you know* are excluded as non-DM function from the previous studies, for example, *you know* in questions such as *do you know*, and *you know* connecting a complement clause (e.g. *you know that she likes him*) (pp. 157-158; see also Redeker, 1991; Biber et al., 1999). Apart from the aforesaid instances, four other non-DM function types of *you know* are discovered in this study as follows:

(1) Part of statement

Example (01a)
i know you guys think **you know** everything but you don't.

(MTOUyk534)

(2) Part of question

Example (01b)
S4: but **how do you know** which is the right one if you think <LAUGH> you

(MSEMyk155)

Example (01c)

you know how you like, when you're young

<div align="right">(MDISyk255)</div>

Example (01d)

S3: **you know what**? i want it. i'm taking it home...

<div align="right">(MSVCyk410)</div>

Example (01e)

the reason why i say that about the, printer question, is cuz like, they all supposed to be working, **you know** what i'm saying?

<div align="right">(MSVCyk641)</div>

(3) Preceding complement clause

Example (01f)

and **you know that** you need to elect either Great Books or Classical Civilization

<div align="right">(MADVyk052)</div>

(4) Part of a phrase

Example (01g)

t: which of these intelligence would be important if you (1.0)

s106: erm maybe the interpersonal intelligence

t: uhu

s106: **as you know** erm we when we have deep talk to other people and we have to understand them well and we have to erm use this intelligence to express ourselves

<div align="right">(C106yk009)</div>

In addition to Example (01e), *you know* also occurs in question *you know what I mean* in MICASE samples. Interrogative sentences of *you know what I'm saying* and *you know what I mean* have one thing in common that they always occur at the final sequence of a set of utterances, checking the hearer's status of understanding. *You know* is not syntactically independent in these instances (same in 01f), thus according to Brinton's List of DM features in Table 2.1 & revised Table 2.2, *you know* in these instances will not be regarded as DMs. The last Example (01g) presents *you know* in the phrase of *as you know*, which also occurs in the MICASE corpus. Although you know do have a function which can be interpreted as the phrase *as you know* and used to introduce the already mutual background knowledge, they are not equivalent (Ö̈stman, 1981, pp. 24-25).

4.2.4 DM function of *you know* at textual level

4.2.4.1 Initiating a new topic

You know can be used as a turn-switching marker in a conversation, functioning as a cohesive device and a "pre-starter" (Ö¨stman, 1981, p. 24), like *well, so, OK*. In Ö¨stman's (1981) point of view, the difference between *you know* and other turn-switching markers is that the turn-taking use of *you know* indicates that the speaker comes up with a new idea at that very moment, in the meantime, implies a reference to mutual knowledge. In Example (02a) the two speakers are talking about people tend to delay a response to an Email. S48 agrees with S52 and further explains that it is not possible for him to write Email to so many people all the time. Then S52 takes the turn with an initial *you know* to switch the topic to his leaving, in which *you know* functions as a pre-starter to end the previous topic and introduce a new idea that the speaker just comes up with. This function can be replaced with the question "you know what?" according to Ö¨stman (1981, p. 25). The turn-initial *you know* also marks a connection to the preceding turn and to the following one (ibid.). In this study, not only turn-initial but also utterance-initial *you know* are discussed in this function which may follow markers such as yeah, OK/okay, etc. Example (02b) shows an initial *you know* preceded by *okay*, a pause filler, and followed by a short pause (marked by a comma, referring to a 1-2 second pause in MICASE corpus) indicating that the speaker S1 is trying to get the attention from S12 and simultaneously introduce a new topic (S1 wants to comfort S12). In (02c), *you know* preceded by a lengthened *ah:* also indicates a new idea.

> Example (02a)
> S48: i have like eighty people on my address list (like i can't write them all the
>   time.)
> S52: you can, [S48: yeah] even if it's just a uh, <LAUGH>
> S48: i just do it when i'm at work.
> S52: **you know**. cuz i think i'm leaving here i think i'll go grab my car and uh, go to
>   this review session.
>
> <div align="right">(MSVCyk391)</div>
>
> Example (02b)

S12: i mean, i don't really believe in the whole point of view thing to begin with [S1: okay.] but i'm saying just, like what i was saying was just kinda in response to your whole dream thing.

S1: okay **you know**, that's okay. you don't have to_

<div align="right">(MSEMyk334)</div>

Example (02c)

S097: er: I'm in Vis-A-Vis I'm a member of Vis-A-Vis it is a organization which ah: **you know** my our slogan is to break the barrier

<div align="right">(C097yk001)</div>

Example (02d)

S3: yeah i don't picture like a [S4: **you know**] big you know in and out in [S4: **you know**] and out heat exchanger or, [S4: right. if between,] reverse osmosis, or <S2: LAUGH>

S4: if if between here and the final report we find something, you know, [S3: better] that would, yeah sounds better, we go with that.

<div align="right">(MSGRyk387) & (MSGRyk388)</div>

The last example is different from the former ones. The speaker S4 fails three times (twice by *you know* and once by *right*) when trying to self-select the turn. Not until cut-off by laughter, does S3 continue to maintain the turn in regardless of S4's signal for a new topic. Finally, S4 find a chance to attain the turn, which can be inferenced from the beginning of his/her utterance where he/she says "if between" that continues after "right. If between" (see the underlines). Although failed, *you know* here has the same function as the other instances in this category.

4.2.4.2 Marking lexical or content search

Lexical or content search refers to a struggling situation that the speaker uses *you know* while the conversation goes to uncertainty of a word or thinking about what to say next (Mü̈ller, 2005, p. 158). Many studies (Erman, 1987, p. 124; Mü̈ller, 2005, p. 158; Ö̈stman, 1981) have stated to discover the relation between pauses and this function. *You know* which is preceded by a pause or filled pause functions as content search, on the contrary, *you know* functions as lexical search when followed by a pause (Ö̈stman, 1981). Mü̈ller (2005) argues that though Ö̈stman's idea is confirmed in many instances (see Example 03a), still exceptions

can be found. Thus it is not necessary to distinguish the sub-functions. As is suggested in Section 4.2.2.3 and Table 4.2, *you know* co-occurs with various lexis, such as *and, but* in this function category which is a coordinating conjunction and functions to continue a cumulative set (Schiffrin, 2006). In Example (03b), *and* connects the preceding idea of the speaker to the following utterance, in which *you know* is used to hold the floor while the speaker searches for the content for speech. In Example (03c), there is a pause (marked by comma) before *and*, indicating the speakers is thinking about what to say next.

Example (03a)

t: okay good so how could graffiti be controlled in the future

s056: en en I think government should er (1.0) **you know** er (1.0) en en en give the permission to er er (1.0) graffiti

(C056yk040)

Example (03b)

I think the crime in business is er should draw attention from the society from the public and **you know** er if er the mm there's always some cheat behaviour or something like the er scandal

(C111yk002)

Example (03c)

S1: so when you've done the analysis, we were realizing and you've got all these data, and **you know** about the speech patterns and those kinds of things, do you do anything else with the information?

(MLAByk023)

In (03d), the *and* preceding *you know* represent another function which is a conjunction coordinating "Public Enemy" and "Chunk D", judging from the context. In this way, you know is inserted when the speaker is recalling the word said before in which he/she is suddenly out of memory.

Example (03d)

s1: it was presented to them by Chuck D and Public Enemy. [S2: Speaker information restricted] and the rest of th- Public Enemy and **you know** and and Chuck D's f- publicly gets up

(MOFCyk025)

4.2.4.3 Marking false start and repair

Terminologies such as "textual monitor" (Erman, 2001), "error correction" (Jefferson 1974, p. 186; Cf. Mü̈ller, 2005, p. 160) and "editing marker" (Buysse, 2017, p. 48) have been used to describe the "false start and repair" function of *you know*. Editing marker in Buysse (2017) is a broader concept, covering the previous function, lexical or content search. However, I prefer to separate them into two functions, for the differences that Mü̈ller distinguished in her study can also be identified in mine. In my samples, two types of cases are distinguished. The first type is repetition without any correction, in which repeating words usually do not exceed two (Fox & Jasperson, 1995; Cf. Mü̈ller, 2005, p. 161). Example (04a) and (04b) present two typical instances that *you know* is inserted between two "I wanna" in (04a) and "the" in (04b).

> Example (04a)
> and i think a thesis can be good but um, i don't want to go into research, i wanna
> **you know** i wanna do something
>
> > (MINTyk612)
>
> Example (04b)
> there're a lot of there are other crime and corruption like the white collar crime and
> er which is no one will ask the **you know** the past scandal or other other some other
> things
>
> > (C051yk078)

The second type includes the instances in which the speaker repairs his/her utterances because of the uncertainty of a word or content. Mü̈ller (2005, p. 161) points out that if the utterance following *you know* "continues with a different syntactic structure", it should not be treated as a repair. In 04c, the speaker doesn't finish the word *cultu-* and soon insert *you know* with repetition of *in another* to lead a different expression. However, the example (04d) violates Mü̈ller's assumption that you know can also function as a repair in a different syntactic structure. Because of the repair of the content, the syntactic structure may be altered sometimes (even if the structure is not grammatically correct). Thus it seems overbold to define such a constraint.

Example (04c)

S7: it's kind of like the situation in Waco. because they thought they were living by, they were living by their own rules and stuff in another cultu- **you know** in another [S1: yeah] set of consequences so, i mean it's kind of,

(MDISyk115)

Example (04d)

um, one step those messages reach the suppliers if the if the suppliers are really, **you know**, are ar- agree that they are quiescent

(MMTGyk371)

4.2.4.4 Marking Approximation

Speakers tend to use you know as a marker of approximation when uncertainty underlines "the linguistic precision of the message" (Cf. Mu¨ller, 2005, p. 163; Stubbe & Holmes, 1995, p. 69). In this case, the speaker uses *you know* to introduce an approximate description of a word or expression and strive to approach to the exact word (Buysse, 2017, p. 51). Buysse (2017) explicates that such function is usually signposted by co-occurred expressions such as *and all that stuff* (05a) and *or something* (05b). In my samples, they also co-occur with *something like that* (05c) and *things like that* (05d). This function only occurs x times in my sample (in CAWSE and in MICASE) and is also stated to be one of the functions with less frequency (Buysse, 2017; Mu¨ller, 2005).

Example (05a)

and Western thought's very like individualistic and **you know** personal rights and all that stuff so.

(MDISyk044)

Example (05b)

S52: i didn't know that. i'm up here thinking she's, i don't know an English major or something, **you know**. hey, i ain't dissing you an English major?

(MSVCyk462)

Example (05c)

but i- it wasn't just difficult for you to see it was difficult for [S2: yeah] you know difficult for everyone to actually get it in perspective or [S2: mhm] or to see that it had consequences or **you know** something like that

Example (05d)

what really is the relationship between the warrior and his, and his leader? **you know** <u>things like that</u>,

(MADVyk267)

4.2.4.5 Introducing an Explanation

You know is also discovered to signal an explanation of a preceding knowledge or to provide a reason for the preceding idea or action. Different terms or descriptions have been found throughout the literature: Erman (2001) describes it as introducing a change of information content and modifying previous discourse, Tree and Schrock's (2002) clarifying a prior utterance, Müller's (2005) introducing an explanation and Buysse's elaboration of a preceding concept (2017). The types of forms in this function include clarification (06a), paraphrase (06b) and exemplification (06c) (Buysse, 2017, p. 46). In this study, Müller's term is adopted for it can define a wider functional use of *you know*. For example in (06d), S3 wants to see S1's notes of calculated numbers and S1 in the fourth turn is explaining why he/she is not (willing) to offer. S3 however laughs when he/she sees what is on the notes. The laughter is questioned by S1, repeating "yes?". Then S3 stops laughing and insert an unintelligible *you know* in the last turn to mark an explanation of the action (keeping the mouth shut) for the laugh which may annoy S1.

Example (06a)

S1: so when you've done the analysis, we were realizing and you've got all these data, and **you know** about the speech patterns and those kinds of things, do you do anything else with the information?

(MLAByk023)

Example (06b)

i can like be involved in and like um controversial things that, **you know** debatable topics not just something that's like, like doing math,

(MADVyk488)

Example (06c)

because the crime the corruption has already er rai- er rise our concerns and **you know** recently a lot of policies has have been made to stop these things er happens again

Example (06d)

S3: enter. <PAUSE:09> Tighe do you have all those numbers? from, when we calculated how much, how many gallons per batch, [S1: i do] and how many pounds of this stuff (we're gonna have?)

S1: i do.

S3: can i see them please?

S1: maybe. i'll think about it. (xx) brewing preparations (xx) i wish i had a, glossary around here or something. um, you're gonna have a difficult time following my notes here. let's see here. <S3: LAUGH> yes? [S3: okay] yes? okay yes? <S2: LAUGH> of course. i have it. <LAUGH> um, <MUMBLES>

S3: i'm, (**you know**,) keeping my mouth shut.
<PAUSE:07>

(MSGRyk287)

4.2.4.6 Transition to reported speech

You know is capable of marking a transition to reported speech (Buysse, 2017, p. 52; see also Redeker, 1991, p. 1163) which in Müller's (2005, p. 167) term is quotative mark. As is concluded by Mu¨ller (2005), this function can mark direct-speech or indirect-speech report from the speaker's own talk or a third person's. She also points out that the speaker tends to change the voice quality to a higher or lower pitch when quoting a third person's talk (2005, p. 168). Additionally, you know often co-occurs with other quotative markers such as *think, say/said*, and *BE+like*, among which *BE+like* appears to be the most frequent (2005, p. 169). Moreover, it is sometimes hard to identify whether you know is a quotative marker or is part of the reported speech. The (07b) shows an example of such ambiguity that the speaker is predicting what the third (he) person would have said because the speaker "haven't seen him" before the quotation, in which you know is defined as a quotative marker rather than a part of the reported speech. Xx instances of *you know* as a transition to reported speech in MICASE corpus but none is discovered in CAWSE.

Example 07a

most of us would think, **you know** what's, what's wrong with this person?

(MDISyk538)

Example (07b)

S72: (is) (xx) (here?)

S52: haven't seen him, i think what he said was, **you know** i need a longer break, i think i'll get here about ten.

<div align="right">(MSVCyk613)</div>

4.2.4.7 Floor-yielding

Turn-final *you know* with declarative intonation (usually followed by a long pause) may realize the true floor-yielding function and can be used to close off a point, as in ''I won't say anything more'' (Ö¨stman, 1981, p. 27). The speaker is expecting the next speaker to take over the turn during the long-time pause. In (08a), S3 is answering S1's question by referring to the computing method and finish the utterance by using you know in the final position, indicating the shared knowledge. Indeed, S1 retrieves the turn and starts to explicate the computing process.

Example (08a)

S1: oh what's the

S3: that's all they did to scale it up [S2: really] it's just a, **you know**,

S1: it's thirty-nine-thirty-six minus the volume of the slurry, gives you the initial, volume of, it gives you [S4: of] the volume of the water

<div align="right">(MSGRyk003)</div>

4.2.5 DM function of *you know* at interpersonal level

Speakers tend to use *you know* as an interpersonal function to show their willingness in appealing to "addressee inferences" (Tree & Schrock, 2002, p. 737). In detail, the speakers' intention can be inferenced from the contexts, including the types of community, the intimacy of relationship between the speaker and hearer and topics in discussion. Tree and Schrock (2002) states that the frequency of *you know* tends to be higher when talking to friends than stranger. Secondly, it may be common to hear *you know* in "opinionated talk" and casual talk. In this study, both corpora contain speeches in the academic setting, i.e. university. The differences are: the samples from CAWSE corpus are one-to-one interviews which are oral exams between teachers and students, while those of MICASE consist of more complex relations of participants, including students, faculty and staff. Although interviews are those types of talk that interviewees can express his/her own opinions, the exams are supposed to

have the interviewees to "fill out all their idea" (Tree & Schrock, 2002, p. 737) rather than leave their ideas to be interpreted by the interviewers (also examiners in this case). Only 7 out of 67 occurrences of *you know* are identified at this level, in which all of them are used to refer to shared knowledge.

4.2.5.1 Securing comprehension

You know in utterance-medial and final position can function as a comprehension-securing device to elicit "backchannels from addressees while maintaining the floor" (Tree & Schrock, 2002, p. 732; see also Müller 2005). Additionally, *you know* can work as a potential floor-yielding function which is usually uttered at the final position usually accompanied by an interrogative intonation contour, meaning that the speaker expects a response from the hearer (Buysse, 2017; O¨stman, 1981, pp. 26-27). Here "potential" is used because the speaker tends to use you know in the utterance-final position to secure the hearer's comprehension but not actually give the turn away. Often the speaker self-selects and continues the utterances after the hearer's response, i.e. backchannel or acknowledgment (but not always necessary) (Tree & Schrock, 2002, pp. 739-740). Instances that the speaker fails to hold the turn can be found, for example in (09b). The speaker S48 receives respond from the hearer S52, meaning that what uttered previously is understood. However, S48 seems surprise that he/she is understood before the information of age is stated. In addition, S52 continues the turn "and you're like eighteen nineteen" to prove the comprehension. As suggested by Buysse (2017), this function reflects the speaker's uncertainty at two different levels: either checking the mutual knowledge hold by the hearer (09a), or the speaker's "accuracy and clarity" (p. 51) of speech (09b) (see also Tree & Schrock, 2002, p. 737). This is different from the floor-yielding function at textual level, in which the speaker (in Example 08a) use *you know* to simply turn the floor to the hearer (Buysse, 2017; Tree & Schrock, 2002). Again, this function is only discovered in the MICASE corpus with totally 22 instances.

> Example (09a)
> S1: and um with a- kind of a, spirit of adventure, i mean he's_ he doesn't seem beleaguered **you know** [S2: mhm] really weighed down by all that he's um responsible for
>
> > (MINTyk076)
>
> Example (09b)
> S48: it's not typical to get along with your parents at this age (**you know**)

S52: <u>yeah, yeah</u>, [S48: y- you know what i mean (kinda) (xx)] and you're like eighteen nineteen?

S48: eighteen.

<div align="right">(MSVCyk018)</div>

4.2.5.2 Referring to shared knowledge

The function of referring to shared knowledge is to insert an utterance-medial *you know* to highlight a particular element, idea or content within a clause (Buysse, 2017, pp. 47-48) rather than to introduce an entire clause as *you know* does at the textual level (see Section xx initiating a new topic). The basic meaning (Schiffrin, 1987; Tree & Schrock, 2002) or prototypical meaning (Ö¨stman, 1981) of *you know* underlines the shared knowledge implied in the speaker's utterances (see previous Section xx for basic meaning of *you know*). In (10a), for example, the speaker inserts *you know* and highlights "Bill Gates" as a reference to one case of "lots of people in the world" who have lots of money but still work, indicating "Bill Gates" is a common ground between the participants. On the other hand, *you know* also draws attention from the hearer because the element highlighted by you know is supposed to be a common ground, even if the information has not been stated previously or is not a shared knowledge of the hearer. In this case, the speaker "merely appeals to the interviewer's knowledge of the world and empathic capacity" (Buysse, 2017, p. 48), which in Mu¨ller's (2005) term is "appeal for understanding" (p.181). *You know* in Example (10b) highlights the speaker's feeling "enjoy" as to refer back to the previous utterances of admiration for a singer, in which the highlighted feeling and the stance implied in it is what the speaker appeals for understanding.

Example (10a)

but um there's lots of people in the world, take Bill Gates **you know** for an example, who um, make lots of money, have lots of money or other people who inherit lots of money. they don't have to work, but they do.

<div align="right">(MDISyk188)</div>

Example (10b)

S1: i i think that um, the, style in which she's doing her music and the, instrumentation and and the um, the, her hispanocentrism, i think is something which is unusual, and it is something which is um, to be commended. and that's what i **you know** enjoy

<div align="right">(MOFCky218)</div>

4.2.6 Summary

By analysing all the instances of *you know* in the selected samples from CAWSE and MICASE corpus, nine categories of discourse marker function are distinguished. Most of the instances (nearly 90% in CAWSE and 70% in MICASE; see Table 4.5) of *you know* function at textual level. As is shown in Figure 4.1.1 and 4.2.2, on average, native speakers use you know in discourse marker function nearly eight times as much as the L1 Chinese speakers do. More importantly, all of the 9 functions present a statistically significance ($p < 0.0001$, with only $p < 0.05$ in XPL and $p < 0.001$ in TRS, see Table 4.4 for abbreviations). Native speakers tend to use *you know* in interpersonal functions in comparison to L1 Chinese speakers. On the other hand, the most frequent function in both corpora is lexical content search, which may due to the informal contexts or lack of planning. For example, the speakers in CAWSE corpus attended face-to-face interviews as an exam, in which they would be asked random questions that are difficult to prepare beforehand and might be nervous than usual. Whereas, the MICASE samples consist of almost all types of spoken data (excluding lectures), which are supposed to provide with a general status of conversation happening in the campus. In other word, these native speakers tended to be more relax when expressing ideas.

In spite of the psychological features, L1 Chinese students are lacking in the knowledge of the discourse marker uses, which may be one of the important reasons to scale up the proportion of lexical or content search function. In Table 4.4, for example, there is no occurrence in functions such as "transition to reported speeches", "floor-yielding" and "securing comprehension" by Chinese speakers.

Table 4.4 Adjusted list of discourse marker function of *you know* with abbreviated label (Based on Mu¨ller, 2005, p. 190)

Full Label	Abbreviated Label
Textual Level:	
➢ Initiating a new topic	INT
➢ Marking lexical or content search	LCS
➢ Marking false start and repair	FSR
➢ Marking approximation	APP
➢ Introducing an explanation	XPL
➢ Transition to reported speech	TRS
➢ Floor-yielding	FYD
Interpersonal Level:	
➢ Securing comprehension	SCH
➢ Referring to shared knowledge	RSK

Table 4.5 Distribution of discourse marker function of *you know* by L1 Chinese speakers in CAWSE and native speakers in MICASE

Function Category		CAWSE		MICASE	
		N	%	N	%
Textual	INT	6	8.96%	53	8.92%
	LCS	28	41.79%	162	27.27%
	FSR	5	7.46%	94	15.82%
	APP	1	1.49%	33	5.56%
	XPL	20	29.85%	42	7.07%
	TRS	-	-	10	1.68%
	FYD	-	-	22	3.70%
	Total%		**89.55%**		**70.03%**
Interpersonal	SCH	-	-	37	6.23%
	RSK	7	10.45%	141	23.74%
	Total%		**10.45%**		**29.97%**
TOTAL		67		594	

Figure 4.1.1 Frequency of *you know* in textual functions (per 10000 words) by L1 Chinese speakers in CAWSE and native speakers in MICASE
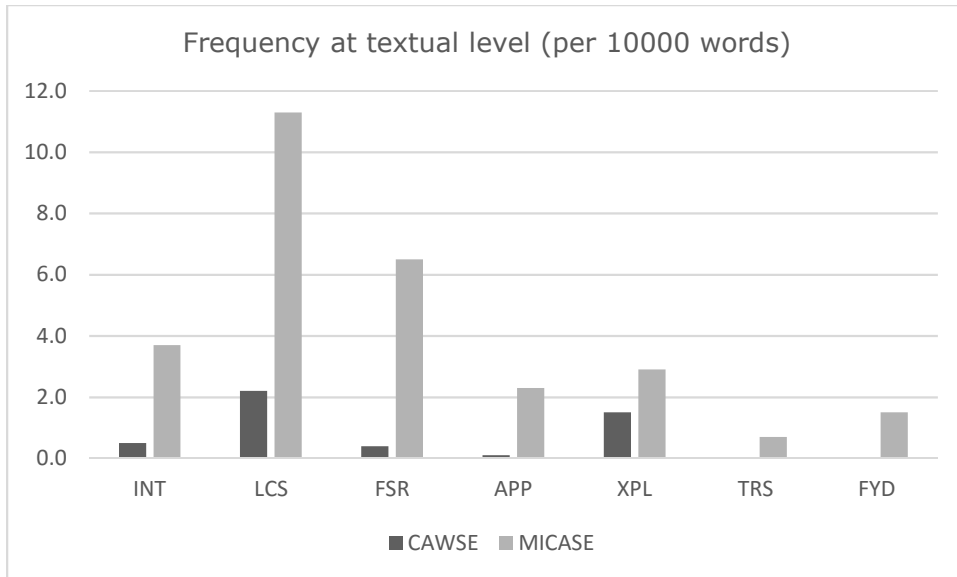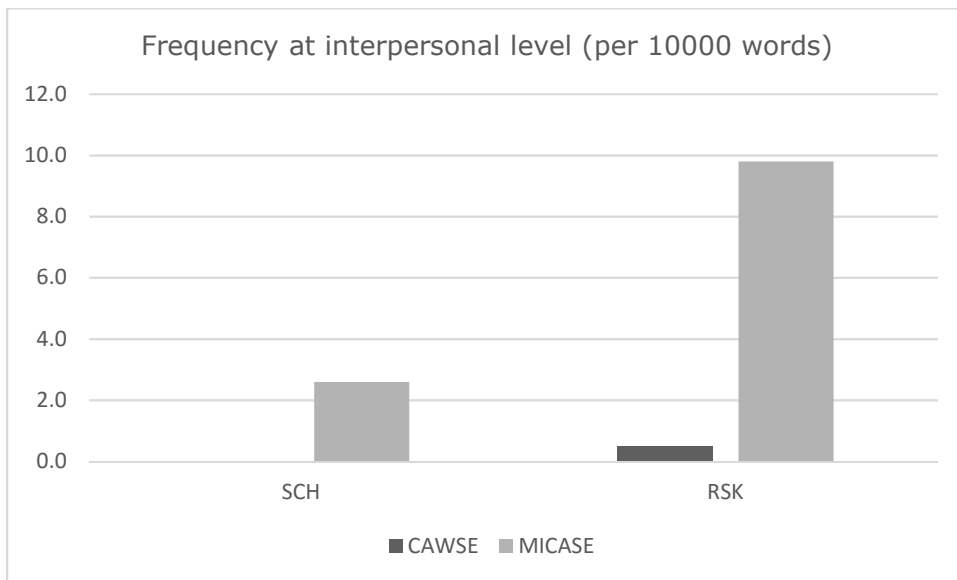

Figure 4.1.2 Frequency of *you know* in interpersonal functions (per 10000 words) by L1 Chinese speakers in CAWSE and native speakers in MICASE

In Figure 4.2, most types of discourse marker function occur in the initial and medial positions. Except for the function of "initiating a new topic" which occurs only in the initial position and "floor-yielding" in the final position, other seven categories are used in both initial and medial positions.

Figure 4.2 Distribution of discourse function *you know* in three positions by L1 Chinese speakers in CAWSE and native speakers in MICASE
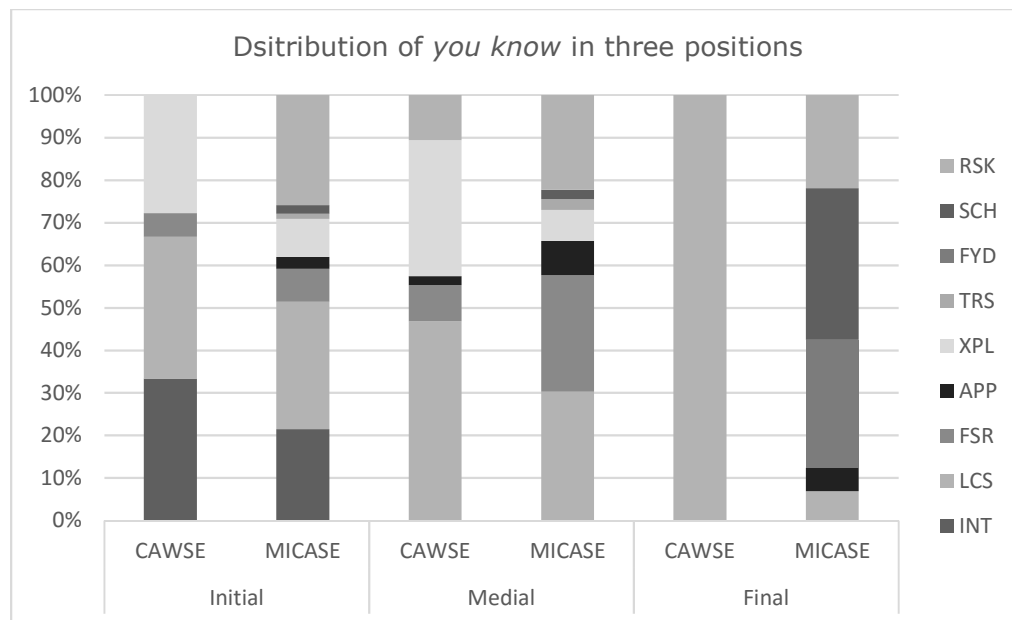
4.3 I think

4.3.1 Previous studies of *I think*

I think in previous studies have been focused on two perspectives, i.e. semantic meaning and functions. Aijmer (1997) defines the prototypical meaning of the verb "think" with a polysemic structure consists of "belief, opinion and intention" (p. 12). The epistemic meanings of *I think* are considered as a grammaticalization by selecting the first person subject *I* and collocating with *think* to derive subjective meanings from its referential meaning (Baumgarten & House, 2010, p. 1189). Therefore, the clause introduced by I think is called epistemic clause (Scheibman, 2001), expressing the speaker's stances, including belief, opinion and subjective evaluation (Aijmer, 1998). The meanings are distinguished by the categorization of functions of I think in spoken discourse by Aijmer (1997) as "tentative and deliberative functions". The tentative function of *I think* reflects the speaker's uncertainty in expressing an idea, in which *I think* usually occurs in utterance-medial or final position; while *I think* functions deliberatively when stating a certain opinion (Aijmer, 1997; Baumgarten & House, 2010; Liu, 2013). However, Liu (2013) compares the uses of *I think* by native speakers and L1 Chinese speakers of English and finds that only Chinese speakers use *I think* to communicate

42

deliberative meaning in regardless of the position, while the native speakers only use it in utterance-initial position.

Similar to *you know*, the meaning and function of *I think* should be discussed by combining the linguistic features and context (Baumgarten & House, 2010). Baumgarten and House (2010) compare the different formal structures of I think used by native and non-native speaker, including the categories of simple clause construction, complement clause construction, and verbal routine. In their study, non-native speakers with L1 in Indonesian, German, Chinese, Korean, French, Nepali and Gujrati tend to utter more *I think* complementary clause with "that-omission" structure, which signals the lack of "structural variability of the collocation" (p. 1190). In terms of contexts, they (2010) find that learners of English tend to employ *I think* in more types of context, which is considered as overuse by second language learners (see also Liu, 2013).

4.3.2 Formal features of *I think*

4.3.2.1 *I think* and positions

Position is an important feature for defining the functions of *I think* (Aijmer, 1997). As pointed out by Aijmer (1997), for example, the initial *I think* with a falling tone signals deliberative function, while the utterance-final *I think* usually shows less uncertainty. *I think* inserted in medial position, on the other hand, is a consequence of the speaker's planning. Most of the instances in this study occurs in the initial position (see Table 4.6) which seems close to the findings in previous studies (Aijmer, 1997, p. 23).

Table 4.6 Distribution of *I think* in three positions by L1 Chinese speakers in CAWSE and native speakers in MICASE

Utterance Position	CAWSE		MICASE	
	N	%	N	%
Initial	985	85.80%	219	72.76%
Medial	137	11.93%	68	22.59%
Final	26	2.26%	14	4.65%
TOTAL	1148	100%	301	100%

Note: Utterance-initial position includes I think that preceded by filled pauses (*ah, oh, er, erm*, etc), markers (*because, but, so*).

4.3.2.1 *I think* and pauses

The statistics of pauses appear before or after (or both) *I think* shows that speakers in CAWSE corpus tend to use more filled pauses than native speakers in MICASE corpus, implying that Chinese learners may be less confident in expressing their opinions. Both filled pauses and unfilled pauses are considered as the boundaries of tone unit, which may also indicate the textual function of content search.

Table 4.7 *I think* with pauses before or after in CAWSE and MICASE

Types of Pause	CAWSE		MICASE	
	N	%	N	%
Unfilled Pauses	64	5.57%	96	31.89%
Filled Pauses	421	36.67%	17	5.65%
No pause	663	57.75%	188	62.46%
TOTAL	1148	100%	301	100%

4.3.3 Non-DM function of *I think*

According to Brinton's (1996) list and my revised list of the features for defining a DM (see Table 2.1 and Table 2.2 in Section 2), *I think* as DM function should be syntactically independent. Therefore, the structure of *I think* connecting prepositions such as *about, on, of* as preposition phrases of *think* are considered as syntactically dependent. In addition, a commonly used positive response to a question, *I think so*, is also regarded as non-DM use. In Aijmer (1997), she considers *I think* following a complementary clause introduced by *that* as a signal of deliberative function, which violates the syntactic feature of a DM and should also be excluded. Consequently, four instances of *I think* are excluded as non-DM function in my samples: *I think* occurs in "*I think so*", "*I think about*", "*I think of*" and "*I think that*", as is suggested by Liu (2013, p. 159).

> Example (12a)
> t: will lead to humans' working for fewer hours or less time and enjoy more (1.1)
> free time
> s014: oh
> t: or leisure time
> s014: er: **I think so**

Example (12b)

s022: er what your require is why what is <ut>x</ut> to the alternative medicines I er first word **I think about** this er <ut>x</ut> er the <ut>x</ut> is not placebo

(C022it010)

Example (12c)

S3: well, water, water and beer is pretty heavy. like if **i think of** carrying a five gallon bucket of beer, i would think that's gonna be heavier than a five gallon bucket of yeast.

(MSGRit155)

Example (12d)

S1: well it's uh i'm also just, **i think that** i i made a horrible decision

(MOFCit199)

From the two corpora in this study, only one instance of *I think about* in CAWSE is function (12b) and two instances of *I think of* in MICASE are non-DM function. Both speakers in the two corpora use I think so as a positive response to a question and use *I think* to lead a complementary clause introduced by *that* when expressing opinions. It is worth mention that *that* following *I think* is not always a complementizer but also a pronoun as subject, which occurs in both corpora.

4.3.4 DM function of *I think* at textual level

4.3.4.1 Marking content search

I think functions at the textual level as an on-line planning device when the speaker is preparing for what to say in stating an opinion (Baumgarten & House, 2010; Liu, 2013). It is usually occurs "before or after tone unit boundaries" marked by filled pauses and unfilled pauses, indicating the speaker's planning (Aijmer, 1997, p. 24). According to the basic meaning of *I think* (see Section 4.3.1), it underlines an idea that the speaker is "*thinking*" (Aijmer, 1997; Baumgarten & House, 2010) while uttering *I think*, although *I think* means more than that. When the speaker utters I think preceded or followed filled pauses or unfilled pauses, for example, in (13a) the speaker is trying to explain the change of water level by gesticulating or pointing (inferred from the context) the estimated depth, i.e. a couple of feet. In this case, *uh* signals the speaker is out of word or content and *I think* indicates that he/she is going through the conceptual work to find a way to make an estimation for that value.

Example (13a)

S1: it is different it's shallower the water level has dropped <u>uh</u>, **i think** as much as a couple of feet.

<div align="right">(MLABit015)</div>

The second type in this functional category is in Example (13b), in which the speaker utters more than two (including two) discourse markers one after another. The co-occurrence of three markers, *you know*, *I mean*, and *I think*, indicates that the speaker is trying to hold the floor while struggling at thinking of or organizing the ideas. Such instance can be found in the samples of both corpora, for instance in Example (13c) *well*, *I think* and *maybe* together function as an on-line planning device.

Example (13b)

S3: the first one with the body, thing <u>you know, i mean **i think**</u>, all the other people who've looked at it have kind of [S4: yeah] disputed

<div align="right">(MSEMit010)</div>

Example (13c)

t: what skills can meet they promote do you think if you use them

s088: er: well **I think** maybe they can be er more personalized

<div align="right">(C088it701)</div>

This function takes up the most proportion among all the four categories (44% in CAWSE and 42% in MICASE). One of the basic meanings of *I think* as *I am thinking* may account for that. In this case, *I think* literally means "I am thinking what is going to say next".

4.3.4.2 Marking false start and repair

In this study, *I think* signposts the textual function of marking false start and repairing the "aborted utterances" (Baumgarten & House, 2010, p. 1192). Again, similar to the repair function of you know, patterns of repetition of words (usually one to three words) preceding and following *I think* are observed. Example (14a) and (14b) shows how *I think* serves as a textual editing device, by repeating the previous words. The speaker thus is able to reorganize the utterance. Noticeably, the syntactic structure does not always stay the same, which is edited according to the context. For instance, in (14c), the structure changes from *I wanna* (colloquial expression of *I want to*) to *I'm gonna* (*I'm going to*) when the speaker reorganizes the utterance.

Example (14a)

t: what would you like to change about this university

s021: en what would you like to change <u>about this university</u> **I think** <u>about this university</u> maybe mm everybody need to mm think more m- maybe to read more about m- read more books

(C021it009)

Example (14b)

and<u>, i really</u>, **i think** <u>i i really</u> would enjoy that lifestyle

(MADVit052)

Example (14c)

i have lots of other interests like um, that are a little bit more like, paleontology or astronomy or [S1: oh] international religion or uh not religion international relations, so, those things <u>i wanna</u> **i think** <u>i'm gonna</u> concentrate more on, i don't think i wanna do the business.

4.3.5 DM function of *I think* at interpersonal level

4.3.5.1 Tentative function

The tentative function of I think at the interpersonal level is to convey the speaker's uncertainty regarding to the content of speeches (Aijmer, 1997). This function usually takes place in medial or final position (Aijmer, 1997; Liu, 2013). Aijmer (1997, pp. 24-25) names a sub-function for final *I think* as "afterthought" which is to qualify the information uttered previously with a rising tone. This sub-function will be classified in this category, since prosodic feature is not available for comparison in current study and final *I think* is usually used when the speaker is uncertain about what he/she says (Aijmer, 1997).

In Example (15a) the speaker is trying to retell the story of Apple's brand image. It seems the speaker is certain about the image of apple with seven colours, however, he/she adds I think in the end of the utterance to show the uncertainty of the expression or lexical choice. In this case, it performs a tentative function in spite of the fact that the speaker has enough information of the story.

Example (15a)

s065: it starts from a complete image of New- of Newton er seen the f- the falling apple and it come and suddenly be idea of gravity <ut>x</ut> and then the brand image gradually <dvl>involves{evolves}</dvl> to an an apple with seven colors **I think**

(C065it832)

Except for the final position, other patterns have been discovered for this category, including co-occurrence of tentative words (maybe, probably, like, just, etc.) and clauses with subjunctive mood (15c).

Example (15b)

S3: i didn't, i don't know. i haven't, and like, does it really make a difference or, is it just a courtesy thing?

S1: **i think** it's <u>probably just</u> a courtesy thing

(MSGRit128)

Example (15c)

i think <u>if</u> you asked her straight out like where are you? she'd be like, on the boat

(MSEMit085)

Example (15d)

because **i think** it would, it would be misleading

(MMTGit141)

In (15b), for example, S1 is stating a positive answer to S3's question, which seems deliberative, introducing by the initial *I think*. Nonetheless, the tentative words, *probably* and *just,* still implie the speaker's uncertainty. Example (15c) shows *I think* preceding an if-clause, meaning that the idea is made up by the speaker instead of true condition. In (15d), *I think* occurs in medial position which also signals a tentative meaning.

4.3.5.2 Deliberative function

When the speaker is confident about what he/she says, I think can be used to convey deliberative meaning (Liu, 2013). The deliberative function can be distinguished from the utterance content. For example, (16a) the speaker S52 initiates another topic by saying *you know* (see Section 4.2.4.1 for explanation) and states his/her leaving. *I think* here is used in a deliberative, because the speaker is certain about the future action. Although it is said that the final *I think* will weaken the speaker's commitment (Aijmer, 1997), the Example (16b) finds a final *I think* which serves deliberative function. The speaker s018 is sharing information about his/her daily life with the teacher. The first *I think* in the initial position is deliberative,

because it is the speaker's subjective evaluation of an object. The final *I think* is also a deliberative function, for the fact that the price the second canteen is lower. In this case, *I think* may serve as a confirmation of the preceding content. Moreover, I think can also function as "a post-nuclear tail pronounced on a low pitch" (Aijmer, 1997, p. 25). Since the lack of prosodic evidence, it remains in this category. Though only three instances have been discovered, native speakers also use medial *I think* to express deliberative meaning, in which it tends to express reassurance (Holmes, 1990; Cf. Aijmer, 1997, p. 22). In (16c) the speaker inserts *I think* before the opinion, but also uses "strictly" to confirm his/her idea.

Example (16a)

S48: i just do it when i'm at work.

S52: you know. cuz **i think** i'm leaving here i think i'll go grab my car and uh, go to this review session.

(MSVCit064)

Example (16b)

t: uhu (1.0) where you usually have you lunch

s018: er: I usually go to the er the second canteen

t: erm uhu

s018: **I think** it's the the the food in the canteen is more delicious

t: uhu

s018: and the and the er (2.8) the price is lower than the third canteen and the three canteen **I think**

(C018it843)

Example (16c)

S1: a- and that f- the problem with the folk chapter **i think** is <u>strictly</u> one of nomenclature

(MOFCit094)

In this category, only one instance of *I think* in utterance-final position from CAWSE is found to serve deliberative function. Nearly one third of initial and part of (less than 10%) medial instances of *I think* conform to this category (see Table 4.6), which seems to be the main function of utterance-initial *I think*.

Table 4.6 Distribution of *I think* in deliberative function by L1 Chinese speakers in CAWSE and native speakers in MICASE

Position	CAWSE			MICASE		
	N	Total	%	N	Total	%
Initial	329	985	33.40%	74	219	33.79%
Medial	12	137	8.76%	3	68	4.41%
Final	1	26	3.85%	0	14	0.00%
TOTAL	342	1148	29.79%	77	301	25.58%

4.3.6 Summary

In the general analyses of *I think*, four categories are distinguished (see Table 4.7), namely marking content search, marking false start and repair at textual level, deliberative and tentative function at interpersonal level, by discussing the relationship with utterance positions and pauses. On average (see Figure 4.3), the frequencies of all the four functions are significantly higher by L1 Chinese speaker than by native speakers ($p < 0.0001$, only $p < 0.01$ in FSR). It presents a statistical evidence for the overuse of *I think* by second language or English as foreign language learners (Liu, 2013). In terms of L1 Chinese speakers of English, Liu (2013) states that the overuse of I think is possibly due to the Chinese equivalence "wo juede" which is supposed to be a common discourse marker use in Chinese opinionative discourse. As is presented in Table 4.8, interestingly, each function shares a similar proportion between two corpora, except for the function of false start and repair in MICASE which is double as much as in CAWSE. It is inferred that both native speaker and non-native speaker in this study have certain common in the discourse marker use of *I think*, i.e. their knowledge, understanding or preference of the functions. For instance, they both use tentative function in final position and only use deliberative in initial and medial position (except for only one instance in CAWSE which is likely a "post-nuclear tail"). In addition, the contexts of the conversation may "encourage" the speakers to use more *I think*. In this study, the teachers in the interviews from CAWSE corpus tend ask students questions about their opinions, ideas, or feelings, with the structure of *do you think*.

Table 4.7 List of discourse marker function of *I think* with abbreviated label

Full Label	Abbreviated Label
Textual Level:	
➢ Marking content search	LCS
➢ Marking false start and repair	FSR
Interpersonal Level:	
➢ Deliberative function	DLB
➢ Tentative function	TTT

Table 4.8 Distribution of discourse marker function of *I think* by L1 Chinese speakers in CAWSE and native speakers in MICASE

Function Category		CAWSE		MICASE	
		N	%	N	%
Textual	CTS	508	44.25%	127	42.19%
	FSR	49	4.27%	27	8.97%
Interpersonal	DLB	341	29.70%	77	25.58%
	TTT	250	21.78%	70	23.26%
TOTAL		1148	100%	301	100%

Figure 4.3 Frequency of discourse marker function *I think* (per 10000 words) by L1 Chinese speakers in CAWSE and native speakers in MICASE
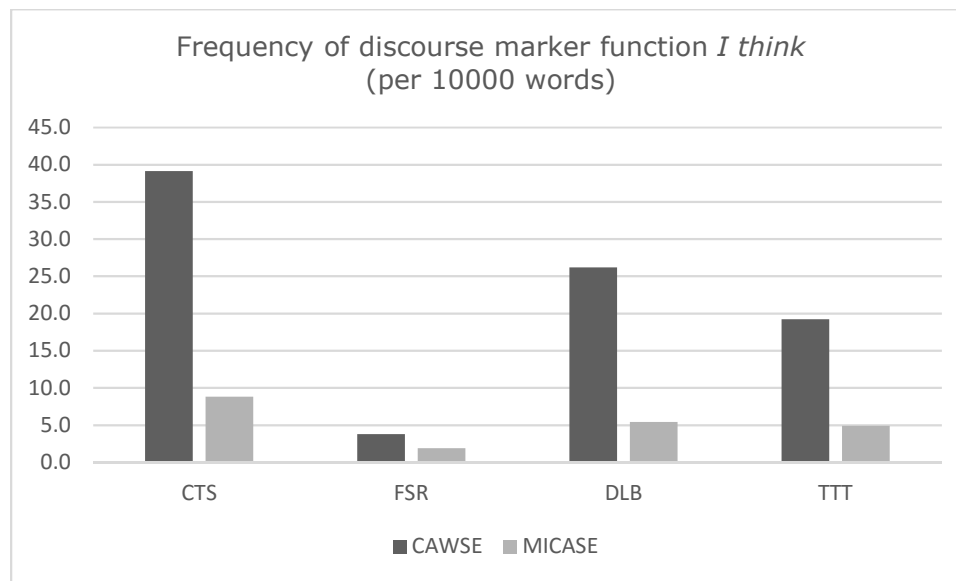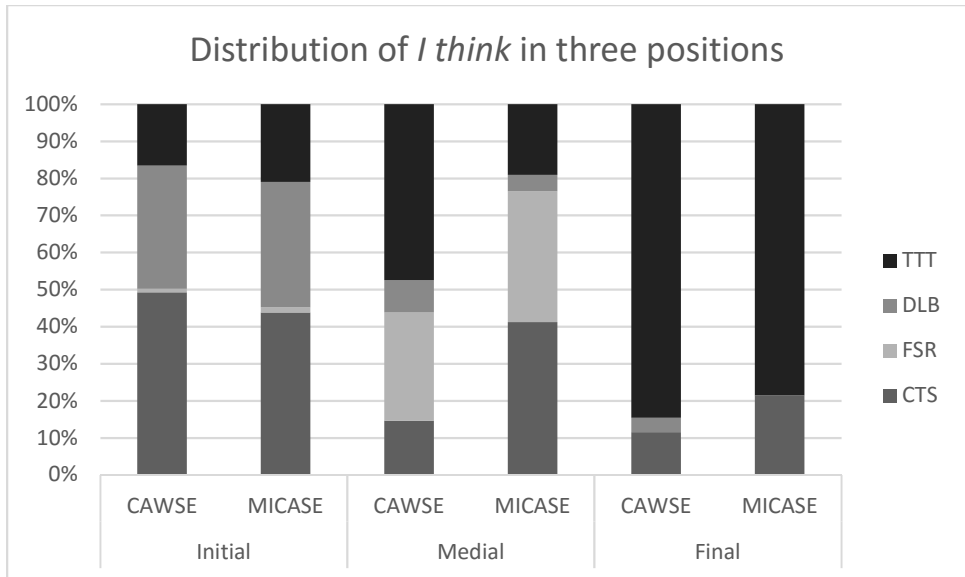
Figure 4.4 Distribution of discourse function *I think* in three positions by L1 Chinese speakers in CAWSE and native speakers in MICASE

Distribution of *I think* in three positions

4.4 Conclusion

In this chapter, two discourse markers, *you know* and *I think* are analysed from their formal features such as utterance position and co-occurrence with pauses to the DM and non-DM functions. In the analyses of *you know*, the DM functions used by native speakers in MICASE corpus are more versatile than by L1 Chinese speakers in CAWSE corpus. The results show that native speakers tend to use more *you know* than non-native speakers in order to draw attention and invite inferences from addressees while expressing ideas, making arguments, stating facts, etc. On the contrary, non-native speakers in this study have not used *you know* in functions such as floor-yielding, marking transition to reported speeches and securing comprehension. In the analyses of *I think*, the results also show significant difference between two groups, implying the overuse of this marker by non-native speakers in this study.

Chapter 5 Conclusion

The functions of discourse markers vary and are influenced by a great many linguistic and sociolinguistic factors, including immersion in target language, stylistic differences, oral accuracy and proficiency (Liu, 2013), as well as social class, ethnicity, gender, age (Aijmer, 2013, p. 148) and individual tendency (Ö¨stman, 1981, p. 31). In the current study, however, the data is not sufficient to examine all the possible factors. It only fulfils the two proposed objectives that are to compare the differences in discourse marker use by native and non-native speakers looking at *you know* and *I think*, and to explore possible routes for setting a starting point as in pedagogical implication of DM teaching in EFL classrooms.

5.1 Suggestions

In order to shed light on the second objective, some notions are provided to EFL teachers and advanced learners. Discourse markers should be considered as an important feature in teaching spoken English, which are "useful contextual coordinates for […] speakers […] to structure and organize speech" at textual and interpersonal level. EFL teachers can be motivated when improving their own speaking skills to a more accurate and proficient level by mastering the function of DMs. Moreover, EFL teachers should be equipped with certain knowledge, including features and functions of DMs so that the students' performance in oral English can be evaluated with linguistic evidence rather than by subjective judgment. At last, teachers should encourage students to practicing different kinds of DMs in communication.

5.2 Limitations

In regard to the limitation of this study, firstly, the main part of this study is qualitative description based on quantitative analyses, whose results may be influenced by many factors as mentioned in this chapter. Secondly, the methodology consumes a great number of time and efforts, as well as demands high proficiency from teachers which may not possible for directly implicating as suggested in Section 5.1. The corpus-driven approach is convincing but difficult to be put into practice, which needs great efforts in observing a large number of concordance lines before identifying certain patterns and also requires a clear understanding of the theories, including definition, framework, etc. Finally, the CAWSE corpus is an on-going project and provides with limited sample data. The results of the current study may not be able to represent the patterns of discourse marker uses by the Chinese students in The University of Nottingham Ningbo China. Consequently, further study in this area can be conducted with the development of the CAWSE project.

References

Aijmer, K. (1997). I think — an English modal particle. In *Modality in Germanic LanguagesHistorical and Comparative Perspectives*.

Aijmer, K. (1998). Epistemic predicates in contrast. *Language and Computers, 24*, 277-296.

Aijmer, K. (2002). *English discourse particles [electronic resource] : evidence from a corpus / Karin Aijmer*. Philadelphia: Philadelphia : John Benjamins Pub. Co.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*(1), 173-190.

Aijmer, K. (2013). *Understanding pragmatic markers : a variational pragmatic approach / Karin Aijmer*. Edinburgh: Edinburgh : Edinburgh University Press.

Aijmer, K., Foolen, A., & Simon-Vandenbergen, A.-M. (2006). Discourse particles in the perspective of heteroglossia. *Approaches to Discourse Particles*, 101-114.

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Retrieved from Available from http://www.laurenceanthony.net/software

Baumgarten, N., & House, J. (2010). I think and I don't know in English as lingua franca and native English discourse. *Journal of Pragmatics, 42*(5), 1184-1200.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99): Cambridge University Press.

Brinton, L. J. (1996). *Pragmatic markers in English [electronic resource] : grammaticalization and discourse functions / Laurel J. Brinton*. Berlin; New York: Mouton de Gruyter.

Brinton, L. J. (2008). *The comment clause in English: Syntactic origins and pragmatic development*: Cambridge University Press Cambridge.

Brinton, L. J. (2010). 10. Discourse Markers. *Historical pragmatics, 8*, 285.

Brown, H. D. (2007). *Teaching by principles : an interactive approach to language pedagogy / H. Douglas Brown* (3rd ed. ed.). White Plains, NY: White Plains, NY : Pearson Education.

Buysse, L. (2017). The pragmatic marker you know in learner Englishes. *Journal of Pragmatics, 121*, 40-57. doi:10.1016/j.pragma.2017.09.010

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge ; New York: Cambridge ; New York : Cambridge University Press.

Chen, Y. H., Harrison, H., Oakey, D., Stevens, M. P., Yang, S., Ioratim-Uba, G., . . . Bruncak, R. (2018). *UNNC Corpus of Academic Written and Spoken Corpus (UNNC CAWSE) Version 1.0.*

Dér, C. (2010). On the status of discourse markers. *Acta Linguistica Hungarica, 57*(1), 3-28.

Erman, B. (1987). *Pragmatic expressions in English: A study of you know, you see and I mean in face-to-face conversation.* Almqvist & Wiksell International,

Erman, B. (2001). Pragmatic Markers Revisited with a Focus on You Know in Adult and Adolescent Talk. *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies, 33*(9), 1337-1359. doi:10.1016/S0378-2166(00)00066-7

Fox, B. A., & Jasperson, R. (1995). A syntactic exploration of repair in English conversation. *Alternative linguistics: Descriptive and theoretical modes, 102*, 77-134.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics, 14*(3), 383-398.

Fraser, B. (1996). Pragmatic markers. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA), 6*(2), 167-190.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*(7), 931-952.

Fraser, B. (2009). An account of discourse markers. *International review of pragmatics, 1*(2), 293-320.

Fung, L. (2003). *The use and teaching of discourse markers in Hong Kong: students' production and teachers' perspectives.* University of Nottingham,

Fung, L., & Carter, R. (2007). Discourse Markers and Spoken English: Native and Learner Use in Pedagogic Settings. *Applied Linguistics, 28*(3), 410-439. doi:10.1093/applin/amm030

Götz, S. (2013). *Fluency in Native And Nonnative English Speech / Sandra Götz, Justus Liebig University, Giessen*. Amsterdam ; Philadelphia: Amsterdam ; Philadelphia : John Benjamins Publishing Company.

Halliday, M. A. (1979). Modes of meaning and modes of expression: types of grammatical structure, and their determination by different semantic functions. *Function and context in linguistic analysis, 1*, 57-79.

Halliday, M. A. K. (1973). Explorations in the Functions of Language.

Heine, B. (2013). On discourse markers: Grammaticalization, pragmaticalization, or something else? In: De Gruyter Mouton.

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication, 10*(3), 185-205.

Hunston, S. (2002). *Corpora in applied linguistics / Susan Hunston*. Cambridge: Cambridge : Cambridge University Press.

Jucker, A. H. (2002). 11 Discourse markers in Early Modern English.

Jucker, A. H., & Ziv, Y. (1998). *Discourse markers [electronic resource] : descriptions and theory / edited by Andreas H. Jucker, Yael Ziv*. Amsterdam, Philadelphia [Pa.]: J. Benjamins.

Lakoff, G. (1975). Hedges: A study in meaning criteria and the logic of fuzzy concepts. In *contemporary Research in Philosophical Logic and Linguistic semantics* (pp. 221-271): Springer.

Lakoff, R. T. (1972). *The pragmatics of modality*: Ann Arbor, Mich.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach*. *Language Learning, 40*(3), 387-417. doi:doi:10.1111/j.1467-1770.1990.tb00669.x

Liu, B. (2013). Effect of First Language on the Use of English Discourse Markers by L1 Chinese Speakers of English. *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies, 45*(1), 149-172. doi:10.1016/j.pragma.2012.11.002

Lutzky, U. (2012). *Discourse markers in Early Modern English [electronic resource] / Ursula Lutzky*. Amsterdam; Philadelphia: John Benjamins Pub. Co.

Maschler, Y., & Schiffrin, D. (2015). Discourse Markers: Language, Meaning, and Context. In D. e. Tannen, H. E. e. Hamilton, D. e. Schiffrin, & C. T. c. Adger (Eds.), *The handbook of discourse analysis / edited by Deborah Tannen, Heidi E. Hamilton & Deborah Schiffrin, contributors Carolyn Temple Adger [and fifty nine others]* (Second edition. ed.).

McEnery, T., & Hardie, A. (2012). *Corpus linguistics : method, theory and practice*. Cambridge: Cambridge : Cambridge University Press.

Müller, S. (2005). *Discourse markers in native and non-native English discourse [electronic resource] / Simone Müller*. Philadelphia: Philadelphia : John Benjamins Pub.

Ostman, J.-O. (1982). The Symbiotic Relationship between Pragmatic Particles and Impromptu Speech.

Östman, J.-O. (1995). Pragmatic particles twenty years after. *Organization in discourse, 14*, 95-

108.

Östman, J.-O. (1981). *You know [electronic resource] : a discourse functional approach / Jan-Ola Östman*. Amsterdam: Amsterdam : Benjamins.

Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics, 29*(6), 1139-1172.

Redeker, G. (1993). *Discourse markers as attentional cues to discourse structure.* Paper presented at the Poster presented at NATO Advanced Research Workshop "Burning issues in Discourse", Maratea, Italy.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes, 14*(4), 423-441.

Scheibman, J. (2001). Local patterns of subjectivity in person and verb type in American English conversation. *Typological Studies in Language, 45*, 61-90.

Schiffrin, D. (1987). *Discourse markers / Deborah Schiffrin*. Cambridge: Cambridge : Cambridge University Press.

Schiffrin, D. (2006). Discourse marker research and theory: revisiting and. In K. Fischer (Ed.), *Approaches to discourse particles [electronic resource] / edited by Kerstin Fischer*. Amsterdam, Boston, London: Elsevier.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales., J. M. (1999). The Michigan Corpus of Academic Spoken English. from The Regents of the University of Michigan https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase

Smith, S. W., & Jucker, A. H. (1998). Interactive aspects of reference assignment in conversations. *Pragmatics & Cognition, 6*(1), 153-187.

Stubbe, M., & Holmes, J. (1995). You know, eh and other" exasperating expressions": An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication*.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6): John Benjamins Publishing.

Tree, J. E. F., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics, 34*(6), 727-747.

Trillo, J. R. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*(6), 769-784. doi:10.1016/S0378-2166(02)00022-X

Zwicky, A. M. (1985). Clitics and particles. *Language*, 283-305.

Appendix

Convention in current study

	Convention for CAWSE	**Convention for MICASE**
Pauses	(1.0)	, . (1-2s) ... (2-3s.) timed (4+s.)
Hesitation markers	ah, er, erm, en,	um, uh
Backchannels	mm, mhm, uhu	hm, hm', huh, mm, mhm, uh, mkay
Reduced because	cos	cuz
personal pronoun I	I	i
Overlap	S: simultaneous utterance T: simultaneous utterance	S: [simultaneous utterance] T: [simultaneous utterance] the utterance continues
Laughter	<laughing>	<LAUGH>, <S LAUGH> <ST LAUGH>, etc.
Unintelligible speech	<ut>x</ut>	(xx)
Deviation	*it said that our* <dvp >*hurt* {*heart*}</dvp> *er have four* *chambers*	-
Serial number for **extractions**	e.g. (C097yk001) C = CAWSE 097 = file name yk = you know 001 = serial number in concordance lines of *you know*	e.g. (MSGRit155) M = MICASE SGR = study group it = I think 155 = serial number in concordance lines of I think