

**DISCOURSE MARKERS IN SPOKEN ENGLISH:
A CORPUS STUDY OF NATIVE SPEAKERS AND
CHINESE NON-NATIVE SPEAKERS**

by

LAN-FEN HUANG

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of English
School of English, Drama and American & Canadian Studies
College of Arts and Law
The University of Birmingham
July 2011

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis explores the use of discourse markers (DMs) in the speech of Chinese non-native speakers (NNSs) of English and native speakers (NSs), using corpus methodologies, the *Linear Unit Grammar* analysis (Sinclair and Mauranen 2006) and text-based analyses. It reports that the DMs for analysis, *like, oh, well, you know, I mean, you see, I think* and *now*, occur more frequently in the dialogic genres than in the monologic genres extracted from the three corpora, SECCL, MICASE and ICE-GB. The co-occurrence of DMs is taken as evidence to determine the categories for discussion with the suggested functions being secondary interpretations. Surprisingly, there are similarities in the use of DMs between Chinese NNSs and NSs. For the differences, some require NSs to become more tolerant and inclusive of the versions of English and some require pedagogical interventions for the Chinese NNSs. This thesis demonstrates that the use of DMs correlates with the genre, context, type of activity and identity of the speaker. All such factors affect the speakers' choice of a DM to use when giving priority to discourse organisation, fluency, the engagement of the listeners, the construction of the speaker's persona and the creation of solidarity.

Dedication

to my dad and mum,
who have been there for me all the time

Acknowledgements

I gratefully acknowledge the support for the first two years of my research from the Ministry of Education, Taiwan, the Republic of China. Without its award of the 2007 Study Abroad Scholarship, I would not have been able to start doctoral research on my own.

I would like to express my gratitude to my family in Taiwan, my father, Min-chiu Huang, my mother, Hsiu-duan Chen, my brother, Yi-ching Huang and my sister, Hsiao-chi Huang for their unfailing support, which has been important in helping me to undertake this research. All my friends in Taiwan are also greatly appreciated for their constant encouragement. They are always prepared to listen to me ramble on about the drudgery of everyday life in Birmingham. Particular thanks are due to Dr Joweet Sun and Prof Chi-yen Shen for their continued encouragement and support in my career, ever since I first met them in 2001.

My thanks also go to my friends in Britain, in particular Mr John Ruck and Mrs Anne Ruck. They had been praying for me. They kindly let out their lovely house to me while they were doing missionary work in Indonesia. I had a very enjoyable stay in Selly Oak. My sincere thanks also go to my fellow research students in the Department of English, who have actively participated in postgraduate seminars and supported me in many different ways. The support of the faculty members in the Centre for English Language Studies and the Department of English is also greatly appreciated. Their instruction in my MA programme in 2004-2005 gave me a solid grounding in research methods and contexts that prepared me for embarking on a doctoral thesis in 2008. I particularly thank Dr Almut Koester for her feedback as my examiner in the upgrading examination from the first-year to second-year PhD programme, Dr Pernilla Danielsson and Dr Oliver Mason for their technical support on corpus methodologies, Dr Allan White for his advice on statistical measurements and Dr Eve Richards for her professional proofreading service.

I would like to thank my external examiner, Dr Michaela Mahlberg (The University of Nottingham) and internal examiner, Dr Jeannette Littlemore for examining this thesis and giving very constructive comment, which have certainly helped to improve it.

Finally, I would like to extend my sincere thanks to my supervisor, Professor Susan Hunston, for her supervision, guidance and valuable criticism from the start of my work. I would not have made steady progress without her feedback. In the course of this research, the most enjoyable times were the supervision meetings with her. I particularly enjoyed the time

when I could raise questions, face difficulties and come up with solutions for my thesis with her. I am always inspired by what she says as well as our conversations, and very much impressed by her responsiveness and resourcefulness. Whatever errors and mistakes there are in this thesis, they assuredly result from my failure to benefit from Susan's critical and constructive comments and wise counsel.

Table of contents

CHAPTER 1: INTRODUCTION	1
1.1 The present study.....	1
1.1.1 Rationale behind the thesis and its general aim.....	1
1.1.2 Hypotheses and research questions.....	3
1.1.3 Significance and contribution of the research	4
1.1.4 Outline of the thesis	4
1.2 The grammar of spoken English.....	5
1.3 Discourse markers in spoken English	7
1.4 The use of native speaker and learner/non-native speaker corpora.....	9
CHAPTER 2: LITERATURE SURVEY	12
2.1 Corpora as a revolution in language study	12
2.1.1 Corpora in use.....	12
2.1.2 Information derived from corpora	14
2.2 Two trends in research on learner corpora	16
2.2.1 Comparative study for English language teaching purposes.....	17
2.2.2 Diachronic study and error analysis for second language acquisition purposes	20
2.3 Towards the characterisation of discourse markers.....	22
2.3.1 Terminology.....	22
2.3.2 Characteristics and definitions of discourse markers.....	23
2.3.3 Discourse markers in this study.....	29
2.4 Previous studies of discourse markers	29
2.5 <i>Linear Unit Grammar</i> : an approach to the description of spoken English..	32
2.6 Chapter summary	36
CHAPTER 3: METHODOLOGY OF THE STUDY	38
3.1 Corpora under investigation	38
3.2 Selection of discourse markers for analysis	41
3.3 Quantitative corpus analysis.....	42
3.3.1 Frequency information	42
3.3.2 Distinguishing between non-discourse use (Type A) and discourse use (Type B) of the words and phrases for analysis	43
3.3.3 Frequency comparisons between the monologic and dialogic genres and between the speech of the Chinese non-native speakers and native speakers	47
3.3.4 Collocates of the words and phrases	50
3.3.5 Identifying the positions of Type B words and phrases	50

3.3.6	Grammatical and discorsal aspects of Type A and Type B words and phrases.....	51
3.3.7	Random sampling procedure.....	51
3.4	Qualitative text-based analysis.....	57
3.5	Key terms used in the study.....	57
3.6	Codes and conventions.....	58
CHAPTER 4: ANALYSIS OF LIKE		60
4.1	Introduction.....	60
4.2	Previous studies of <i>like</i>	61
4.2.1	Word classes of <i>like</i>	61
4.2.2	Syntactical aspect of <i>like</i>	61
4.2.3	<i>Linear Unit Grammar</i> analysis of <i>like</i>	63
4.2.4	Previous studies of <i>like</i> as a discourse marker.....	64
4.3	Frequency information in the speech of the non-native speakers and native speakers.....	65
4.3.1	Overall frequency of <i>like</i>	65
4.3.2	Collocates of <i>like</i>	68
4.4	Discourse aspects of Type B <i>like</i>	73
4.4.1	Positions in an utterance/turn.....	73
4.4.1.1	<i>Like</i> in extra-clausal position.....	74
4.4.1.2	<i>Like</i> in intra-clausal position.....	75
4.4.2	Contexts where Type B <i>like</i> tends to occur.....	76
4.4.2.1	<i>Like</i> co-occurring with hesitation markers, pauses and restarts.....	76
4.4.2.2	<i>Like</i> co-occurring with numerical expressions and indications of location.....	77
4.4.2.3	<i>Like</i> co-occurring with reported speech.....	79
4.4.2.4	<i>Like</i> co-occurring with expressions of uncertainty.....	79
4.4.2.5	<i>Like</i> co-occurring with expressions of certainty and key points.....	80
4.4.2.6	<i>Like</i> co-occurring with exemplifications.....	81
4.4.2.7	<i>Like</i> co-occurring with explanations.....	82
4.4.2.8	Unclassified instances of <i>like</i>	83
4.4.2.9	Summary of the contexts where Type B <i>like</i> tends to occur.....	84
4.5	Further investigation.....	89
4.5.1	Variations in the use of <i>like</i> across texts and among speakers: Revisiting the sub-corpus of the American NSs' highly interactive discourse mode in MICASE.....	89
4.5.2	Analysis of <i>just like</i>	92
4.6	Chapter summary and conclusions.....	93
CHAPTER 5: ANALYSIS OF OH		96
5.1	Introduction.....	96
5.2	Previous studies of <i>oh</i>	97
5.2.1	Word class of <i>oh</i>	97

5.2.2	Syntactical structure of <i>oh</i>	97
5.2.3	The use of <i>oh</i> in spoken English	98
5.2.4	<i>Oh</i> as a marker of change-of-state	99
5.2.5	<i>Oh</i> for marking reported speech	101
5.3	Frequency information in the speech of the non-native speakers and native speakers.....	102
5.3.1	Overall frequency of <i>oh</i>	102
5.3.2	Collocates of <i>oh</i>	104
5.4	Discourse aspects of <i>oh</i>	108
5.4.1	Positions in an utterance/turn	109
5.4.1.1	<i>Oh</i> in extra-clausal position.....	110
5.4.1.2	<i>Oh</i> in intra-clausal position	111
5.4.2	Contexts where <i>oh</i> tends to occur	112
5.4.2.1	<i>Oh</i> co-occurring with reported speech	112
5.4.2.2	<i>Oh</i> co-occurring with hesitation markers, pauses and repetitive words	113
5.4.2.3	<i>Oh</i> co-occurring with repairs and rephrasing	114
5.4.2.4	<i>Oh</i> occurring at the opening and changing of a topic	115
5.4.2.5	<i>Oh</i> as a (preface to a) response to a question and new information.....	116
5.4.2.6	<i>Oh</i> to show emotions.....	118
5.4.2.7	<i>Oh</i> to indicate a cognitive process has been completed	119
5.4.2.8	Problematic and unclassified instances of <i>oh</i>	120
5.4.2.9	Summary of the contexts where <i>oh</i> tends to occur	121
5.5	Further investigation.....	126
5.5.1	The over-representation of <i>oh</i> in extra-clausal turn-initial position in the sub-corpora of the dialogic genres.....	126
5.5.2	Revisiting <i>oh</i> as a (preface to a) response	128
5.6	Chapter summary and conclusions	132
CHAPTER 6: ANALYSIS OF WELL		135
6.1	Introduction.....	135
6.2	Previous studies of <i>well</i>	136
6.2.1	Semantic aspect of <i>well</i>	136
6.2.2	Syntactical aspect of <i>well</i>	137
6.2.3	<i>Linear Unit Grammar</i> analysis of <i>well</i>	137
6.2.4	Previous studies of <i>well</i> as a discourse marker.....	138
6.3	Frequency information in the speech of the non-native speakers and native speakers.....	142
6.3.1	Overall frequency of <i>well</i>	142
6.3.2	Collocates of <i>well</i>	144
6.4	Discourse aspects of <i>well</i>	149
6.4.1	Positions in an utterance/turn	150
6.4.1.1	<i>Well</i> in extra-clausal position.....	151
6.4.1.2	<i>Well</i> in intra-clausal position	152

6.4.2	Contexts where Type B <i>well</i> tends to occur	152
6.4.2.1	<i>Well</i> co-occurring with hesitation markers, pauses, repetitive words and restarts	153
6.4.2.2	<i>Well</i> co-occurring with reported speech	154
6.4.2.3	<i>Well</i> co-occurring with repairs and rephrasing	154
6.4.2.4	<i>Well</i> co-occurring with the opening and closing of a topic and concluding remarks	155
6.4.2.5	<i>Well</i> prefacing questions	156
6.4.2.6	<i>Well</i> co-occurring with transitions and shifts of topic	157
6.4.2.7	<i>Well</i> co-occurring with disagreement and negative evaluation	158
6.4.2.8	<i>Well</i> as a preface to a response	159
6.4.2.9	<i>Well</i> co-occurring with key information	161
6.4.2.10	<i>Well</i> as a continuer	162
6.4.2.11	Unclassified instances of <i>well</i>	163
6.4.2.12	Summary of the contexts where Type B <i>well</i> tends to occur	163
6.5	Chapter summary and conclusions	168
CHAPTER 7: ANALYSES OF YOU KNOW, I MEAN AND YOU SEE		170
7.1	Introduction	170
7.1.1	Hypotheses and research questions	171
7.1.2	Ways of distinguishing <i>you know</i> , <i>I mean</i> and <i>you see</i> between non-discourse use (Type A) and discourse use (Type B)	172
7.2	Previous studies of <i>you know</i> , <i>I mean</i> and <i>you see</i>	173
7.2.1	Grammatical aspect: Syntactical structure	173
7.2.2	<i>Linear Unit Grammar</i> analysis of <i>you know</i> , <i>I mean</i> and <i>you see</i>	175
7.2.3	Previous studies of <i>you know</i> and <i>I mean</i> as discourse markers	176
7.2.4	Previous studies of <i>you see</i> as a discourse marker	180
7.3	Frequency information about <i>you know</i> , <i>I mean</i> and <i>you see</i> in the speech of the non-native speakers and native speakers	181
7.3.1	Overall frequencies of <i>you know</i> and <i>I mean</i>	181
7.3.2	Overall frequency of <i>you see</i>	188
7.3.3	Collocates of <i>you know</i>	189
7.3.4	Collocates of <i>I mean</i>	194
7.3.5	Collocates of <i>you see</i>	197
7.4	Discourse aspects of <i>you know</i> , <i>I mean</i> and <i>you see</i>	200
7.4.1	Positions in an utterance/turn	200
7.4.1.1	<i>You know</i> in extra-clausal position	204
7.4.1.2	<i>I mean</i> and <i>you see</i> in extra-clausal position	205
7.4.1.3	<i>You know</i> in intra-clausal position	206
7.4.1.4	<i>I mean</i> and <i>you see</i> in intra-clausal position	206
7.4.2	Contexts where Type B <i>you know</i> tends to occur	206
7.4.2.1	<i>You know</i> co-occurring with hesitation markers, pauses and restarts	207
7.4.2.2	<i>You know</i> co-occurring with reported speech	207
7.4.2.3	<i>You know</i> co-occurring with repairs	208
7.4.2.4	<i>You know</i> co-occurring with emphatic lexis and key information	208
7.4.2.5	<i>You know</i> co-occurring with exemplifications	210

7.4.2.6	<i>You know</i> co-occurring with clarifications and explanations	210
7.4.2.7	<i>You know</i> co-occurring with contrasting and negative points	211
7.4.2.8	<i>You know</i> occurring at the opening of a topic or narrative	212
7.4.2.9	<i>You know</i> prefacing concluding remarks	212
7.4.2.10	<i>You know</i> co-occurring with shared knowledge presumed by the speaker	213
7.4.2.11	Problematic and unclassified instances of <i>you know</i>	214
7.4.2.12	Summary of the contexts where Type B <i>you know</i> tends to occur	216
7.4.3	Contexts where Type B <i>I mean</i> tends to occur	221
7.4.4	Contexts where Type B <i>you see</i> tends to occur	225
7.5	Chapter summary and conclusions	231
CHAPTER 8: ANALYSIS OF <i>I THINK</i>		232
8.1	Introduction	232
8.1.1	Hypotheses and research questions	232
8.1.2	Ways of distinguishing <i>I think</i> between non-discourse use (Type A) and discourse use (Type B)	232
8.2	<i>I think</i> in the literature	233
8.2.1	Grammatical aspect: Syntactical structure	234
8.2.2	<i>Linear Unit Grammar</i> analysis of <i>I think</i>	235
8.2.3	Previous studies of <i>I think</i>	237
8.2.4	Previous studies of <i>I think</i> in the speech of Chinese NNSs	239
8.3	Frequency information in the speech of the non-native speakers and native speakers	240
8.3.1	Overall frequency of <i>I think</i>	240
8.3.2	Collocates of <i>I think</i>	243
8.4	Discourse aspects of <i>I think</i>	248
8.4.1	Positions in an utterance/turn	248
8.4.1.1	<i>I think</i> in extra-clausal position	249
8.4.1.2	<i>I think</i> in intra-clausal position	250
8.4.2	Contexts where Type B <i>I think</i> tends to occur	250
8.4.2.1	<i>I think</i> co-occurring with hesitation markers, pauses and restarts	251
8.4.2.2	<i>I think</i> co-occurring with personal opinions and evaluation	251
8.4.2.3	<i>I think</i> co-occurring with factual information	254
8.4.2.4	<i>I think</i> prefacing concluding remarks	256
8.4.2.5	<i>I think</i> prefacing questions	257
8.4.2.6	Summary of the contexts where Type B <i>I think</i> tends to occur	257
8.5	Further investigation	262
8.5.1	Importance of the type of activity	262
8.5.2	Alternatives to <i>I think</i> used by the non-native speakers	262
8.6	Chapter summary and conclusions	263
CHAPTER 9: ANALYSIS OF <i>NOW</i>		265
9.1	Introduction	265

9.2	Previous studies of <i>now</i>	267
9.2.1	Grammatical aspect: Word classes.....	267
9.2.2	Grammatical aspect: Syntactical structure	267
9.2.3	<i>Linear Unit Grammar</i> analysis of <i>now</i>	268
9.2.4	Previous studies of <i>now</i> as a discourse marker.....	268
9.3	Frequency information in the speech of the non-native speakers and native speakers.....	270
9.3.1	Overall frequency of <i>now</i>	270
9.3.2	Collocates of <i>now</i>	273
9.4	Discourse aspects of <i>now</i>	278
9.4.1	Positions in an utterance/turn	278
9.4.1.1	<i>Now</i> in extra-clausal position	279
9.4.1.2	<i>Now</i> in intra-clausal position	280
9.4.2	Contexts where Type B <i>now</i> tends to occur	280
9.4.2.1	<i>Now</i> occurring at the opening/closing of topic and concluding remarks.....	281
9.4.2.2	<i>Now</i> occurring at a shift of (sub)topic and viewpoint	281
9.4.2.3	<i>Now</i> prefacing a question	282
9.4.2.4	<i>Now</i> co-occurring with elaborations, explanations and exemplifications.....	282
9.4.2.5	<i>Now</i> co-occurring with listing items and sequence of events	283
9.4.2.6	<i>Now</i> co-occurring with contrasting items	283
9.4.2.7	<i>Now</i> co-occurring with emphatic lexical items and structure.....	284
9.4.2.8	<i>Now</i> co-occurring with indications of location/object	284
9.4.2.9	Problematic and unclassified instances of <i>now</i>	285
9.4.2.10	Summary of the contexts where <i>now</i> tends to occur.....	286
9.5	Chapter summary and conclusions	290
CHAPTER 10: TEXT-BASED ANALYSIS OF DISCOURSE MARKERS IN THE SPEECH OF THE NATIVE SPEAKERS		292
10.1	Introduction.....	292
10.2	Texts for analysis.....	292
10.3	Text-based analysis of discourse markers in the native speakers' speech.....	293
10.3.1	Unscripted speeches in the ICE-GB corpus	293
10.3.2	Private direct conversations in the ICE-GB corpus.....	296
10.3.3	Texts of highly monologic discourse mode in the MICASE corpus	299
10.3.4	Texts of highly interactive discourse mode in the MICASE corpus.....	309
10.4	Chapter summary and conclusions	313
CHAPTER 11: TEXT-BASED ANALYSIS OF DISCOURSE MARKERS IN THE SPEECH OF THE NON-NATIVE SPEAKERS		314
11.1	Introduction.....	314
11.2	Texts for analysis.....	314
11.3	Text-based analysis of discourse markers in the non-native speakers' speech.....	315

11.3.1 Monologues in the SECCL corpus.....	315
11.3.2 Dialogues in the SECCL corpus	319
11.4 Chapter summary and conclusions	324
CHAPTER 12: SUMMARY AND DISCUSSION	327
12.1 Introduction.....	327
12.2 Summary of the analyses of discourse markers.....	327
12.2.1 The proportions of the words/phrases as discourse markers	327
12.2.2 The frequencies of the words/phrases as discourse markers	328
12.2.3 Positional preference of discourse markers	329
12.2.4 Co-occurrence and suggested functions of discourse markers.....	330
12.2.5 Discourse marker collocations.....	334
12.2.6 The usefulness and limitations of <i>Linear Unit Grammar</i> in the analysis ..	336
12.3 The use of discourse markers	337
12.3.1 Factors in the use of discourse markers: genre dependency, context and activity type sensitivity and identity construction	338
12.3.2 Different uses of discourse markers in the Chinese NNSs' and NSs' speech.....	340
12.4 Implications for the investigation of discourse markers in a non-native variety of English.....	342
12.5 Implications for pedagogy in English language teaching	344
12.5.1 Pedagogical aspects of discourse markers.....	344
12.5.1.1 Making learners aware of the use of discourse markers in speech	345
12.5.1.2 Areas which may require pedagogical interventions	348
12.5.2 Pedagogical approaches to the instruction of discourse markers	350
12.5.2.1 Consciousness-raising activities and data-driven learning.....	350
12.5.2.2 Controlled exercises	354
CHAPTER 13: CONCLUSIONS	356
13.1 Introduction.....	356
13.2 Achievements and strengths of the thesis.....	356
13.3 Limitations and weaknesses of the research.....	357
13.4 Implications for future research	359
13.5 Concluding remarks	359
APPENDIX 1: Details of the two sub-corpora extracted from the Spoken English Corpus of Chinese Learners (SECCL)	361
APPENDIX 2: Details of the two sub-corpora extracted from the Michigan Corpus of Academic Spoken English (MICASE)	364
APPENDIX 3: Details of the two sub-corpora extracted from the International Corpus of English – Great Britain (ICE-GB)	367

APPENDIX 4: Labels in <i>Linear Unit Grammar</i> (Sinclair and Mauranen 2006)	371
APPENDIX 5: Frequency comparison of discourse markers in the manual classification and in the tagged ICE-GB	372
APPENDIX 6: The log-likelihood test of discourse markers between two types of genre and between Chinese non-native speakers and native speakers	373
APPENDIX 7: The z test for two proportions of discourse markers between two types of genre and between Chinese non-native speakers and native speakers	374
REFERENCES	375

List of tables

Table 2.1: Characteristics of discourse marker in the three approaches	28
Table 3.1: Corpora under investigation	38
Table 3.2: Discourse markers identified in the Chinese non-native speakers' speech.....	41
Table 3.3: Raw and normalised frequencies of the words and phrases for analysis in the six sub-corpora	43
Table 3.4: Distribution of <i>like</i> in the three sets of 100-line concordance samples.....	53
Table 3.5: <i>Like</i> as a discourse marker in MICASE: Highly interactive discourse mode (The 1st set of 100-line concordance samples)	55
Table 3.6: <i>Like</i> as a discourse marker in MICASE: Highly interactive discourse mode (The 2nd set of 100-line concordance samples).....	55
Table 3.7: <i>Like</i> as a discourse marker in MICASE: Highly interactive discourse mode (The 3rd set of 100-line concordance samples)	56
Table 3.8: <i>Like</i> as a discourse marker in MICASE: Highly interactive discourse mode (The three sets of 100-line concordance samples).....	56
Table 3.9: Table of codes	59
Table 4.1: Frequency information of <i>like</i> in the non-native speakers' and native speakers' speech.....	67
Table 4.2: Pattern of <i>like</i> in the non-native speakers' monologues in SECCL	69
Table 4.3: Pattern of <i>like</i> in the non-native speakers' dialogues in SECCL.....	69
Table 4.4: Pattern of <i>like</i> in the native speakers' highly monologic discourse mode in MICASE	70
Table 4.5: Pattern of <i>like</i> in the native speakers' highly interactive discourse mode in MICASE	70
Table 4.6: Pattern of <i>like</i> in the native speakers' unscripted monologues in ICE-GB	71
Table 4.7: Pattern of <i>like</i> in the native speakers' private direct conversations in ICE-GB.....	71
Table 4.8: Pattern of Type B <i>like</i> in the non-native speakers' monologues in SECCL	72
Table 4.9: Pattern of Type B <i>like</i> in the non-native speakers' dialogues in SECCL.....	72
Table 4.10: Pattern of Type B <i>like</i> in the native speakers' highly monologic discourse mode in MICASE	72
Table 4.11: Pattern of Type B <i>like</i> in the native speakers' highly interactive discourse mode in MICASE	72
Table 4.12: Pattern of Type B <i>like</i> in the native-speakers' unscripted monologues in ICE-GB	72

Table 4.13: Pattern of Type B <i>like</i> in the native-speakers' private direct conversations in ICE-GB.....	73
Table 4.14: Distribution of the positions of Type B <i>like</i> in an utterance/turn	74
Table 4.15: Distribution of co-occurrence of <i>like</i> as a discourse marker in SECCL: Monologues	86
Table 4.16: Distribution of co-occurrence of <i>like</i> as a discourse marker in SECCL: Dialogues	86
Table 4.17: Distribution of co-occurrence of <i>like</i> as a discourse marker in MICASE: Highly monologic discourse mode	87
Table 4.18: Distribution of co-occurrence of <i>like</i> as a discourse marker in MICASE: Highly interactive discourse mode	87
Table 4.19: Distribution of co-occurrence of <i>like</i> as a discourse marker in ICE-GB: Unscripted monologues	88
Table 4.20: Distribution of co-occurrence of <i>like</i> as a discourse marker in ICE-GB: Private direct conversations.....	88
Table 4.21: Concordance for <i>just like</i> in the Bank of English (Accessed on 10 August 2009).....	93
Table 5.1: Examples of the use of <i>oh</i> in the literature.....	100
Table 5.2: Frequency information of <i>oh</i> in the non-native speakers' and native speakers' speech.....	103
Table 5.3: Pattern of <i>oh</i> in the non-native speakers' monologues in SECCL	105
Table 5.4: Pattern of <i>oh</i> in the non-native speakers' dialogues in SECCL	106
Table 5.5: Pattern of <i>oh</i> in the native speakers' highly monologic discourse mode in MICASE	106
Table 5.6: Pattern of <i>oh</i> in the native speakers' unscripted monologues in ICE-GB	107
Table 5.7: Pattern of <i>oh</i> in the native speakers' highly interactive discourse mode in MICASE	107
Table 5.8: Pattern of <i>oh</i> in the native speakers' private direct conversations in ICE-GB	108
Table 5.9: Distribution of the positions of <i>oh</i> in an utterance/turn.....	110
Table 5.10: Distribution of co-occurrence of <i>oh</i> in SECCL: Monologues.....	123
Table 5.11: Distribution of co-occurrence of <i>oh</i> in SECCL: Dialogues.....	123
Table 5.12: Distribution of co-occurrence of <i>oh</i> in MICASE: Highly monologic discourse mode	124
Table 5.13: Distribution of co-occurrence of <i>oh</i> in MICASE: Highly interactive discourse mode	124
Table 5.14: Distribution of co-occurrence of <i>oh</i> in ICE-GB: Unscripted monologues	125
Table 5.15: Distribution of co-occurrence of <i>oh</i> in ICE-GB: Private direct conversations	125

Table 5.16: Percentages of types of co-occurrence of <i>oh</i> in the three sub-corpora of the dialogic genres	129
Table 5.17: Proportion of <i>oh</i> prefacing <i>what</i> - and <i>which</i> -questions in the three sub-corpora of the dialogic genres.....	130
Table 6.1: Frequency information of <i>well</i> in the non-native speakers' and native speakers' speech.....	143
Table 6.2: Pattern of <i>well</i> in the non-native speakers' monologues in SECCL	145
Table 6.3: Pattern of <i>well</i> in the non-native speakers' dialogues in SECCL.....	145
Table 6.4: Pattern of <i>well</i> in the native speakers' highly monologic discourse mode in MICASE	146
Table 6.5: Pattern of <i>well</i> in the native speakers' unscripted monologues in ICE-GB.....	146
Table 6.6: Pattern of <i>well</i> in the native speakers' highly interactive discourse mode in MICASE	146
Table 6.7: Pattern of <i>well</i> in the native speakers' private direct conversations in ICE-GB....	147
Table 6.8: Pattern of Type B <i>well</i> in the non-native speakers' monologues in SECCL.....	147
Table 6.9: Pattern of Type B <i>well</i> in the non-native speakers' dialogues in SECCL	148
Table 6.10: Pattern of Type B <i>well</i> in the native speakers' highly monologic discourse mode in MICASE	148
Table 6.11: Pattern of Type B <i>well</i> in the native speakers' unscripted monologues in ICE-GB	148
Table 6.12: Pattern of Type B <i>well</i> in the native speakers' highly interactive discourse mode in MICASE	149
Table 6.13: Pattern of Type B <i>well</i> in the native speakers' private direct conversations in ICE-GB.....	149
Table 6.14: Distribution of the positions of Type B <i>well</i> in an utterance/turn	151
Table 6.15: Distribution of co-occurrence of <i>well</i> as a discourse marker in SECCL: Monologues	165
Table 6.16: Distribution of co-occurrence of <i>well</i> as a discourse marker in SECCL: Dialogues	165
Table 6.17: Distribution of co-occurrence of <i>well</i> as a discourse marker in MICASE: Highly monologic discourse mode	166
Table 6.18: Distribution of co-occurrence of <i>well</i> as a discourse marker in MICASE: Highly interactive discourse mode	166
Table 6.19: Distribution of co-occurrence of <i>well</i> as a discourse marker in ICE-GB: Unscripted monologues	167
Table 6.20: Distribution of co-occurrence of <i>well</i> as a discourse marker in ICE-GB: Private	

direct conversations.....	167
Table 7.1: Frequency information of <i>you know</i> in the non-native speakers' and native speakers' speech	182
Table 7.2: Frequency information of <i>I mean</i> in the non-native speakers' and native speakers' speech.....	183
Table 7.3: Raw frequencies of <i>you know</i> and <i>I mean</i> in three different discourse modes in MICASE	186
Table 7.4: Frequency information of <i>you see</i> in the non-native speakers' and native speakers' speech.....	188
Table 7.5: Pattern of <i>you know</i> in the non-native speakers' monologues in SECCL	190
Table 7.6: Pattern of <i>you know</i> in the native speakers' highly monologic discourse mode in MICASE	190
Table 7.7: Pattern of <i>you know</i> in the native speakers' unscripted monologues in ICE-GB..	190
Table 7.8: Pattern of <i>you know</i> in the non-native speakers' dialogues in SECCL	191
Table 7.9: Pattern of <i>you know</i> in the native speakers' highly interactive discourse mode in MICASE	191
Table 7.10: Pattern of <i>you know</i> in the native speakers' private direct conversations in ICE-GB	192
Table 7.11: Pattern of Type B <i>you know</i> in the non-native speakers' monologues in SECCL	192
Table 7.12: Pattern of Type B <i>you know</i> in the non-native speakers' dialogues in SECCL ..	192
Table 7.13: Pattern of Type B <i>you know</i> in the native speakers' highly monologic discourse mode in MICASE.....	193
Table 7.14: Pattern of Type B <i>you know</i> in the native speakers' highly interactive discourse mode in MICASE.....	193
Table 7.15: Pattern of Type B <i>you know</i> in the native speakers' private direct conversations in ICE-GB.....	194
Table 7.16: Pattern of Type B <i>you know</i> in the native speakers' unscripted monologues in ICE-GB.....	194
Table 7.17: Pattern of <i>I mean</i> in the non-native speakers' dialogues in SECCL	195
Table 7.18: Pattern of <i>I mean</i> in the native speakers' highly interactive discourse mode in MICASE	195
Table 7.19: Pattern of <i>I mean</i> in the native speakers' private direct conversations in ICE-GB	196
Table 7.20: Pattern of Type B <i>I mean</i> in the non-native speakers' dialogues in SECCL	196
Table 7.21: Pattern of Type B <i>I mean</i> in the native speakers' highly interactive discourse	

mode in MICASE.....	197
Table 7.22: Pattern of Type B <i>I mean</i> in the native speakers' private direct conversations in ICE-GB.....	197
Table 7.23: Pattern of <i>you see</i> in the non-native speakers' dialogues in SECCL.....	198
Table 7.24: Pattern of <i>you see</i> in the native speakers' highly interactive discourse mode in MICASE	198
Table 7.25: Pattern of <i>you see</i> in the native speakers' private direct conversations in ICE-GB	198
Table 7.26: Pattern of Type B <i>you see</i> in the non-native speakers' dialogues in SECCL	199
Table 7.27: Pattern of Type B <i>you see</i> in the native speakers' highly interactive discourse mode in MICASE.....	199
Table 7.28: Pattern of Type B <i>you see</i> in the native speakers' private direct conversations in ICE-GB.....	200
Table 7.29: Distribution of the positions in an utterance/turn of Type B <i>you know</i>	202
Table 7.30: Distribution of the positions in an utterance/turn of Type B <i>I mean</i>	203
Table 7.31: Distribution of the positions in an utterance/turn of Type B <i>you see</i>	204
Table 7.32: Distribution of co-occurrence of <i>you know</i> as a discourse marker in SECCL: Monologues	218
Table 7.33: Distribution of co-occurrence of <i>you know</i> as a discourse marker in SECCL: Dialogues.....	218
Table 7.34: Distribution of co-occurrence of <i>you know</i> as a discourse marker in MICASE: Highly monologic discourse mode	219
Table 7.35: Distribution of co-occurrence of <i>you know</i> as a discourse marker in MICASE: Highly interactive discourse mode	219
Table 7.36: Distribution of co-occurrence of <i>you know</i> as a discourse marker in ICE-GB: Unscripted monologues.....	220
Table 7.37: Distribution of co-occurrence of <i>you know</i> as a discourse marker in ICE-GB: Private direct conversations	220
Table 7.38: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in SECCL: Monologues	222
Table 7.39: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in SECCL: Dialogues.....	222
Table 7.40: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in MICASE: Highly monologic discourse mode	223
Table 7.41: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in MICASE: Highly interactive discourse mode	223
Table 7.42: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in ICE-GB:	

Unscripted monologues.....	224
Table 7.43: Distribution of co-occurrence of <i>I mean</i> as a discourse marker in ICE-GB: Private direct conversations.....	224
Table 7.44: Distribution of co-occurrence of <i>you see</i> as a discourse marker in SECCL: Monologues.....	228
Table 7.45: Distribution of co-occurrence of <i>you see</i> as a discourse marker in SECCL: Dialogues.....	228
Table 7.46: Distribution of co-occurrence of <i>you see</i> as a discourse marker in MICASE: Highly monologic discourse mode.....	229
Table 7.47: Distribution of co-occurrence of <i>you see</i> as a discourse marker in MICASE: Highly interactive discourse mode.....	229
Table 7.48: Distribution of co-occurrence of <i>you see</i> as a discourse marker in ICE-GB: Unscripted monologues.....	230
Table 7.49: Distribution of co-occurrence of <i>you see</i> as a discourse marker in ICE-GB: Private direct conversations.....	230
Table 8.1: Frequency information of <i>I think</i> in the non-native speakers' and native speakers' speech.....	241
Table 8.2: Frequencies of <i>do you think</i> across corpora.....	243
Table 8.3: Pattern of <i>I think</i> in the non-native speakers' monologues in SECCL.....	244
Table 8.4: Pattern of <i>I think</i> in the non-native speakers' dialogues in SECCL.....	244
Table 8.5: Pattern of <i>I think</i> in the native speakers' highly monologic discourse mode in MICASE.....	245
Table 8.6: Pattern of <i>I think</i> in the native speakers' unscripted monologues in ICE-GB.....	245
Table 8.7: Pattern of <i>I think</i> in the native speakers' highly interactive discourse mode in MICASE.....	245
Table 8.8: Pattern of <i>I think</i> in the native speakers' private direct conversations in ICE-GB.....	246
Table 8.9: Pattern of Type B <i>I think</i> in the non-native speakers' monologues in SECCL.....	246
Table 8.10: Pattern of Type B <i>I think</i> in the non-native speakers' dialogues in SECCL.....	246
Table 8.11: Pattern of Type B <i>I think</i> in the native speakers' highly monologic discourse mode in MICASE.....	247
Table 8.12: Pattern of Type B <i>I think</i> in the native speakers' unscripted monologues in ICE-GB.....	247
Table 8.13: Pattern of Type B <i>I think</i> in the native speakers' highly interactive discourse mode in MICASE.....	247
Table 8.14: Pattern of Type B <i>I think</i> in the native speakers' private direct conversations in ICE-GB.....	247

Table 8.15: Distribution of the positions in an utterance/turn of Type B <i>I think</i>	249
Table 8.16: Distribution of co-occurrence of <i>I think</i> as a discourse marker in SECCL: Monologues	259
Table 8.17: Distribution of co-occurrence of <i>I think</i> as a discourse marker in SECCL: Dialogues	259
Table 8.18: Distribution of co-occurrence of <i>I think</i> as a discourse marker in MICASE: Highly monologic discourse mode	260
Table 8.19: Distribution of co-occurrence of <i>I think</i> as a discourse marker in MICASE: Highly interactive discourse mode	260
Table 8.20: Distribution of co-occurrence of <i>I think</i> as a discourse marker in ICE-GB: Unscripted monologues.....	261
Table 8.21: Distribution of co-occurrence of <i>I think</i> as a discourse marker in ICE-GB: Private direct conversations.....	261
Table 8.22: Frequencies of <i>I think</i> and <i>I thought</i> in the non-native speakers' speech	262
Table 8.23: Frequencies of alternatives to <i>I think</i> used by the non-native speakers	263
Table 9.1: Frequency information of <i>now</i> in the non-native speakers' and native speakers' speech.....	271
Table 9.2: Pattern of <i>now</i> in the non-native speakers' monologues in SECCL	273
Table 9.3: Pattern of <i>now</i> in the non-native speakers' dialogues in SECCL.....	273
Table 9.4: Pattern of <i>now</i> in the native speakers' highly monologic discourse mode in MICASE	274
Table 9.5: Pattern of <i>now</i> in the native speakers' highly interactive discourse mode in MICASE	274
Table 9.6: Pattern of <i>now</i> in the native speakers' unscripted monologues in ICE-GB.....	275
Table 9.7: Pattern of <i>now</i> in the native speakers' private direct conversations in ICE-GB....	275
Table 9.8: Pattern of Type B <i>now</i> in the non-native speakers' monologues in SECCL	276
Table 9.9: Pattern of Type B <i>now</i> in the non-native speakers' dialogues in SECCL.....	276
Table 9.10: Pattern of Type B <i>now</i> in the native speakers' highly monologic discourse mode in MICASE	277
Table 9.11: Pattern of Type B <i>now</i> in the native speakers' highly interactive discourse mode in MICASE	277
Table 9.12: Pattern of Type B <i>now</i> in the native speakers' unscripted monologues in ICE-GB	277
Table 9.13: Pattern of Type B <i>now</i> in the native speakers' private direct conversations in ICE-GB.....	278
Table 9.14: Distribution of the positions in an utterance/turn of Type B <i>now</i>	279

Table 9.15: Distribution of co-occurrence of <i>now</i> as a discourse marker in SECCL: Monologues	287
Table 9.16: Distribution of co-occurrence of <i>now</i> as a discourse marker in SECCL: Dialogues	287
Table 9.17: Distribution of co-occurrence of <i>now</i> as a discourse marker in MICASE: Highly monologic discourse mode	288
Table 9.18: Distribution of co-occurrence of <i>now</i> as a discourse marker in MICASE: Highly interactive discourse mode	288
Table 9.19: Distribution of co-occurrence of <i>now</i> as a discourse marker in ICE-GB: Unscripted monologues	289
Table 9.20: Distribution of co-occurrence of <i>now</i> as a discourse marker in ICE-GB: Private direct conversations	289
Table 10.1: Texts of the native speaker speech for analysis	293
Table 10.2: Devices signalling listeners' engagement in Excerpt 10.5.....	304
Table 10.3: Devices signalling listeners' engagement in Excerpt 10.6.....	307
Table 11.1: Texts of the non-native speaker speech for analysis	314
Table 12.1: Percentage of the words/phrases as discourse markers across corpora	327
Table 12.2: Frequency comparisons of the discourse markers in the monologic and dialogic genres under investigation (Normalised frequency per 10,000 words (times)).....	328
Table 12.3: Summary of the positions in an utterance/turn of the discourse markers in the monologic and dialogic genres under investigation	330
Table 12.4: Summary of the most frequent types of co-occurrence of the discourse markers in the monologic and dialogic genres under investigation	331
Table 12.5: Functions of discourse markers identified on the basis of co-occurrence	333
Table 12.6: Discourse marker collocations in the speech of the non-native speakers and native speakers under investigation	335
Table 12.7: Example material for consciousness-raising activities.....	352

List of figures

Figure 4.1: Comparison of normalised frequencies of <i>like</i> across sub-corpora	68
Figure 4.2: Plot of <i>like</i> in the sub-corpus of the NSs' highly interactive discourse mode in MICASE	89
Figure 5.1: Comparison of normalised frequencies of <i>oh</i> across sub-corpora.....	103
Figure 6.1: Comparison of normalised frequencies of <i>well</i> across sub-corpora	144
Figure 7.1: Comparison of normalised frequencies of Type B <i>you know</i> and <i>I mean</i> across sub-corpora.....	184
Figure 7.2: Comparison of normalised frequencies of Type B <i>you see</i> across sub-corpora.....	189
Figure 8.1: Normalised frequencies of <i>I think</i> across corpora	242
Figure 8.2: Normalised frequencies of Type B <i>I think</i> across corpora.....	242
Figure 9.1: Comparison of normalised frequencies of <i>now</i> across corpora	272
Figure 12.1: Concordance for <i>oh</i> , <i>well</i> and <i>you know</i> marking reported speech in MICASE (<i>WordSmith 4</i> , Scott 2004)	353
Figure 12.2: Example material for the use of <i>well</i> (McCarthy, McCarten and Sandiford 2006:39).....	355

List of abbreviations

BASE	The British Academic Spoken English
BAWE	The British Academic Written English
BNC	The British National Corpus
CANCODE	The Cambridge and Nottingham Corpus of Discourse in English
<i>CGE</i>	<i>Cambridge Grammar of English</i>
CIA	Contrastive Interlanguage Analysis
COBUILD	Collins Birmingham University International Language Database
COLSEC	The College Learners' Spoken English Corpus
DDL	data-driven learning
DM	discourse marker
ELT	English language teaching
HKCSE	The Hong Kong Corpus of Spoken English
ICE	The International Corpus of English
ICE-GB	The International Corpus of English – Great Britain
ICLE	The International Corpus of Learner English
KWIC	key-word-in-context
L1	first language
L2	second language
<i>LGSWE</i>	<i>Longman Grammar of Spoken and Written English</i>
LOB	The Lancaster-Oslo/Bergen Corpus
LOCNESS	The Louvain Corpus of Native English Essays
LSWE	The Longman Spoken and Written English Corpus
<i>LUG</i>	<i>Linear Unit Grammar</i>
MI	Mutual Information
MICASE	The Michigan Corpus of Academic Spoken English
MICUSP	The Michigan Corpus of Upper-level Student Papers
NNS	non-native speaker
NS	native speaker
SEC	The Lancaster/IBM Spoken English Corpus
SECCL	The Spoken English Corpus of Chinese Learners
SLA	second language acquisition
SWECCCL	The Spoken and Written English Corpus of Chinese Learners
TBTL	task-based teaching and learning
TEM 4	The Test for English Majors Band 4
VOICE	The Vienna-Oxford International Corpus of English

CHAPTER 1: INTRODUCTION

This thesis opens with an introduction to the present study and then provides a background for the research, including a discussion of the grammar of spoken English, discourse markers (hereafter DMs) and the use of corpora.

1.1 The present study

1.1.1 Rationale behind the thesis and its general aim

In the emergence of a number of corpora dedicated to spoken English (e.g. The Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens and Swales 2002), the British Academic Spoken English (BASE) Corpus (Nesi and Thompson 2006) and the Hong Kong Corpus of Spoken English (HKCSE) (Cheng and Warren 1999)), a voluminous literature has grown up on the investigation of spoken English. The use of DMs in speech, as opposed to and distinct from written English, has become a popular topic for research. The use of DMs made by native speakers (NSs) has been intensively examined with fruitful results (e.g. Schourup (1985), Schiffrin (1987), Fraser (1990, 1999), Jucker (1993), Lenk (1995), Biber, Finegan, Johansson, Conrad and Leech (1999), Aijmer (2002) and Carter and McCarthy (2006)). Some of the studies are corpus-based, some based on a small set of data and others use contrived examples. Little attention, however, has been paid to the DMs used by learners or non-native speakers (NNSs)¹.

Among the studies of learner language, it has been shown that learner corpora have for some considerable time been used to study the written mode of English (e.g. studies edited by Granger (1998c), Granger, Hung and Petch-Tyson (2002) and Meunier and Granger (2008)). However, fewer studies have been made of learners' spoken English. The compilation of a spoken language corpus consumes much more time and effort than one comprising written language, because spoken data have to be manually transcribed and keyboarded. Yet it is widely acknowledged that this is an area which needs to be further explored (Carter and McCarthy 1995). In addition, features of spoken English, which are very different from those of written English, have been neglected in English language teaching (ELT). On this aspect,

¹ The use of the two terms *learners* and *non-native speakers* relates to the issue of ideology, which is further discussed in Section 1.4.

the present thesis accordingly discusses some of the pedagogical implications for the teaching of DMs.

As noted above, the use of DMs is one of the distinct features in spoken English. At the time of writing, DM is a commonly-used term, but its terminology and definition are still open to debate. Moreover, previous studies have not empirically established the functions of DMs. Nor have they comprehensively investigated the use of DMs in Chinese learners'/NNSs' speech.

Researchers agree that DMs are used for particular functions, with no one-to-one correspondence between form and function. In other words, a DM can have several functions and the same function can be expressed by several DMs. Furthermore, DMs are produced without conscious knowledge. Their uses are not easily available to introspection and the empirical evidence provided in the literature is not clear, either. For these reasons, it is difficult to demonstrate unequivocally the function of any instance of a DM. Researchers therefore tend to use intuition in interpreting this function.

Given the background sketched out above, I present in this thesis a new process that uses collocation phenomena and co-text analyses to empirically derive the functions rather than interpreting them intuitively. I aim to use quantitative corpus methodologies, combining with *Linear Unit Grammar (LUG)* (Sinclair and Mauranen 2006) analysis and qualitative text-based analysis, to identify the use which Chinese learners/NNSs and NSs make of DMs.

This thesis starts bottom-up with a linguistic description of the words and phrases under investigation from the standpoints of grammar and discourse. Contexts and types of co-occurrence of DMs are taken as evidence to determine the categories for discussion; subsequently, the functions of DMs are suggested on the basis of the collocation phenomena and contextual information.

The use of DMs is discussed in connection with their positions in an utterance, which are described with the newly-established device, *LUG*, since traditional grammars are not able to accommodate a satisfactory description of spoken English and the labels in the grammars often imply that DMs have been used incorrectly. For example, DMs may be described as between fragments and between incomplete clauses. The terms *fragment* and *incomplete clause* seem to suggest that DMs are ungrammatical. In my research, the *LUG* analysis is employed to assign units in spoken English and to describe the positions of DMs in an utterance. This is one of the innovative aspects of my research; there have not so far been any

studies which use *LUG* in the investigation of spoken English.

1.1.2 Hypotheses and research questions

The research sets out to test hypotheses about the use which Chinese NNSs and NSs make of DMs. I hypothesise that the use of DMs is genre-dependent and culture-sensitive. The two general hypotheses proposed in this thesis are: 1) across the monologic and dialogic genres under investigation, the more interactive the genre or type of activity is, the more DMs occur and 2) the uses of DMs in the Chinese NNSs' speech under investigation are different from and are possibly not as varied as those in the speech of the NSs.

To test my hypotheses, the following questions are addressed in the analysis of each word/phrase for analysis:

1. What are the non-discourse use (hereafter referred to as Type A) and discourse use (hereafter referred to as Type B) of the words/phrases for analysis in the literature?
2. What is the distribution of Type A words/phrases and Type B words/phrases across the corpora under investigation?

Question 1 addresses the uses of the words and phrases for analysis identified in the previous studies. The answer to Question 2 provides evidence for whether or not DMs (Type B words and phrases) occur more often in interactive types of activity (i.e. the discussions and conversations used in the present study) than in the less interactive types of activity (i.e. the unscripted speeches, lectures, etc).

3. What is the distribution of the positions in an utterance/turn of DMs (Type B words/phrases) in the speech of the Chinese NNSs and the NSs under investigation?
4. What do collocates of the words/phrases for analysis reveal about their use in speech?
5. What DMs tend to occur together in the NNSs' and NSs' speech?

The answers to Questions 3 to 5 show the overall sense of the use of the words and phrases in the speech of the two groups of speakers.

6. What lexical items and collocation phenomena do the DMs tend to co-occur with?

The identified types of co-occurrence are used as discussion categories and they are also the basis for interpreting the functions of DMs.

7. How can *LUG*, Sinclair and Mauranen's (2006) linear approach to the description of discourse, be used to describe spoken English and applied to language teaching?

Using this newly-established device, *LUG*, is one of the innovative aspects of the present research. It demonstrates that the elements in *LUG* can be used to describe units in spoken English.

8. How and why do DMs occur more in one text than another?
9. Are successful users of DMs also fluent speakers?

Questions 8 and 9 are answered in text-based analysis, which is employed to examine DMs from a complementary standpoint in relation to broader contexts and to test some hypotheses which cannot be tested with corpus methodologies.

10. How can the findings be applied to classroom use? What pedagogies can be used to teach DMs in spoken English?

Not all the findings are of relevance to learners of English. Some findings may have implications for pedagogy and some for the improvement of both NSs' and NNSs' understanding of the use of DMs in a (non-)native variety of global English.

In addition to the above hypotheses and research questions, specific hypotheses for each word/phrase for analysis and for text-based analyses will be formulated in the relevant chapters.

1.1.3 Significance and contribution of the research

This research is highly significant because it is one of the first studies to investigate DMs in the NNS speech. It contributes to a small but growing amount of research into the speech of Chinese speakers of English. It is innovative because of the use of corpus methodologies and the *LUG* analysis (Sinclair and Mauranen 2006). The research will benefit English teaching to Chinese speakers of English, help raise awareness among NNSs and prevent misunderstanding as it facilitates inter-cultural communication in English. In a broader sense, this research will add to the knowledge of under-researched areas and, it is hoped, make a major contribution to the modelling of NNS English and the investigation of global English varieties.

1.1.4 Outline of the thesis

Chapter 2 presents a survey of the literature on DMs and finds the five characteristics of DMs dealt with in this study. It also provides an overview of the previous studies of DMs and introduces the *LUG* analysis (Sinclair and Mauranen 2006), which is a new approach to the

description of the English language.

Chapter 3 describes the six sub-corpora under investigation, which are extracted from three publicly available corpora, the Spoken English Corpus of Chinese Learners (SECCL), the Michigan Corpus of Academic Spoken English (MICASE) and the International Corpus of English – Great Britain (ICE-GB); it describes how DMs are selected for analysis and how the quantitative and qualitative methods are pursued.

Chapters 4 to 9 are devoted to the eight words and phrases *like* (Chapter 4), *oh* (Chapter 5), *well* (Chapter 6), *you know*, *I mean*, *you see* (Chapter 7), *I think* (Chapter 8) and *now* (Chapter 9) respectively. Each chapter begins with the review of relevant literature on the given word/phrase, followed by frequency information from corpus analysis in the six sub-corpora before further examining the discursal aspects of the word/phrase as a DM (Type B).

Text-based analyses of DMs in the NNSs' and NSs' speech are reported in Chapters 10 and 11. Each chapter has its own hypotheses to be tested. Chapter 12 provides a summary of the description and analyses of DMs as well as discussing the implications of the research. The final chapter assesses the thesis as a whole, pointing to its achievements, strengths and limitations. It also provides suggestions for future work with concluding remarks.

1.2 The grammar of spoken English

In Applied Linguistics, there has been a bias towards the grammar of written English. The written mode of English has been the focus of research and ELT. In spite of the fact that speech precedes writing (Fromkin and Rodman 1974, Fromkin, Rodman and Hyams 2007), the study of language has mostly been based on written forms. Because of this, such researchers as Carter and McCarthy (1995) argue that a fair number of common uses in spoken English have not been fully described or have been treated as inferior to written forms. Similarly, as Sinclair and Mauranen (2006) maintain, the spoken mode of a language is used before the written mode exists and the emphasis in descriptions of language should be put on spoken language.

Some attempts to elucidate the grammar of spoken English include Brazil (1995), who offers an “exploratory”, “purpose-driven” and “linear” grammar of speech which is opposed to sentence-oriented grammars, Biber *et al.* (1999), who provide contrasting distributions of actual use in four spoken and written registers (academic prose, conversation, news and

fiction) and Carter and McCarthy (2006), who offer a comprehensive explanation of different grammatical usages in spoken and written modes, paying greater attention to spoken grammar.

Brazil (1995) argues that features in speech cannot be described by the standard of grammatical correctness in traditional sentence-based grammars. His exploratory grammar begins by testing assumptions against observable evidence. The assumption made for sentence grammars is that speakers use strings of words to produce linguistically correct sentences. In contrast, the assumption made for the grammar of speech is that speakers use strings of words to accomplish particular communicative purposes. The grammar proposed by Brazil is driven by purposes, not free-standing sentences. The grammar also has the notion of linearity, which refers to the fact that the meanings are created increment by increment as the speech goes on. Brazil's linear description of syntax influences Sinclair and Mauranen's *LUG* (2006), which is adopted in this thesis and is discussed in more detail in the next chapter.

Biber *et al.*'s work, *Longman Grammar of Spoken and Written English (LGSWE)* (1999), is based on the Longman Spoken and Written English (LSWE) Corpus, consisting of more than 40 million words from Britain and the United States in four registers: spoken conversation, fiction, news and academic prose. Word classes, syntactical constructions and lexical expressions are discussed in terms of a quantitative analysis and a functional analysis of the quantitative results. The corpus findings draw distinctions in the use of particular items between the four registers, demonstrating how an item tends to appear more in one register than another. In this grammar book, there is a chapter dedicated to the grammar of conversation which characterises the grammar of spoken English. Features exclusive to conversational English are highlighted, such as the phenomena of "dysfluency", "dislocation", "hesitations", "incompletion" and "ellipsis" (Biber *et al.* 1999: 1037-1125). Some terms of the meta-language used in this work seem to imply that spoken language in use is flawed. Carter and McCarthy use more neutral terms in their *Cambridge Grammar of English (CGE)* (2006). For example, "prefaces" are discussed under the heading of "dislocation" in Biber *et al.*'s *LGSWE* (1999), but are called "heads" in Carter and McCarthy's *CGE* (2006).

Based on naturally-occurring data in the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), Carter and McCarthy (2006: 4-10) take a descriptive approach to the grammar of standard British English and make distinctions between deterministic and probabilistic grammar rules and between grammar as structure and as

choice. They also point out the important influence of discourse and context on the grammar of spoken English.

These two writers (2006: 164) define four main characteristics of spoken language, which have distinguished themselves from the written form: 1) “spoken language happens in real time and is typically unplanned”; 2) “spoken language is most typically face to face”; 3) “spoken language foregrounds choices which reflect the immediate social and interpersonal situation”; and 4) “spoken language and written language are not sharply divided but on a continuum”. Biber (2006) in his multi-dimensional approach adds that it is not a single continuum but several interlocking continua.

Common features of spoken English include the following five categories: 1) deictic expressions, 2) situational ellipsis, 3) headers, tails and tags, 4) discourse markers and 5) polite and indirect language, vague language and approximations (Carter and McCarthy 2006). The use of these five categories is common in spoken English but rare in written English.

Among the above-mentioned features of spoken English, not every category can be investigated using corpus methodologies which depend on automatic retrieval in the (un)annotated corpora under investigation in this study. Deictic expressions are very context-dependent and -bound; it is not, therefore, easy to examine them in their immediate contexts in corpus study. The nature of situational ellipsis cannot be investigated, since corpus methodologies often involve an item to be searched for and the absence of words provides nothing to look at in a corpus. Headers, tails and tags are not easy to identify as specific items to look for in a corpus. The two remaining categories, 1) DMs and 2) polite and indirect language, vague language and approximations, seem better candidates for corpus analysis. DMs in NNS and NS speech are selected for investigation in the present thesis, in part because DMs are frequently used in spoken English and very unlikely to be found in formal English writing and in part because they are probably among the most difficult features of spoken English to explain to students and have consequently been typically ignored in curricula in places where English is taught as a foreign or second language.

1.3 Discourse markers in spoken English

The use of DMs is one of the key features of spoken English. The use of some DMs (e.g. *well*, *you know* and *I mean*) is often criticised as indicating poor fluency or speech in general (Crystal 1988, Watts 1989, O'Donnell and Todd 1991). However, DMs are constantly used by

speakers and play a significant role in speech, in particular in spontaneous speech. Crystal (1988) maintains that if used appropriately, DMs act as a lubricant to refine the interaction between speakers and that DMs should not be associated with an undesirable and overused style of speaking.

This thesis maintains that DMs are frequently used in speech and does not criticise their overuse in speech. In Biber's project of the investigation of spoken and written language used in university (2006: 66-70), such DMs as *ok*, *well*, *now* and *so* rarely occur in the written mode of English, but they are commonly used in all of the five spoken registers (classroom teaching, class management, office hours, study groups and service encounters).

In addition to the fact that DMs occur more often in spoken than in written language, it has also been found that they are more likely to be used in situations where more than one speaker is involved. In Stenström's study (1990), the occurrences of DMs are more than ten times as frequent in dialogues as in monologues. The author concludes that in conversations, lack of DMs makes speech dull and in monologues, it makes it unappealing to listeners.

DMs, as a matter of fact, serve as more than attention catchers. They are not optional extras in speech. DMs function more than decorations, for example, contributing to fluency. Hasselgren (2002) reports that there is a correlation between the use of "smallwords" (including DMs) and fluency. She provides evidence that in learners' speech the more "smallwords" in the native-like way, the greater the fluency. The working definition of fluency in Hasselgren's study is as follows:

Fluency: the ability to contribute to what a listener, proficient in the language, would normally perceive as coherent speech, which can be understood without undue strain, and is carried out at a comfortable pace, not being disjointed, or disrupted by excessive hesitation (2002: 148).

In this sense, fluent speakers are able to make coherent speech with appropriate segments and links to facilitate listeners' understanding. This usually involves marking or signposting the discourse and DMs have been found to perform vital functions, such as signalling, rephrasing, marking information and organising discourse.

Considering the value of DMs in speech and their contributions to interaction, fluency and the organisation of discourse, I would therefore maintain that NNSs should be able to interpret NSs' use of DMs and to use them in appropriate contexts, since the misuse or non-use of DMs would possibly be interpreted wrongly by NSs and lead to a weak interaction between speakers. The importance for NNSs to be competent in using DMs is highlighted by

Svartvik:

[I]f a foreign language learner says *five sheeps* or *he goed*, he can be corrected by practically every native speaker. If, on the other hand, he omits a *well*, the likely reaction will be that he is dogmatic, impolite, boring, awkward to talk to etc, but a native speaker cannot pinpoint an 'error' (1980: 171).

1.4 The use of native speaker and learner/non-native speaker corpora

Corpora have for a long time been used to study the written mode of a language, as written languages are easier to compile than spoken languages, which call for time-consuming and laborious transcribing process. Recently, the emergence of a number of corpora dedicated to spoken language (e.g. MICASE, BASE and CANCODE) has rapidly increased studies of spoken English. Corpora have been applied in several areas in Applied Linguistics, for example, the production of reference books and packages for learners and writers, critical linguistics, translation studies, literary studies and stylistics, forensic linguistics and ELT (Hunston 2002). Results from corpus research have made a substantial contribution to four areas in ELT: syllabus design, course design, materials and language teaching methodology.

Corpora have been available for more than forty years. The studies of corpora have changed the ways in which we look at language and what the language is like. Until the recent decade, corpora have received the attention they deserve and have informed syllabus design (Sinclair 2004b: 1). For instance, the lexical syllabus originally proposed by Sinclair and Renouf (1988) and later highlighted by Willis (1990) lists lexical items for learning in order of frequency. The frequently-used items are presented with their phraseologies and contexts.

In addition to contributing to syllabus design, corpus research helps teachers and curriculum developers tailor courses to meet learners' needs. Comparing learner corpora with NS ones has identified the aspects which learners have difficulty with and the aspects they already know (e.g. studies in Granger (1998c), Tankó (2004)). Small and specialised corpora have been used to investigate language in a particular domain and this helps learners to understand the disciplinary use of English (Tribble 1997: 112-113).

Corpora have been used to develop teaching and learning materials, which have enriched language teaching and learning. Based on corpora and the task-based learning and teaching (TBLT) approach, *Collins COBUILD English Course* series (Willis and Willis 1988) is probably the first coursebook of this kind. Recently, corpus-informed coursebooks, such as the *Touchstone* series (McCarthy, McCarten and Sandiford 2006) and *Top Notch* series (Saslow and Ascher 2006) have become available. Studies of NS and learner corpora have

also aided the development of dictionaries for learners, e.g. Collins COBUILD (e.g. *Collins COBUILD Advanced Learner's English Dictionary* (Collins COBUILD Advanced Learner's English Dictionary 2006)) and Longman (e.g. *Longman Dictionary of Contemporary English* (Longman Dictionary of Contemporary English 2009)) dictionaries. In the development of these teaching and learning materials, NS corpus studies offer information about frequency, collocations and NS usages. Learner corpus studies reveal frequent learner errors for teachers to look out for. Additionally, corpora are good sources for making one's own materials to enrich learning (Aston 1997). Materials prepared by teachers, if based on concordance lines, give learners chances to explore the language and notice what can be learned about the node word(s) (Hunston 2002: 177-179).

In addition to the areas of syllabus design, course design and materials, corpora influence teaching methodology. An often-quoted approach relating to the use of corpora is Johns's data-driven learning (DDL) (1991). The principle of DDL is for learners acting as researchers and working on corpus data to make discoveries about language instead of learning from the explicit instructions of the teacher. As Hunston (2002: 184-187) suggests, DDL can be used as a consciousness-raising activity in the Language Focus stage of the TBTL cycle.

Not only do corpora make contributions to the areas of ELT, but they can be used also to investigate the localisation of English. In a non-native variety of English, the use of DMs in speech is one of the important features. It is a good candidate for investigating the localisation of English because 1) using DMs says nothing about right and wrong in relation to syntax and semantics; 2) neither sentence grammar nor semantics has had much to say about DMs; and 3) the functions of DMs are difficult to define and they are probably culture specific and context dependent.

I am now to draw a distinction between the two terms *learners* and *NNSs*. The Chinese speakers of English in the SECCL corpus used in this study are, on the one hand, learners and on the other, NNSs because the variety of language they are speaking is quite possibly a non-native variety of English. Two views are taken in this thesis. In some contexts, in particular in the research which considers the subjects' competence vis-à-vis NS competence and in the discussion of pedagogy, it makes sense to talk about the speakers as learners. However, in my study of DMs, the problem with addressing the Chinese speakers of English as learners is that this implies that their use of DMs is not taken as a target feature of non-native English and they are incompetent at using DMs. To avoid implanting prejudice to

my Chinese speaker data, I tend then to call them *NNSs*, which means that I can examine the use of DMs without making a judgment about accuracy and correctness. In the chapter of pedagogical implications and applications, they are called *learners*.

The issue of whether NNSs should speak like NSs in learning spoken English is still under debate. This has been shown in the increasing interest in English as a lingua franca (e.g. Mauranen (2003b, 2006, 2007), Prodromou (2008) and Seidlhofer (2001, 2004)). I argue in this thesis that while NNSs should not be punished for not speaking like NSs, competent NNSs should use the target language appropriately in the given context and if they wish, they can keep their cultural identities, as long as their NNS language causes no misunderstanding and impedes the development of no interpersonal relations. Meanwhile, they should be aware of how NSs express themselves in a particular context.

This thesis argues that in addition to teaching Chinese users of English to use DMs differently, NSs should be encouraged to acknowledge the use of DMs made by NNSs and the global community of English speakers can be educated in the acceptance of a local variety. Taking the use of DMs made by NNSs as a feature of a local variety raises the question of what the target norm of the teaching and learning of English should be. As has been argued, NNSs do not need to sound like NSs (Prodromou 1996) and since English has become the lingua franca in Asia and other regions of the world, it is worth reconsidering whether the NS usages of DMs should be taken as the targets for NNS who want to improve their communicative competence or whether NNS usages should be acknowledged as acceptable features of an emerging variety of English in their own right.

This chapter has outlined the present study and given background information; it will be followed by a comprehensive survey of the relevant literature on Corpus Linguistics, DMs and the *LUG* analysis (Sinclair and Mauranen 2006).

CHAPTER 2: LITERATURE SURVEY

There are five sections in this chapter. It begins with an introduction to corpora and the common techniques used to consider the information derived from corpora. Next, it reports two main trends observed in the studies of learner corpora. It goes on to provide a detailed discussion of the characteristics of DMs and the approaches to the definitions of DMs in the literature. This discussion then turns to the characteristics of the DMs for analysis in this thesis. The last two sections discuss previous studies of DMs and the *LUG* analysis (Sinclair and Mauranen 2006), which is used in the present study to assign units in spoken English and describe the positions of DMs in an utterance.

2.1 Corpora as a revolution in language study

2.1.1 Corpora in use

Over the last few decades, the surge in recent corpus studies has come about in part because of the development of information technology. The storage capacity of computers, together with today's sophisticated concordancing software tools, has begun to have a significant impact on language research (Sinclair 1991).

Early research on corpora focused on written English and contributed a great deal of corpus-based grammatical description and explanation. Arguably, the most influential research has been the Collins Birmingham University International Language Database (COBUILD) project, which has changed the ways of describing language and contributed to the publication of a series of Collins COBUILD reference books and materials (e.g. *Collins COBUILD English Language Dictionary (Collins COBUILD English Language Dictionary)*, *Collins COBUILD English Grammar* (1990) and *Collins COBUILD Advanced Learner's English Dictionary* (5th ed.) (2006)). The COBUILD project produced the first corpus-based language dictionary in the world. Its influence on lexicography can easily be seen in the widespread applications of corpora in the dictionaries and reference books published by major publishers, such as Cambridge, Oxford, Longman and Macmillan.

There are several types of corpora. Corpora, compiled for unspecific research purposes, are usually referred to as general corpora, or reference corpora, for example, the Bank of

English² and the British National Corpus (BNC). This type of corpus is usually larger than other corpora. The former is composed of 450 million words of written and spoken texts and the latter 100 million words. The first-generation general corpora are the Brown Corpus and the Lancaster-Oslo/Bergen (LOB) Corpus, compiled in the 1960s. The Brown Corpus consists of 500 samples of written American English, each of about 2,000 words, and the LOB Corpus is a British English counterpart (Kennedy 1998, Hunston 2002).

Some corpora are designed for specific purposes. Comparable corpora of different languages or different varieties of English (e.g. the International Corpus of English (ICE), one million words each of different varieties of English) are compiled for language learning and the investigation of English varieties. Diachronic corpora (e.g. the Helsinki Corpus) are used for the investigation of the longitudinal development of a language. Specialised corpora (e.g. the British Academic Written English (BAWE) corpus and the Michigan Corpus of Upper-level Student Papers (MICUSP)) aim to collect representative data of a given set of genres. Learner corpora (e.g. the International Corpus of Learner English (ICLE)) are for the analysis of inter-language, primarily for pedagogical purposes. Learner corpus research will be further discussed in Section 2.2.

On the whole, in terms of quantity, corpora of written English are larger and more numerous than those of spoken English. However, the beginning of spoken English corpora can be traced back to 1959, when the Survey of English Usage, led by Quirk, began to investigate spoken English manually on paper. In the late 1970s, the spoken component of Quirk's Survey was computerised by Svartvik and became known as the London-Lund Corpus (Krishnamurthy 2004b). The first electronic corpus of spoken English is arguably the corpus of conversation for the COBUILD project (Krishnamurthy 2004a: xiv). Since then, a number of spoken English corpora have been compiled, for example, the Lancaster/IBM Spoken English Corpus (SEC) (Knowles, Taylor and Williams 1992), the Wellington Corpus of Spoken New Zealand English (Holmes, Vine and Johnson 1998), the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (Carter and McCarthy 1995, 1997), the Michigan Corpus of Academic Spoken English (MICASE) (Simpson *et al.* 2002) and the Vienna-Oxford International Corpus of English (VOICE, version 1.0 online) (Seidlhofer,

² The Bank of English corpus is jointly owned by HarperCollins Publishers and the University of Birmingham. In 2007 the corpus stood at 450 million words. The Collins Wordbanks Online offers a 553-million-word corpus of contemporary written and spoken text, which can be subscribed at <http://www.collinslanguage.com/wordbanks/> (accessed on 24 February 2011).

Breiteneder, Klimpfinger, Majewski and Pitzl 2009).

Before the exploitation of corpora, language research was based mainly on intuited and elicited data, termed the “armchair” and “laboratory” methods by Clark and Bangerter (2004). Corpora have made it possible to investigate large quantities of naturally-occurring data, called the “field” method. Corpus analysis gives us a whole new perspective on language descriptions. As Sinclair maintains, from the beginning of the COBUILD project, the research team has found results which were different from, or even contradicted, NSs’ intuition (Krishnamurthy 2004b). The “field” method has been used to investigate many areas, for example:

- lexicography and the development of dictionaries, with particular focus on collocation (e.g. Sinclair (1991, 2004c))
- the interdependence of lexis and grammar (e.g. pattern grammar (Hunston and Francis 1999)),
- the structure of language (e.g. Altenberg (1993)),
- register variations (e.g. Biber (1988), Biber and Finegan (1991) and Carter and McCarthy (2006)),
- disciplinary variations (e.g. studies in the volume edited by Hyland and Bondi (2006)) and
- inter-language (e.g. studies in the volumes edited by Granger (1998c), Granger, Hung and Petch-Tyson (2002) and Meunier and Granger (2008)).

2.1.2 Information derived from corpora

Corpus methodologies depend mainly on automatic searches of the linguistic item(s) in question. A raw corpus allows researchers only to search for individual wordforms rather than lemmas. A tagged corpus can be used to search for words in terms of their parts of speech and a parsed corpus can be used to look for words/phrases in syntactical structures (Hunston 2002: 18-20). Either raw or annotated corpora, investigated using software packages, such as *WordSmith Tools* (Scott 2004, 2008) and *AntConc* (Anthony 2007), can provide useful statistics about the frequency of linguistic item(s) in question, as well as sorting and re-ordering the item(s) to facilitate the identification of collocation and phraseology (Hunston 2002: 3-13). These processes can be carried out quickly and accurately by software tools on

computer, thus avoiding discrepancies from human interventions in the data-processing.

Quantitative measurements are an important aspect of corpus analysis. In general, frequency lists tell us what words occur more often in a particular corpus. By comparing the frequency lists of two corpora, distinctive words can be identified, which helps make interpretations or further investigations. For example, in Biber *et al.*'s (1999) description of language, frequency information about the word classes, syntactical constructions or lexical expressions under investigation is compared across four registers: spoken conversation, fiction, news and academic prose, demonstrating how a certain item has a tendency to appear in a given register. In lexicography, frequency data help in the selection of words and phrases and clarifies why priority given to one sense or another. In learner language studies, frequency information is usually compared with that in an NS corpus or in a learner corpus of different first language (L1).

Corpora have made it possible to identify linguistic phenomena which cannot be empirically observed in one or a few instances and only become noticeable when a wide range of evidence is presented all at once. The most widely used format for corpus analysis is a key-word-in-context (KWIC) concordance, in which a searched item (referred to as a node word) is displayed in the centre of each concordance line and the co-texts on either side of the node word are presented. All the concordance lines of a node word can easily be re-listed in alphabetical order according to the collocates to the right or left of the node word (Sinclair 1991). The concordance processing can primarily facilitate two kinds of observation: collocation and phraseology.

Collocation is the statistically significant co-occurrence of two or more words (Sinclair 1991: 170). Two of the most commonly used measures of statistical significance are the Mutual Information (MI) score and t-score. The value of an MI score gives the strength of collocation and that of t-score the certainty. The t-score is dependent on and affected by the size of the corpus, but the MI score is not; therefore, MI scores can be used to compare the values across corpora (for more discussion of the calculation and uses of MI score and t-score, see Hunston (2002: 69-75)). Krishnamurthy (1987) stresses the importance of collocation in the development of the dictionary for making decisions on the semantic categories of a headword and their order for listing and also identifying phrases.

In addition to the identification of significant collocates of the node word, phraseology can be observed in the concordance lines (Hunston 2002: 9). Standard software tools, such as

Concord in *WordSmith Tools* (Scott 2004), can be used to produce patterns of the node word(s), which are usually displayed in a 5:5 span³ (five words on either side) to facilitate the observation of phraseology. The tool also offers the function of automatic counting of collocational patterns, i.e. recurrent strings of words, called *clusters* in *WordSmith Tools* (Scott 2004).

The present study uses several functions in software tools, such as counting tokens of a corpus and frequency of the node word(s), sorting concordancing lines and listing collocates and clusters. In addition, what I have done manually is to classify concordance lines and set them in groups for re-sorting. I also used the source text retrieval function, as most software tools permit, to see bigger co-texts. Referring back to source texts made it easier to identify co-occurring linguistic items of DMs. When the instances of DMs were too numerous for manual analysis, software tools were used to extract random samples. More details about this process are given in the next chapter. As previously mentioned, these functions can handle a large number of data and make it possible to discover linguistic phenomena when many instances are presented together. These corpus methodologies also help to direct me to selective texts for qualitative text-based analyses (see Chapters 10 and 11).

2.2 Two trends in research on learner corpora

The research into the NS corpora can be traced back to the 1960s (Sinclair 1991, Kennedy 1998). In the late 1980s, learner corpora began to be compiled and to attract research interest. Learner corpora contain a large amount of learners' language output. Studies of learner corpora, as argued by Mark ((1998) cited in Granger (2002: 6-7)), have established their position in English language teaching (ELT) and second language acquisition (SLA) and indeed are closely related to the three areas of mainstream research – descriptions of the target language, characteristics of learners and language teaching methods. Over the past decade, it has been recognised that learner corpora are of use in the fields of ELT and SLA, mainly because they show how the language is actually used by learners (Granger 2002: 5).

Since the late 1980s, there has been a growing amount of work on learner corpora. Two research trends have emerged: first, comparative study between NS and learner language for

³ Sinclair *et al.* in the *OSTI Report* (Sinclair, J., R. Daley and S. Jones. 1970. *English Lexical Studies*. Report No. 5060. London: Office of Scientific and Technical Information.) explain that the optimal span was 4:4, as had been statistically calculated by Bob Daley. In the late 1990s, the research team, based on a much larger corpus, re-calculated and found that a 5:5 span might improve the semantic relevance of the node word(s) (Krishnamurthy 2004b).

ELT purposes and, second, error analysis and diachronic study for SLA purposes.

2.2.1 Comparative study for English language teaching purposes

In learner language research, there has been a consistent focus on two types of comparison: 1) between NS and learner languages and 2) between speakers of different mother tongues. This is what Granger (1998a: 12-13, 2002: 12-13) terms *Contrastive Interlanguage Analysis* (CIA). Learner corpora may be compiled in conjunction with NS corpora to identify underuse and overuse of linguistic items and to gauge learners' problems with usages.

Much research using corpora has been done by comparing the International Corpus of Learner English (ICLE) with the Louvain Corpus of Native English Essays (LOCNESS), a 300,000-word corpus of essays written by NSs. The data of both corpora are university students' argumentative writing (Granger 1998a: 13). The ICLE corpus and the LOCNESS corpus have been employed to investigate aspects of lexis, discourse and the grammar of learners' English (Granger 1998b: xxi) (see various studies in the volumes edited by Granger (1998c) and Gilquin, Papp and Díez-Bedmar (2009)). Other pairs of corpora have also been compared. Shirato and Stapleton (2007) compare the vocabulary in Japanese learners' conversations with that in the conversation component of BNC. Their study reveals that learners underuse certain lexical items, which are distinctive features in spoken English and overuse some auxiliary verbs and adjectives. In addition to the exploitation of the publicly available learner corpora, small specialised learner corpora and their comparable NS corpora have been compiled for particular pedagogical purposes, such as Ackerley's study (2008) on learners' report writing and Dalziel and Helm's investigation (2008) of learners' use of modal verbs in online writing.

Similar to NS corpus research, there is an increasing emphasis on phraseology and collocation in learner language research. There are many examples in the literature. Altenberg and Granger's contrastive study (2001) combines the two types of comparison, between NS and learners and between learners with different first language (L1) backgrounds. They compare French- and Swedish-speaking advanced learners' use of *make* with their NS counterparts in LOCNESS. They identify the different overuse and underuse of the lexical and grammatical construction of *make* in these two groups of learners. A follow-up study using two parallel corpora, English-Swedish and Swedish-English, by Altenberg (2002) ascribes Swedish learners' overuse to intra-lingual influence or inter-lingual influence. Other examples

are Nesselhauf's investigation (2004) of verb-noun collocations in the German-learner component of ICLE, with reference to dictionaries and BNC and Aerselaer's contrastive study (2008) of English-Spanish interpersonal discourse phrases in novice and expert writers' argumentative texts.

Comparison between two or more groups of learners with different L1s also attracts research attention. For instance, Tankó's investigation (2004) of adverbial connectors indicates that, compared with French and Swedish learners and NSs, Hungarian learners have similar usages. They use fewer types of connectors than NSs and overuse enumerative (e.g. *first* and *second*) and additive adverbials (e.g. *also* and *furthermore*). However, they also tend to use a large number of resultive (e.g. *therefore* and *thus*) and contrastive connectors (e.g. *however*), which are typical of argumentative discourse. Another example of comparative studies of learner language with different L1s is Paquot's work (2008), which examines the phraseological patterns of such lexical items as *for example*, *for instance*, *example*, *illustrate* and *exemplify* in five sub-corpora of the ICLE corpus and compares the uses in the learner data with the NS counterparts in the LOCNESS corpus. Paquot also discusses the underuse and overuse phenomena found in different groups of learners and identifies uses which are probably due to L1 transfer.

For ELT purposes, learner corpora studies inform English learning coursebooks and reference books. For example, the writers of the *Top Notch* series coursebooks (Saslow and Ascher 2006) claim that their books are informed by both NSs' and learners' usages, identified in the 328-million-word Longman Corpus Network. The teacher's edition of *Top Notch* (Saslow, Ascher and Tiberio 2006) provides frequent learner errors, so that teachers can be alert to the vocabulary and structures which give learners difficulty. Learner corpora research also contributes to the development of reference books (e.g. dictionaries published by Longman and Cambridge). The second edition of the *Macmillan English Dictionary for Advanced Learners* (2007) collaborates with the Centre for English Corpus Linguistics (Université catholique de Louvain, Belgium) to produce a 30-page academic writing section, which is based on sixteen learner corpora of argumentative essays in the ICLE corpus and the academic sub-corpus of BNC.

Most of the comparative studies offer insights into learners' linguistic knowledge and suggest that the findings from learner corpora research have pedagogical implications and applications. Nevertheless, not many studies report empirical evidence of the actual impact of

learner corpora research. Granger and Meunier (2008) have issued a call for such evidence. They stress the importance of phraseology in language teaching and learning and urge that more action should be taken in the classroom and also discuss the challenges which lie in the relevant fields of language teaching and learning.

In studies which compare NS and learner languages, one of the key notions is that of underuse and overuse of language features, which have been studied in research into learners' language since the pioneering work by the contributors to Granger's edited volume (1998c). The use of these terms, *underuse* and *overuse*, seems to assume that NS language is taken as the target norm for language learning. Saying a particular linguistic item is overused or underused by learners means that learners use the target item too much or too little. If this assumption is valid for comparing a learner variety with a particular NS language, it raises the issue of comparability, which is important and under-discussed in the literature.

Strictly speaking, it is very difficult to compile two comparable corpora, ensuring the same method of data collection, genre and context for a given use of language. This observation attracts some reservations about the interpretation of frequency information, either approximate to or deviating from the frequency of NS usages. In the light of this, further examination of the data can be of great help.

The underuse and overuse phenomena in frequency comparisons are not always reliable for suggesting any pedagogical implications. Take Guo's study (2006), for example. He compares a learner corpus of Chinese students' essays with the LOCNESS corpus and reports that although the pattern *KEEP + noun phrase* occurs with similar frequency in the two corpora, very few nouns and noun phrases in the pattern are shared by the two groups of writers (Guo 2006: 183). He also points out that the overused items in the learner corpus could be the result of the different writing topics across corpora; for example, Chinese students use the key words, *keep fit*, which appear in the essay rubric, thereby increasing the total counts of this phrase (Guo 2006: 199).

A recent study by Granger and Paquot (2009) discusses the issue of comparability. They compare ICLE with a reference corpus, consisting of NS experts' writing from the MicroConcord corpus (Johns and Scott 1993) and the Baby BNC corpus and also with NS novice writers' essays in LOCNESS; according to their report, some differences in frequency comparison are due to the nature of the type of text. The ICLE corpus consists of learners' argumentative essays, while the reference corpus includes expository prose. They also

identify some problems of underuse and overuse shared by both NS novice writers and NNS learners; inevitably the latter have other problems also. The researchers conclude that the comparability of the corpora under investigation influence the results and conclusions drawn from them.

The use of the terms *underuse* and *overuse* seems to be more appropriate in the comparison between NNSs with different L1s. If comparing the sub-corpora in ICLE with a reference corpus, it makes more sense to discuss underuse and overuse phenomena across learner corpora of different L1s, as it aims to compare very similar data produced by different groups of learners to the standards set in the reference corpus. For the studies that are concerned with comparing a learner corpus and an NS corpus, I would suggest as alternatives the neutral terms, *under-* and *over-representation*, to discuss differences in frequency across corpora. The aim underlying *under-* and *over-representation* is to keep frequency information as linguistic evidence in focus and avoid over-generalising differences to learners' performance before more careful examination is undertaken and more supporting evidence is found.

The present study does not primarily aim to compare frequency information about DMs in NSs' and NNSs' spoken English. The frequency information is taken as an entry point into the data, as is done in most corpus studies. DMs are investigated with the notion of *appropriateness* in relation to context, not the notion of *correctness* and *accuracy*. This is why it is not appropriate, as in other learner corpus research, to discuss the underuse and overuse of DMs in the NNS corpus under investigation. The other reason that *under-* and *over-representation* are suitable terms for the present study is related to the issue of comparability. The corpora under investigation are of different kinds, chosen for comparison across the monologic and dialogic genres and for the investigation of the uses of DMs by the two groups of speakers, not for frequency comparison alone. Two broad types of genre, monologic and dialogic, are selected to test the hypothesis that there are more instances of DMs in the dialogic genres than in the monologic genres.

2.2.2 Diachronic study and error analysis for second language acquisition purposes

Compared to synchronous studies, diachronic studies of the longitudinal development of learners' proficiency are relatively few, mainly because they require a long period of time for data collection. A small portion of Housen's study (2002) on inter-language grammar is

longitudinal. However, there is a growing interest in *quasi-longitudinal* studies, i.e. comparing learners of the same L1 at different proficiency levels. For example, Dagneaux, Denness and Granger's study (1998) investigates the written English produced by French-speaking learners of intermediate and advanced levels. The data are manually processed with error-editing software by an NS to identify errors and by an NNS expert user of English to annotate error tags. The error-tagged learner corpora can be compared using standard corpus investigation software for further examination, such as counting errors and error types and analysing errors in context. Research of this kind evaluates learners' proficiency at different levels and can be applied in the production of automatic grammar and style checkers on computer.

The traditional error analysis, pointed out by Dagneaux *et al.* (1998), has been criticised for using heterogeneous data and fuzzy categories. It cannot identify phenomena of absence and avoidance. It focuses on learners' incompetence and remains product-oriented, which cannot demonstrate the dynamic aspect of L2 learning. In spite of the criticism of its limitations, Granger (2002: 14) argues that learner corpus studies of different L1 backgrounds identify underused linguistic phenomena, which indicate avoidance; error analysis helps us to understand the development of learner language and provides valuable information about learners' proficiency. Granger also maintains that error analysis using learner corpora is different from traditional error analysis, in that learner corpora can be well designed and error categories can be well defined and also that co-texts and broader contexts can be analysed.

Another example of the corpus-based error analysis is Flowerdew's work (2006). He presents a taxonomy of error types in the use of *signalling nouns*, such as *attitude*, *assistance*, *difficulty*, etc. Based on a corpus of Cantonese learners' argumentative essays, he reports a significant correlation between the number of signalling nouns and the grades given to the essays and between the number of signalling noun errors and grades. In the essays with higher grades, there are more instances of signalling nouns, which improve the lexical coherence of a text.

From the perspective of phraseology, Osborne (2008) examines four common errors: omission of 3rd person -s, inappropriate positions of adverbs, pluralised adjectives and plural form of mass nouns in the Chambéry Corpus (French-speaking university students' essays) and the ICLE corpus. Looking at the phraseological patterns where errors appear, he identifies three types of phraseology effect and argues that the four types of error do not occur randomly but are triggered by the phraseology where they occur.

Error analysis with corpus analysis techniques helps researchers look at learner language on a large scale and enables them to report results statistically (e.g. Flowerdew (2006)) and to investigate errors which might pass in an individual text but are obvious when repeated (e.g. Osborne (2008)). The computer-assisted error analysis contributes to the teaching and learning contexts where grammatical correctness is made the target. The identification of errors suggests areas where learners need more instruction.

However, error analysis is less useful for the present study, mainly because of the nature of DMs. The use of DMs says nothing about grammatical correctness in relation to syntax and semantics. Therefore, in a sense, taking error analysis as a starting position to investigate DMs will presume the use of DMs made by the NNSs under investigation to be incorrect. In addition, DMs have not yet been recognised as a category and academic opinion has not agreed yet what linguistic items are DMs. The relevant issues about DMs in the NS speech are in dispute, thereby questioning the use of NS examples as the target variety for analysing *errors* in the NNS speech. The next section will discuss the various classifications and approaches to the defining DMs in the literature.

2.3 Towards the characterisation of discourse markers

The general consensus in the literature is that DMs are difficult to fit into traditional grammatical categories. However, little agreement has been reached on their terminology, definition and classification. This section begins with the justification for the use of the term *discourse marker*, followed by a review of the classifications of DMs; it then moves towards the characteristics of the DMs chosen for analysis in this study.

2.3.1 Terminology

In terms of terminology, DMs are also known by a variety of other names, such as “sentence connectives” (Halliday and Hasan 1976), “discourse particles” (Schourup 1985, Aijmer 2002), “utterance particles” (Luke 1987, 1990), “semantic conjuncts” (Quirk, Greenbaum, Leech and Svartvik 1985), “pragmatic expressions” (Erman 1987), “discourse operators” (Redeker 1991), “continuatives” (Romero Trillo 1997), “discourse connectives” (Blakemore 1987, 1992) and “discourse markers” (Fraser 1990, 1999, Carter and McCarthy 2006). This last term is most widely used. Schourup (1999), in his later work, adopts the term *discourse marker*, as does Blakemore (2002).

The above terms are based on different theories and varying assumptions, which are not discussed here due to lack of space. In this study, I adopt the more popular and theoretically neutral term *discourse marker* (DM).

2.3.2 Characteristics and definitions of discourse markers

The classification of DMs and approaches to defining them are under debate. There is general agreement about some words and phrases, such as *well* and *you know*. They are classified as central DMs. For these DMs, it is not difficult to give characteristics and definitions. It is less certain whether some words and phrases, such as *oh*, *right* and *but*, are in the same category. For the time being, it is almost impossible to come to a dividing line between DMs or non-DMs for the words and phrases discussed in the literature and to give defining criteria. This section discusses the features which tend to belong to DMs rather than suggesting the criteria for deciding DMs.

The characteristics of DMs are based on work by Schourup (1999) and Fung and Carter (2007). Schourup (1999) discusses seven characteristics: 1) connectivity, 2) optionality, 3) non-truth-conditionality, 4) weak clause association, 5) initiality, 6) orality and 7) multi-categoriality. He claims that the first three are frequently taken as central characteristics of DMs. Fung and Carter (2007) list five criteria: 1) position, 2) prosody, 3) multi-grammaticality, 4) indexicality and 5) optionality. They are more like the characteristics in Schourup's work than criteria that could be used reliably to judge DMs.

This section discusses five characteristics: 1) optionality, 2) flexibility of position, 3) prosodic independence, 4) connectivity and 5) multi-grammaticality. In three common definitions of DMs, these are often considered prominent features: grammatical-pragmatic definition, coherence-based functional definition and relevance-based definition. Table 2.1 below summarises the literature discussed in this section.

The first characteristic, optionality, is generally accepted. DMs are syntactically and semantically optional. In other words, the presence or absence of DMs does not affect syntactical structures in the semantic relationships in discourse. This raises a question of what we mean by *optionality*, because words classed as adverb and adjective are also syntactically optional. For example, no word in Example (2.3.1) below can be omitted and still leave the utterance syntactically correct. In Example (2.3.2), *better* can be deleted and the clause is still syntactically correct but it loses some semantic meaning. In Example (2.3.3), *you know* can be

omitted without affecting the utterance's syntactical or semantic correctness.

- (2.3.1) I like this term (MICASE: COL605MX039)
- (2.3.2) I like this term better (MICASE: COL605MX039)
- (2.3.3) **you know** I like this term better

In spite of being syntactically and semantically optional, however, DMs are not taken as useless or redundant elements, but ones which facilitate the process of interpretation and interaction (Brown and Yule 1983, Fraser 1990, Carter and McCarthy 2006, Fung and Carter 2007). As pointed out in Section 1.3 in the introductory chapter, DMs are not merely decorations. They perform certain functions in speech. For example, *well* prefacing a repair would give the listener a signal in the process of making sense of the utterance.

The second characteristic is the flexibility of their position in an utterance. DMs appear at any point, utterance-initial, -medial or -final, depending on their preferences and functions (Fung and Carter 2007). Another radical argument is that DMs are commonly used in utterance-initial position (Schiffrin 1987: 328). However, this is not the feature of such central DMs as *well* and *you know*, which are found more often to occur utterance-medially.

DMs are also prosodically independent (Fung and Carter 2007). A DM “has to have a range of prosodic contours e.g. tonic stress and followed by a pause, phonological reduction”, is suggested by Schiffrin (1987: 328). This criterion is not true for every DM. For example, DMs *and* and *but* are often not prosodically independent. It can be argued that either *and* and *but* are peripheral DMs which do not share this characteristic, or that they are not DMs.

The fourth characteristic is connectivity. Fraser (1990: 383) defines DMs as signals of “a sequential relationship between the current basic message and the previous discourse”. While the connectivity of this kind refers to the link between two adjacent textual units, Blakemore's (1987) *discourse connectives*, based on relevance theory (Sperber and Wilson 1986) and within the scope of cognitive connectivity, go beyond linguistic resources to the background or contextual assumptions.

- (2.3.4) [On seeing someone carrying lots of parcels]
So you've spent all your money. (Blakemore 1987: 86, 106)

In Example (2.3.4) above, the speaker uses *so* to connect his or her utterance to the context. Carter and McCarthy (2006: 218) also emphasise the linking function of DMs in the organisation of discourse. However, this is not one of the criteria listed in Fung and Carter's

work (2007), as they find a number of DMs, such as *right*, *OK* and *now*, are used to open a topic, which does not connect to any previous discourse or context. Despite the discontinuity signalled by these DMs (e.g. a teacher beginning a lecture with *OK* or *now*), they seem to suggest a different kind of relationship of connectivity, which marks the disconnection from the previous utterance (e.g. students' chats) and the context (e.g. a break between lectures) which are relevant to the participants (e.g. the teacher and students). Moreover, this characteristic would probably cause confusion in the distinction between *but* and *and* as coordinating conjunctions and *but* and *and* as DMs. For example, *and* in Example (2.3.5) is unarguably a conjunction. *And* in Example (2.3.6) can be a DM to link two parts of a narrative. In Example (2.3.7), it is problematic whether *and* should be categorised as a conjunction or a DM.

(2.3.5) fish **and** chips

(2.3.6) when i was logging off like, **and** suddenly like all the lights went out (MICASE: SGR565SU144)

(2.3.7) she just applied for a better job **and** got it instantly (MICASE: MTG400MX008)

The last characteristic, multi-grammaticality, is discussed by both Schourup (1999) and Fung and Carter (2007). The inclusion of different grammatical classes is still open to debate. DMs are neither a word class nor a grammatical unit. They can include items of grammatical units of different kinds. The range of DMs in Carter and McCarthy's work (2006) is probably one of the widest, including adverbials, phrases, clauses and interjections.

From the discussion above, it can be seen that the five features are not all necessary for distinguishing DMs. Central DMs would have more tendency to exhibit the five characteristics above. Peripheral DMs may be difficult to identify using these characteristics. Overall, it is less difficult to characterise known DMs than to develop criteria for determining DMs.

The above characteristics are prominent but not all apply to the three common approaches to defining DMs, coherence-based functional definition, grammatical-pragmatic definition and relevance-based definition, discussed below and summarised in Table 2.1.

Based on the concept of coherence, Schiffrin's work (1987) has captured considerable attention. She defines DMs as "**sequentially dependent** elements which bracket units of talk" (1987: 31). The "units of talk", which she identified, are deliberately broad and she then analyses the DMs in her interview conversations. She argues that "markers are devices that work on a discourse level; they are not dependent on the smaller units of talk of which

discourse is composed” (Schiffrin 1987: 37) and concludes her study by defining DMs theoretically as “members of a **functional** class of verbal (and non-verbal) devices which provide contextual coordinates for ongoing talk” (Schiffrin 1987: 41). DMs function in a framework which embraces five “planes of talk”: *exchange structure*, *action structure*, *ideational structure*, *participation framework* and *information state* (Schiffrin 1987: 24-25).

Redeker (1991: 1139) offers the criticism that some of Schiffrin’s instances of *oh*, *y’know* and *I mean* do not function as DMs and that her exclusion of uses of such adverbials as *now* and *then* is not appropriate. Redeker, on the basis of the coherence function of DMs, prefers the term *discourse operator* and defines it as “a word or phrase – for instance, a conjunction, adverbial, comment clause, interjection – that is uttered with the primary function of bringing to the listener’s attention a particular kind of linkage of the upcoming utterance with the immediate discourse context” (1991: 1168).

Lenk (1998) comments on the space constraint of Schiffrin’s definition of DMs (1987), which builds a relationship between two adjacent segments. Lenk argues that DMs bring coherence on a more “global” level within the discourse; for example, linking to earlier topics, topics to be followed, or knowledge outside the context of the conversation. She points out that coherence in conversations depends not only on what Schiffrin refers to as relationships between two immediately adjacent items but also on global relationships linked by DMs for facilitating the interlocutors’ process of establishing coherence.

From the grammatical-pragmatic perspective, DMs have been taken as a pragmatic class (Fraser 1990, 1996). Fraser (1990) notes that a sentence contains propositional meaning and pragmatic meaning, which may be signalled by pragmatic markers to express the speaker’s communicative intentions. Pragmatic markers can include expressions in other syntactical categories, which presuppose central DMs, if they meet the condition of carrying pragmatic meaning. DMs are categorised as *commentary pragmatic markers*, which signal “a sequential relationship between the current basic message and the previous discourse” (Fraser 1990: 383). They have five properties: pragmatic function, non-truth-conditionality, flexible positions (but typically in utterance-initial position), optionality and a clear distinction between other *commentary pragmatic markers* (e.g. *certainly*, *frankly* and *according to her*), which create a certain degree of varied meaning and interjections, which are claimed to be a separate message and do not signal a relationship in discourse.

Fraser’s definition is relatively broad. However, it surprisingly excludes two

generally-accepted DMs, *oh* and *well*. The former is categorised as a basic marker and the latter a focusing marker. Neither of them signals a relationship in discourse; therefore, they are not DMs in Fraser's study (1999: 942).

Carter and McCarthy (2006) see DMs as a sub-class of pragmatic markers, which also include stance markers, hedges and interjections. With a broader perspective than Fraser's (1990), the DMs, seen as a lexical category, can be of any grammatical form used to "link segments of the discourse to one another in ways which reflect choices of monitoring organization and management exercised by the speaker" (Carter and McCarthy 2006: 208) and to "indicate degrees of formality and people's feelings towards the interaction" (Carter and McCarthy 2006: 212).

Apart from grammatical-pragmatic and coherence-based definitions, the relevance-theoretical definition tries to offer a unified framework for accounts of the use of DMs. DMs are defined within relevance theory and termed *discourse connectives*, i.e. "expressions that constrain the interpretation of the utterances that contain them by virtue of the inferential connections they express" (Blakemore 1987: 105). In other words, in understanding an utterance, a person draws inferences based on the relevance of the person's assumptions and the contexts and this process is constrained by DMs. For instance, in Example (2.3.8) below (the same as Example (2.3.4) above, which is repeated here for convenience's sake), the utterance *so you've spent all your money* is based on what the speaker has seen. The listener's comprehension process is constrained by the DM *so*, which seems to suggest the speaker is drawing a conclusion after seeing the listener's parcels.

(2.3.8) [On seeing someone carrying lots of parcels]
So you've spent all your money. (Blakemore 1987: 86, 106)

Two arguments for the use of DMs appear in the framework of relevance theory: 1) An utterance connects the listener's background knowledge and contextual assumptions and 2) DMs are non-truth-conditional, making no contribution to the proposition carried in an utterance but functioning pragmatically (Blakemore 1987, Andersen 1998: 147). The relevance-theoretical definition of DMs instead functions explanatorily, seeking to explain why DMs are used. The main disadvantage of relevance theory is that contrived examples are taken for analysis, or works with only few examples are taken into account. This explanatory way of investigating DMs is very different from corpus analysis, which is descriptive and

based on a great deal amount of linguistic evidence.

Table 2.1 below summarises the above discussion of the three approaches in relation to the five characteristics of DMs. The table shows that the coherence-based definition and the grammatical-pragmatic definition resemble one another most closely, in the sense that they broadly agree on the characteristics of DMs. The relevance definition has a different, theory-based, starting point and is more distinct. All the approaches discuss DMs comprising a variety of grammatical units.

Table 2.1: Characteristics of discourse marker in the three approaches

	Coherence-based functional definition	Grammatical-pragmatic definition		Relevance-theoretical definition
Researcher(s)	Schiffirin 1987 Redeker 1990, 1991 Lenk 1998	Fraser 1990, 1996, 1999	Carter & McCarthy 2006	Blakemore 1987, 1992 Andersen 1998
Characteristics				
1) Optionality	Yes	Yes	Yes	n/a
2) Flexibility of position	Yes	Yes (typically in utterance-initial position)	Yes	n/a
3) Prosodic independence	Yes	n/a	Yes	n/a
4) Connectivity	Yes (connect two 'units of talk', Schiffirin) (connect utterance and context, Redeker) (connect discoursal segments, Lenk)	Yes (connect two messages)	Yes (connect discoursal segments)	Yes (not necessarily connecting two textual segments; may be background or contextual assumptions)
5) Multi-grammaticality	Yes	Yes	Yes	Yes

Although there is a consensus in the literature that DMs constitute a class of items performing particular functions, there has been no successful attempt to produce a definitive list of DMs. This is a very difficult and complex issue, relating to unclear definitions. Most of the studies in the above three approaches to defining DMs discuss the characteristics of DMs rather than giving defining criteria and therefore it has not been clearly pointed out which words and phrases are DMs and which are not. Some of the studies include words and phrases (e.g. *however* and *so* in Fung and Carter's work (2007)) which are labelled differently in other models (e.g. *however* and *so* are conjunctions in Halliday and Hasan (1976)).

Another problem in the investigation of DMs is the identification of the functions of DMs. As argued earlier in this thesis, the functions of DMs are not easily identified by introspection. This feature turns out to be highly problematic in the discussion of functions of DMs in previous studies. The lack of clarity and objective description of functions can result

in varying interpretations of the use of a particular DM.

2.3.3 Discourse markers in this study

There has been no consensus over defining DMs and which linguistic items are DMs. Researchers who intend to investigate DMs usually find themselves facing different definitions and a huge range of DMs. The first step used in this study to narrow the range of DMs is to use the textual and interactive distinction which Sinclair and Mauranen (2006) make in *Linear Unit Grammar* (see Section 2.5, where it is dealt with in some detail). DMs are among interactive organisational (OI) elements. OI elements are claimed by Sinclair and Mauranen (2006: 73) to be second-level ordering devices, which look outward to the context and interactional aspects of the text. Sinclair and Mauranen (2006: 78) also point out that there are more OI elements in conversations than formal speech and OI elements are less important and almost absent in written English. This view is consistent with the characteristic of *orality* in Schourup's article (1999: 234).

As noted above, it is difficult to reach any criteria for determining DMs. I would here give five characteristics of the eight DMs analysed in this study, *like, oh, well, you know, I mean, you see, I think* and *now*: 1) semantic and syntactical optionality, 2) flexibility of position, 3) frequent prosodic independence, 4) connectivity and 5) multi-grammaticality.

2.4 Previous studies of discourse markers

Questionable answers and answerable questions, Lakoff's paper (1973), is according to Müller (2005: 101) probably the beginning of the investigation of DMs. Lakoff points out that research moved from syntactical correctness towards appropriateness in relation to contexts. She then discusses the appropriateness of using *well* and *why* in turn-initial position to preface an answer to a question in different conversational conditions. Since then, a considerable amount of research has been done on the DMs used by NSs, such as Östman (1981), Schourup (1985), Holmes (1986), Schiffrin (1987), Erman (1987), Fraser (1990, 1999), Jucker (1993), Lenk (1995, 1998), Biber *et al.* (1999), Fox Tree and Schrock (2002), Aijmer (2002), Macaulay (2002), Müller (2005) and Carter and McCarthy (2006), *inter alia*. Detailed reviews of the literature on the words and phrases under investigation are given in the appropriate sections of later chapters.

Most studies of DMs, either theory-based or data-based studies, discuss DMs in terms of

their functions but it is seldom clear how the researchers determine what the functions of DMs are, as they could be through a process of either guessing or logic. In other words, it is difficult to discuss the functions of DMs without reading the speakers' minds. It is reasonable to argue that there is a certain amount of speculation attached to the functions of DMs in the literature. The functions discussed in the previous studies are usually derived from intuitively interpreting the context in which a DM occurs. The dilemma is that when presenting the uses of DMs, it is probably better to apply the term *functions* mostly to learners of English, who are learning what DMs are for. The approach employed in this study is, first, to use the collocation phenomena surrounding a DM to determine the co-occurrence categories, thereby providing a basis for logically progressing to the identification of functions, which are secondary interpretations. The research procedure in this thesis is unpacked in the chapter on methodology.

Most studies of DMs apply to NS speech. Relatively few researchers have considered comparative uses of the DMs by NSs and NNSs. Müller (2005), Wang and Zu (2005) and Fung and Carter (2007) are among the few exceptions.

Müller's research (2005), based on a 350,000-word corpus of spoken English by American NSs and German NNSs, provides a detailed analysis of the frequencies and functions of the four DMs, *so*, *well*, *you know* and *like*. Her study was well-designed for collecting comparable data from silent movie narratives and discussions. Rather than adopting an existing framework, Müller manually identified the functions of the four DMs and classified them at two levels, the textual and the interactional level. The functions at the textual level are not used to address the listeners but focus on lexical expressions (e.g. a speaker's search for words, restarting and repairing), the structure of propositional contents (e.g. explanations and exemplifications) and the distinction between the speaker's voice and reported speech⁴. The functions at the interactional level work for the relationship between speakers and listeners by marking a speech act, response, opinion, evaluation, appeal to the listener, etc. *So*, *you know* and *well* have been found to serve a number of functions at both the textual and interactional levels. *Like* functions only at the textual level⁵. It has been suggested that German students use the DMs *so*, *you know* and *like* less frequently than American

⁴ I prefer to use the term *reported speech*, which is often used in grammars, because the use of discourse markers is part of spoken English grammar.

⁵ In the text-based analysis (Chapter 10), *like* as a discourse marker is found in use as a means of expressing solidarity among young people. This use of *like* is at the interactional level.

students do and the DM *well* with similar frequency; some functions are used only by American students and some only by German students.

Unlike Müller's compilation of two comparable corpora, the investigation by Wang and Zhu (2005) concerns fifteen types of DM in the Spoken English Corpus of Chinese Learners (SECCL) (Wen, Wang and Liang 2005) and the spoken component of the British National Corpus (BNC). The SECCL corpus consists of Chinese-speaking NNSs' monologues and dialogues, while the BNC corpus includes NSs' informal conversations. Three main conclusions are drawn to mark differences between Chinese NNSs and NSs in the use of DMs. First, the NNSs and NSs use different types of DM; second, the NNSs underuse DMs in terms of frequency and type and third, the NNSs overuse some of the additive and emphatic DMs, such as *and*, *but* and *very* and fillers with semantic meaning, such as *I think*. These findings raise the question of comparability in learner corpus research and hence the usefulness of underuse and overuse phenomena. The researchers randomly extracted about 460,000 tokens from SECCL and from BNC. The frequency comparison ignored the constraints relating to genre. It is questionable whether the frequency informs us of the underuse and overuse of DMs. As noted in Section 2.2.1, using the terms *under-* and *over-representation* of a particular DM is probably more appropriate than the terms *underuse* and *overuse* in this case.

The above studies by Müller (2005) and Wang and Zhu (2005) differ in the respect of scope and research methods. Müller compiled two spoken corpora of NS and NNS and investigated four DMs. Her study is qualitative, in that it makes sense to discuss underuse and overuse phenomena in two comparable corpora and qualitative, in that the contexts of the four DMs are analysed. In contrast, Wang and Zhu's study is purely quantitative, comparing a wide range of DMs (76 DMs in 15 categories) in two corpora of similar size but different genre. This study provides an overview of the representation of the DMs in the two corpora, but does not further examine of the use of DMs. Overall, Wang and Zhu's study can give only a general picture of the frequency of DMs in the NS and Chinese NNS speech and Müller's study offers better understanding of the use of DMs. As DMs are found to be context-dependent, looking at the frequency information without analysing their (immediate) contexts is of less help in comparing NS speech with NNS.

The present study combines some elements of Müller's (2005) and Wang and Zhu's work (2005). It is both quantitative and qualitative. In the first stage, the DMs used by the Chinese NNSs are manually identified and the instances of the selected DMs for analysis in the six

sub-corpora under investigation are retrieved with corpus techniques. The results are used to discuss the representation of the DMs across corpora. The identification of collocation phenomena of DMs is qualitative, looking at the immediate contexts. It is neither a limited study of a few DMs, as Müller's is (2005), nor a purely quantitative study like Wang and Zhu's (2005).

Another of the few studies carried out on the use by Chinese NNSs of DMs is Fung and Carter's (2007). They compared the use of DMs by NSs and Chinese NNSs in pedagogic settings. Based on a 460,055-word sub-corpus of the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and a 14,157-word learner corpus of interactive classroom discourse, they examined a wide range of DMs in a framework of interpersonal, referential, structural and cognitive categories and found that the Chinese NNSs extensively use referential and structural DMs (*and, but, because, I think*) but have limited use of others that are frequently used by the NSs (*and, right, yeah, well, so, now, sort of, you know, actually, see, say and cos*). Fung and Carter abandon the terms *underuse* and *overuse* and discuss the representation of DMs in the two corpora. They use a simple mathematical subtraction to arrive at a comparative analysis. The problem of this study is that the researchers did not remove the instances of non-discourse use of the words and phrases, e.g. the instances of *you know* may include those in *do you know* and *as you know*. The present study distinguishes between non-discourse use and discourse use of the words and phrases under investigation. This provides more accurate counts of those words and phrases which are DMs and not in other grammatical categories.

The above three studies are examples which have contributed to the understanding of the use of DMs by Chinese NNSs and NSs. To some extent they reveal the underuse and overuse of DMs in NNS speech. However, this is not the main purpose of the present thesis. As noted earlier in Section 1.4 of Chapter 1, I adopt a different perspective on NNS speech, viewing it as a variety of English.

2.5 Linear Unit Grammar: an approach to the description of spoken English

Traditional grammars cannot describe spoken English satisfactorily. Even where a grammatical description is possible, the labels in the description (e.g. *fragment* and *incomplete clause*) often imply incorrectness. This may be why Brazil (1995) argues for a "purpose-driven" and "linear" grammar of speech (see discussion in Section 1.2 of Chapter 1).

Similarly, Sinclair and Mauranen's (2006) *Linear Unit Grammar (LUG)* is designed as a descriptive bottom-up approach to syntagmatic grammar. It is intended to be compatible with most conventional grammars. It completely abandons traditional word classes and syntactic labels.

The *LUG* analysis consists of four steps. The procedure is shown below with Sinclair and Mauranen's example (2006: 151), extracted from a conversation. First, the extract is chunked, separated into units. The vertical bars indicate the boundaries between units.

I wondered | what happens | when you go | from one island | to the other | no | the train goes | on the ferry |
oh | I see | yes

Each unit is then assigned one of two main types, a message-oriented element (M) or an organisational element (O) (see Appendix 4 for the labels used in *LUG*).

I wondered (M) | what happens (M) | when you go (M) | from one island (M) | to the other (M) | no (O) |
the train goes (M) | on the ferry (M) | oh (O) | I see (O) | yes (O)

Third, O elements are further divided into two sub-categories, interaction-oriented organisational (OI) elements and text-oriented organisational (OT) elements. M elements are divided into varied sub-types: message fragment (MF), incomplete message unit (M-), completion of message unit (+M), partial completion of message unit (+M-), supplement to message unit (MS), adjustment to message unit (MA) and revision to message unit (MR).

I wondered (M-) | what happens (+M-) | when you go (+M-) | from one island (+M-) | to the other (+M) |
no (OI) | the train goes (M-) | on the ferry (+M) | oh (OI) | I see (OI) | yes (OI)

The final step is to synthesise by following a set of procedures: 1) remove OI elements, 2) remove MF elements, 3) reconcile MA with the following +M, 4) reconcile M- to the following +M, 5) add MS with the preceding M, 6) merge MR with the M elements of which they are reformulations, 7) adjust text to take account of notes and 8) adjust text towards written norms. This last step results in two *Linear Units of Meaning (LUMs)*.

- 1) I wondered what happens when you go from one island to the other
- 2) the train goes on the ferry

The final product of the *LUG* analysis does not seem very exciting. The above two

LUMs can be taken a revision of the draft in spoken form. The main contribution of *LUG* is its ability to process authentic language, which is often viewed as un-grammatical and incorrect in traditional grammars.

The words and phrases under investigation in the present study can be classified into two major functional categories, an M element and an OI element. The latter functions as a DM, mainly contributing to aspects of the interaction, such as initiating, maintaining and structuring the interaction and controlling the timings. The former increases the shared knowledge of the interlocutors. As informed by Brazil's speech grammar (1995), the speaker and hearer in real-time communication process meaning incrementally.

In this thesis, *LUG* proceeds in two steps. First, the DMs for analysis are selected through manual examination. The two major categories in *LUG*, message-oriented elements and organisational elements, support my distinction between Type A word/phrase and Type B. Each word/phrase for analysis should be classifiable both as a DM used by NSs and as an OI element in an *LUG* analysis. Second, *LUG* is used to assign units in spoken English and its labels are used to describe where DMs occur in the intra-clausal positions in an utterance.

In principle, applying the *LUG* analysis is useful because it offers a clear distinction between message-oriented (M) elements and organisational (O) elements and this coincides with the distinction between Types A and B in this thesis. In practice, *LUG* provides another way of thinking about the problem of distinguishing phrases (e.g. *you know* and *I think*) between DMs and reporting clauses, but it must be admitted that this apparatus does not solve the problem; it merely provides a different way of looking at it. When I faced an ambiguous example, thinking about it from the viewpoint of *LUG* helped me to see the problem in a different way; however, I still had to decide whether the element was M or O. This was essentially the same judgement about whether a word or phrase is a DM or not. It is often the case that rephrasing the problem helps to solve it. Nonetheless, it has to be recognised that the *LUG* analysis is sometimes not mechanically developed enough to deal with the ambiguity. The usefulness and limitations of *LUG* are further discussed in Section 12.2.6 of Chapter 12.

The occurrences of the selected words/phrases have to be classified into an OI element and an M- element, which is a complicated process, and the distinction often relies on analysts' judgement on the contexts where they occur. In Example (2.5.1), Sinclair and Mauranen (2006: 73-75) illustrate the distinction between OI and OT and how the decision was reached that *I think* was in this case an OI, not an M-. To segment M elements, OT

elements are of first-level ordering, which look inwardly and contribute to the coherence, while OI elements, being second-level ordering, look outwardly to a larger stretch of discourse and circumstances. *But* is designated an OT because it sets up a contrasting relationship between the preceding M element (*in certain areas*) and the subsequent M element (*in service*). To designate *I think* as an OI element, the analysts take into consideration the previous utterance, *you can't even get a job officially*, and assume that the shared uncertainty continues; therefore, it is unlikely that the utterance after *I think* states personal opinions. In this case, *I think*, as an OI element, “controls timing and presentation, and it just extends and slightly emphasises the cushioning effect of *well*” (Sinclair and Mauranen 2006: 74).

(2.5.1)

well (OI) | **i think** (OI) | in certain areas (M) | you can (M) | but (OT) | for example (OT) | in service (M) |
you can't (M)

(Sinclair and Mauranen (2006: 75))

The above example shows that the identification of two-word DMs (e.g. *I think*) is more complex than one-word DMs (e.g. *well*). More examples of this classification are given in later chapters.

The other use of *LUG* analysis in this thesis is to assign units in spoken English and to describe the positions of DMs in an utterance/turn. Traditional syntactical structures may be used to describe spoken English, but it is very likely that spoken English does not follow the syntactical rules and DMs cannot be identified in syntactical structures. DMs can occur in any position in an utterance, making them difficult to describe. In Example (2.5.2) below, *like* is an OI element and therefore a DM. It is placed between an OT (*that*) element and a +M (*why most people*) element.

(2.5.2)

yeah | i asked my dad | that | **like** | why most people | but | he said | something that | um | the U-S was
looked at | as a better place | to go to | that | it was harder to get here,

(MICASE: OFC115SU060)

It is worth noting that the unit in the *LUG* analysis is a chunk, not a clause. For instance, in *Sounding nice* (M-)| *is no longer enough*, (+M)| *he argued* (M) (Sinclair and Mauranen 2006: 83). *Sounding nice is no longer enough* is analysed as two elements. This leads in *LUG*

to some analyses which may seem peculiar, in particular when a DM occurs between M- and +M elements. In cases of this kind, a DM seems to separate an M element; however, it is noted in subsequent chapters that some DMs tend to occur between M- and +M elements rather than the other way around.

Possible applications of *LUG* in the areas of Applied Linguistics, according to Sinclair and Mauranen (2006), are foreign language teaching and translation studies. In language teaching, the model of *LUG* helps to bridge the gap between the naturally-occurring language which learners encounter outside class and well-formed language in the hierarchical model of pedagogical grammars. The chunking activity handles lexis and structure together and improves learners' ability to process on-going speech. The authors suggest that explicit instruction on chunking short extracts of unscripted speech with the *LUG* approach can make language learning more effective better than expecting learners to acquire language through exposure to authentic data, which is barely feasible in an environment where English is used as a foreign language.

The authors point out that the distinction between O and M elements is important for learners. When they are trying to understand the proposition in an utterance, they should be able to focus on M elements. For making sense of the connection in discourse, they should be able to make use of OT elements. These strategies are helpful in contexts where facts are required. However, the authors argue that OI elements help interpret speakers' attitudes, feelings, degree of commitment, certainty and reservations. This is especially helpful for the teaching of DMs, a point I shall return to in the chapter on pedagogical implications and applications (Chapter 12).

In addition to the applications in language teaching and learning, *LUG* can be applied in the training of interpreters. The distinction between M and O elements and the separation of OT and OI elements can be an important skill to facilitate translating and interpreting interpersonal meanings (Sinclair and Mauranen 2006).

2.6 Chapter summary

In this chapter I survey the literature on corpus data and linguistic items for analysis, analytical approaches and methods. Corpora in use and common corpus techniques are introduced, as related to the quantitative method used in this study. I also discuss two main interests in the studies of learner corpora and the characteristics of DMs and the approaches to

defining DMs. In addition, previous studies of DMs and the *LUG* analysis (Sinclair and Mauranen 2006) are introduced. This chapter provides relevant backgrounds. I now turn to a detailed description of methodology in Chapter 3.

CHAPTER 3: METHODOLOGY OF THE STUDY

3.1 Corpora under investigation

This research uses three publicly available corpora of spoken English: The Spoken and Written English Corpus of Chinese Learners (SWECCL) available via the Foreign Language Teaching and Research Press Beijing, China (Wen *et al.* 2005), the Michigan Corpus of Academic Spoken English (MICASE), developed by the University of Michigan and the International Corpus of English – The British Component (ICE-GB) (2006), published by University College London. Each of the three corpora had two subsets extracted and these six sub-corpora were processed using a standard corpus investigation software, *WordSmith 4* (Scott 2004), to search for the relevant items and scrutinise their co-texts. Table 3.1 below shows the number of texts, word counts and average words per text in the six sub-corpora under investigation.

Table 3.1: Corpora under investigation

Corpus	Number of texts	Word counts (tokens)*	Average words per text (tokens)
SECCL: Monologues (Chinese NNSs)	1,143	336,303	294
SECCL: Dialogues (Chinese NNSs)	1,143	596,639	522
MICASE: Highly monologic discourse mode (American NSs)	13	134,096	10,315
MICASE: Highly interactive discourse mode (American NSs)	48	577,996	12,042
ICE-GB: Unscripted monologues (British NSs)	70	153,646	2,195
ICE-GB: Private direct conversations (British NSs)	90	185,000	2,056

* All texts are processed in *WordSmith 4* (Scott 2004).

The NNS data used in this study comprised the monologue (Task B) and dialogue (Task C) sections in the Spoken English Corpus of Chinese Learners (SECCL), which was the spoken component of the SWECCL (Wen *et al.* 2005). The SECCL corpus was compiled from the Test for English Majors Band 4 (TEM 4) in China. This test of spoken English was taken by second-year English majors between 1996 and 2002; its three tasks comprised retelling a story, talking on a given topic and role-playing (see Appendix 1 for the details of the tasks in TEM 4). There were 1,460,042 tokens in total and 327,199 tokens, 378,862 tokens and 753,981 tokens in Tasks A to C respectively (Wen *et al.* 2005). (Wen *et al.*'s word counts were generated by *WordSmith 3* (Scott 1998). This study uses *WordSmith 4* (Scott 2004) which issued word counts of 337,926 tokens, 336,303 tokens and 596,639 tokens

respectively.)

The present study used the monologues in Task B, in which each speaker had three minutes of preparation time and was asked to talk for three minutes on a given topic and the dialogues in Task C, in which two speakers talked with each other for four minutes on the basis of a given prompt. The NNSs' speech under investigation was elicited in a rather restricted test-taking context. However, since people in the environment of English as a foreign language do not use English elsewhere, it is very difficult to collect naturally-occurring speech produced by NNSs. The chosen corpus was one of a few existing corpora of spoken English by Chinese people at the time of this research.

It was difficult to obtain a comparable corpus of NS speech. It seemed appropriate to recruit NSs to do the same tasks as NNSs had done in the oral tests. Nevertheless, in these circumstances, the speech of NSs would be different from naturally-occurring speech, because NSs are not trained to take an oral exam in their L1 and have never practised for one. Moreover, some topics for discussion by the Chinese NNSs may be inappropriate for NSs; for example, giving advice to freshmen at university and the pros and cons of going abroad for college education (for topics discussed by the Chinese NNSs, see Appendix 1). It was also difficult to compile a corpus of the NS speech which would resemble in size and number of participants (more than 1000) the NNS corpus used in this study. The chosen compromise was to compare the uses of DMs in the NNSs' spoken English in SECCL with those in the NS speech in such publicly available corpora as MICASE and ICE-GB. An advantage of using existing corpora is that it allows a better spread of data and larger corpora for robust descriptions of DMs than building my own NS corpora could provide.

Two NS corpora are selected in order to obtain a wider range of types of activity and context. The speech data in the MICASE corpus are produced solely in academic settings on campus, while those in the ICE-GB corpus are in various business-related contexts and casual conversations. (See Appendices 2 and 3 for the types of activity in MICASE and ICE-GB.)

The SECCL corpus consists of elicited speech, which has its limitations. In the corpus studies where immediate contexts were examined, the un-naturalness of the role-play activity in the NNSs' dialogues for analysis sometimes made the NNSs' use of DMs sound odd, because their discourse in getting information from each other was idiosyncratic. The nature of these NNS data was taken into account whenever possible. (See Section 13.3 of Chapter 13 for an acknowledgement of the limitations of using elicited data.)

The NS data were extracted from the MICASE corpus and the ICE-GB corpus. The MICASE corpus used in this thesis consists of two sub-corpora: 13 transcripts of highly monologic discourse mode and 48 transcripts of highly interactive discourse mode (see Appendix 2 for the fact sheet of the two sub-corpora extracted from MICASE).

The online search engine of MICASE provided such speaker attributes as gender, age, academic position/role, NS status and first language and such transcript attributes as speech event type, academic division, academic discipline, participant level and interactivity rating. Setting the NS status at *native speaker, American English* and the transcript attributes at *highly monologic* and *highly interactive*, elicited 14 and 48 transcripts respectively (retrieved on 18 February 2009). These transcripts were downloaded and processed with *WordSmith 4* (Scott 2004). One of the highly monologic texts, LES495JU063, was on closer inspection found to be a lecture produced by a senior graduate with near-native proficiency. More than half of the occurrences of *you know* (225 out of 388) and *I mean* (64 out of 130) came from this text. The speaker's overly familiar use of *you know* and *I mean* could be non-native-like. (An extract from this text is discussed in Section 7.3.1.) Therefore, this text was taken out in order not to skew the results from the NS data. The remaining 13 texts were used for analysis.

The second NS corpus consisted of two subsets of the ICE-GB corpus, the 70 unscripted monologues and 90 private direct conversations (see Appendix 3 for the fact sheet of the two sub-corpora extracted from ICE-GB). ICE-GB is a fully tagged and parsed corpus and is accompanied by a text analysis program, *ICECUP*. The two selected sections of texts were processed by the text processors *SED* and *AWK* to clear all kinds of annotations except the identification of speakers and markups of short and long pauses.

The problem with using the three different corpora was that they were arranged differently. In the SECCL corpus, each monologue was spoken by one speaker. The dialogues were produced by two speakers who were identified in the transcripts. Speaker turns (identified by *a* and *b*, which can be searched) were represented as continuous paragraphs. In the MICASE corpus, the speaker information, speech segments which overlapped and pauses were shown in the markup language. In the ICE-GB corpus, an utterance was separated by pauses and a grammatical unit was isolated for grammatical annotation. Each utterance was usually a finite main clause. For example, Excerpt (3.1.1) was produced by a single speaker, whose utterance was divided into segments.

(3.1.1)

A: And it's an example of how something totally unexpected may result from media coverage because <, > we've helped by this without our realising it people all over England <, > and to preserve and enhance their areas or their houses or their <, > little villages or whatever

A: We've had a huge mail bag which we actually can't hope to deal with <, >

A: We ourselves are having problems again at the moment because something else has threatened to be built next us

A: But whatever happens the principle that was established by this court case has made a huge difference

(ICE-GB: S2A-027)

My thesis would need to be consistent in identifying units in order to describe the positions of DMs in an utterance. Thus the above example in the ICE-GB was taken as a single utterance, even though it was set out in segments.

3.2 Selection of discourse markers for analysis

The data were not analysed within any framework, in part because there was no agreed taxonomy of DMs available and in part because existing frameworks might not be appropriate for the data used in this study. To avoid limiting the point of entry into the data to a list of DMs, the proposed DMs for analysis were selected on the basis of the NNS data. Of 1,143 texts in the sub-corpus of monologues of the SECCL, 114 (10%) of the texts, in the similar proportion for each year of the corpus, were read. It was found that twelve words and phrases had the characteristics of DMs and belonged to the category of OI elements in *LUG* (Sinclair and Mauranen 2006). They are listed in Table 3.2 in descending order of the frequency in the monologues.

Table 3.2: Discourse markers identified in the Chinese non-native speakers' speech

	Discourse markers	Raw frequency*	
		SECCL: Monologues	SECCL: Dialogues
1	<i>and</i>	11,612	12,515
2	<i>so</i>	3,544	5,645
3	<i>but</i>	2,806	7,681
4	<i>I think</i>	1,019	9,465
5	<i>well</i>	512	1,384
6	<i>now</i>	475	1,163
7	<i>oh</i>	397	2,882
8	<i>you know</i>	303	3,263
9	<i>yes (yeah)</i>	118 (30)	5,683 (3,804)
10	<i>right</i>	106	773
11	<i>I mean</i>	56	319
12	<i>you see</i>	39	452

*including both non-discourse use and discourse use of the word/phrase

The limited space in this thesis dictated that only seven of these twelve DMs – *I think*,

well, now, oh, you know, I mean and *you see* – could be selected for analysis. The remaining five DMs were not chosen because their highly frequent non-discourse uses were highly likely; for instance, *and* as a conjunction and *so* as an adverb would be unsurprisingly frequent.

In addition to the seven words and phrases, the special case *like* was added for analysis. *Like* as a DM was not found in the manual search in the Chinese NNSs' monologues. However, it had become a common DM and studies of *like* in NS speech have emerged lately, for example, Dailey-O'Cain (2000), Cukor-Avila (2002), Siegel (2002), Fuller (2003b), Fox Tree (2006, 2007), Fox Tree and Tomlinson Jr. (2008), Barbieri (2009) and Jones and Schieffelin (2009). As noted above (Section 2.2.2 of Chapter 2), corpus studies cannot identify phenomena of absence and avoidance. The hypothesis formulated at this point is that the Chinese NNSs do not use *like* as a DM, as this use is fairly recent. It was therefore interesting to investigate how the NNSs and NSs use *like* in speech.

3.3 Quantitative corpus analysis

The present study started at the bottom with a linguistic description of the words and phrases under investigation from the standpoints of grammar and discourse. The frequency information was used as a point of entry into the data. The frequencies of the eight words and phrases showed their distribution across the six sub-corpora. Manual classification of the non-discourse use (Type A) and discourse use (Type B) of the words/phrases indicated whether they were primarily used as DMs in the speech of the Chinese NNSs and NSs. The frequencies of Type B words/phrases were compared to test the hypothesis that DMs occurred more frequently in the dialogic genres than in the monologic genres. The collocates of the words/phrases revealed an overall picture of their use. The major part of the analysis was the new process, which used collocation phenomena and co-text analyses to empirically derive the functions of DMs rather than interpreting them intuitively. The following subsections include a detailed step-by-step discussion of the approaches used in the analysis chapters (Chapters 4 to 9).

3.3.1 Frequency information

Frequency information on the words and phrases for analysis came from standard corpus investigation software, *WordSmith 4* (Scott 2004). It was used to search for the relevant items and then their co-texts were scrutinised. The *Concord Tool* in *WordSmith 4* (Scott 2004) was

used to generate frequency information and collocates of the words and phrases. Table 3.3 below lists the raw frequencies and normalised frequencies (in brackets) of the eight words and phrases for analysis across the six sub-corpora under investigation.

Table 3.3: Raw and normalised frequencies of the words and phrases for analysis in the six sub-corpora

Corpus/word and phrase	Raw frequency* (Normalised frequency per 10,000 words)							
	<i>like</i>	<i>oh</i>	<i>well</i>	<i>you know</i>	<i>I mean</i>	<i>you see</i>	<i>I think</i>	<i>now</i>
SECCL: 1,143 monologues (Chinese NNSs)	736 (22)	397 (12)	512 (15)	303 (9)	56 (2)	39 (1)	1,019 (30)	475 (14)
SECCL: 1,143 dialogues (Chinese NNSs)	984 (17)	2,882 (48)	1,384 (23)	3,263 (55)	319 (5)	452 (8)	9,465 (159)	1,163 (20)
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	284 (21)	24 (2)	200 (15)	163 (12)	66 (5)	58 (4)	89 (7)	367 (27)
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	8,513 (147)	2,795 (48)	2,116 (37)	2,671 (46)	1,702 (29)	228 (4)	1,841 (32)	1,175 (20)
ICE-GB: 70 unscripted monologues (British NSs)	235 (15)	45 (3)	353 (23)	79 (5)	21 (1)	35 (2)	170 (11)	620 (40)
ICE-GB: 90 private direct conversations (British NSs)	913 (49)	1,123 (61)	1,521 (82)	819 (44)	865 (47)	144 (8)	662 (36)	372 (20)

*including both non-discourse use and discourse use of the word/phrase

3.3.2 Distinguishing between non-discourse use (Type A) and discourse use (Type B) of the words and phrases for analysis

Apart from *oh*, the words and phrases for analysis, *like*, *well*, *you know*, *I mean*, *you see*, *I think* and *now*, have two very distinct uses, non-discourse use (Type A) and discourse use (Type B). The instances of these words and phrases were manually classified into Type A and Type B. *Oh* has no content meaning and can serve only in discourse. It is further discussed in Chapter 5.

The words *like*, *well* and *now* can be of different word classes. The phrases *you know*, *I mean*, *you see* and *I think* are consequences of the two decisions, made from the paradigm of *pronouns* (e.g. *you*, *I*, *he*, *she*, *we*, etc.) + *know(s)/mean(s)/ see(s)/ think(s)*. Each of these is independently selected and grammatically constructed. This is quite different from Type B *you know*, *I mean*, *you see* and *I think*, which are chosen by the “idiom principle” (Sinclair 1991). Type B *you know*, *I mean*, *I think* and *you see* do not vary as phrases, no independent choice or paradigmatic selection. There is no negative form or past tense for Type B phrases,

which occur as fixed phrases.

To distinguish Type A *like*, *well* and *now* from Type B was mostly a straightforward reference to the word classes and syntactical structures where they occurred. For instance, *like* in Excerpt (3.3.1) was obviously a verb, preceded by the auxiliary verb *did* and followed by the gerund *being*. In contrast, *like* in Excerpt (3.3.2) did not belong to any word classes and fitted into no syntactical structure, either. Instances such as *like* in the former were classified into Type A and those such as *like* in the latter into Type B.

(3.3.1)

.....he didn't **like** being| on the deck of a ship| during the sea battles|.....

(MICASE: LEL215SU150)

(3.3.2)

.....if there's an insertion or deletion it's going to be shorter or longer you can detect it when when you do P-C-R, you're gonna get **|like|** replications of uh the chromosomes

(MICASE: SGR175MU126)

The *LUG* analysis offered a taxonomy in which Type B words/phrases could be classified, although the use of *LUG* did not make this classification easier. *Like* in the first case was part of the M- element *he didn't like being*, while the one in the second case was a single OI element.

Similar to the words *like*, *well* and *now*, some of the instances of *you know*, *I mean*, *you see* and *I think* were easily classified with reference to the co-occurring syntactical structure, as exemplified in Excerpts (3.3.3) and (3.3.4). *I mean* in the first example was grouped in Type A, because it was not a fixed phrase and it could be replaced by other pronouns (e.g. *he*, *she*, *we*, etc.) + conjugations of *MEAN*. *You know* in the second example was of Type A, because it followed a *that*-clause and in the Subject-Verb-Object structure.

(3.3.3)

.....so, let's take a concrete example. |you'll see what **i mean**|.

(MICASE: LEL485JU097)

(3.3.4)

.....so I think this is a good teaching... eh... .. method, because **|you know|** that| language is not just a tool,

(SECCL: B01-123-21)

The positions in the utterance/turn of the phrases were among the criteria for making the distinction. When the instances of the phrases occurred in extra-clausal utterance/turn-medial

and -final positions and in intra-clausal positions, these were classified as Type B. The instances of the phrases in extra-clausal utterance/turn-initial position created great difficulty due to their syntactical ambiguity. Take Excerpt (3.3.4) above, for example. The omission of *that* would make the classification difficult. In such cases, three things – punctuation in the transcripts, Biber *et al.*'s (1999: 1076-1078) three criteria for determining “utterance launchers” as DMs and the *LUG* analysis (Sinclair and Mauranen 2006) – contributed to the determination.

To differentiate the phrases between clausal functions and DMs, prosodic information was of help. Type B phrases were usually a tone unit by themselves followed by a brief pause (Biber *et al.* 1999: 1076, Carter and McCarthy 2006: 211). This kind of pause might be indicated by a comma, but punctuation signs were not fully reliable because they were the subjective choice of the transcribers. In the present study, prosodic transcription was not provided in the corpora under investigation. The SECCL corpus was accompanied by recordings of all the texts; the MICASE corpus offered access to recordings of selected transcriptions and the ICE-GB corpus provided no sound tracks of the texts. To apply the same standard in the three corpora, therefore, prosodic information was not used for reference in the classification of Type A and Type B phrases.

Biber *et al.* (1999: 1076-1078) listed three criteria for determining “utterance launchers” as DMs, in which *you know*, *I mean*, *you see* and *I think* functioned alone and not as part of the following clause. First, that *that* could be added to examine the grammatical and discursal functions. Second, that *that* could not be added when “utterance launchers” were followed by a non-declarative clause, such as an interrogative. Third, that “utterance launchers” maintained the same interactive function in medial and final positions as they did in initial position. In Excerpt (3.3.5), inserting *that* after *you know* made the utterance grammatically correct but discursively inappropriate. Speaker 4 used the modal verb *could* and seemed uncertain about what s/he was saying and therefore this instance of *you know* did not assume that Speaker 4 knew what was going on. Moreover, *you know* seemed to function in a similar way when placed in the clause-final position. On the basis of these two criteria, this instance was classified as Type B.

(3.3.5)

S1: no no it's saying it doesn't exist.

S4: oh okay.

S1: but i was just, i was having a problem thinking like you know if somebody did say that, like the

wave thing and the, one, i can't see (both) i can, i felt like i could call that something. **[you know]** i could call that something. that something could exist, without being a particular one thing.
S4: oh you see that's why i was having trouble. i was i was thinking it could be a particular two things.....

(MICASE: SGR999MX115)

I mean in Excerpt (3.3.6) prefaced an interrogative and *that* could not be added here. This instance belonged to Type B, on the basis of Biber *et al.*'s second criterion.

(3.3.6)

C: Oh he's a baritone

D: Are you employed in a job like that **[I mean]** does he sing

C: He's a schoolteacher

D: Well that's what I thought yes I was suggesting

(ICE-GB: S1A-032)

In the cases where the above process was of no use for classification, the *LUG* analysis (Sinclair and Mauranen 2006) was used. For the instances of the phrases in extra-clausal utterance/turn-initial position, subjective judgement often had to be used as a last resort. (See Section 2.5 for the introduction to *LUG* and Section 12.2.6 for its usefulness and limitations.)

In the analysis chapters (Chapters 4-9), more instances of the words/phrases and their classification are discussed and exemplified in more detail. In the analyses of *like*, *well* and *now* where the distinction between Types A and B is clear, the discussion of Type A is relatively short, but in the case of phrases where the distinction is not obvious, for example, *I think*, more extensive discussion of Type A is needed in order to explain what Type B phrases are.

Since the tagging of DM (a function label in ICE-GB (Greenbaum 1996: 140)) is available in the ICE-GB corpus, it is intriguing to compare my manual classification with the tagging in ICE-GB. My manual tagging of Type A and Type B and the tagging of DMs in ICE-GB show very similar results. The frequencies of Type B *like*, *well*, *now*, *you know*, *I mean* and *you see* in the manual classification mostly resemble the numbers retrieved by the accompanying software *ICECUP* (for the frequency comparisons of DMs between the manual classification and the tagged ICE-GB, see Appendix 5). This resemblance adds confidence and credibility to both my manual examination process and the tagging in ICE-GB, which is likely to have pursued automatically by a computer program. This comparison is not made in the investigation of *oh* and *I think* because there is no classification in the analysis of *oh* and because *I think* is not categorised as a DM in ICE-GB.

The above process of manual classification between Types A and B words/phrases was carried out in all instances of the words/phrases under investigation. (When the raw frequency was over 400 times, random sampling procedure was used. See Section 3.3.6 below for more discussion.) The *Set* column in the *Concord* in *WordSmith Tools* (Scott 2004) allowed user-defined categories to be entered. At this early stage, *A* was entered when the node word was classified into Type A. *E* was entered for Type B words/phrases in extra-clausal position and *I* was entered for those in intra-clausal position. These categories were re-sorted, counted and subsequently categorised further.

3.3.3 Frequency comparisons between the monologic and dialogic genres and between the speech of the Chinese non-native speakers and native speakers

The frequency information of *like* across the six sub-corpora is taken as a model analysis in this section. In Table 4.1 in Chapter 4 (repeated here for ease of reading), the raw frequencies of *like* were produced using *WordSmith 4* (Scott 2004). The frequencies of *like* were normed on a basis of 10,000 words and the normalised frequencies of *like* revealed the distribution of *like* across the six sub-corpora, which answered one of the research questions set out in Section 1.1.2. The results of the manual classification were used to calculate the proportions of the words/phrases as DMs in order to ascertain whether they were primarily used as DMs in the sub-corpora. For example, in the sub-corpus of the highly monologic discourse mode, 15 out of 284 instances were categorised into Type B and therefore 5.3 per cent of the instances were used as DMs ($284/15 \times 100 = 5.3$). The percentages in the table indicated that *like* was not primarily used as a DM, except in the sub-corpus of the highly interactive discourse mode in MICASE, in which 57.3 per cent of the instances of *like* were used as DMs.

Table 4.1: Frequency information of *like* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percent-age (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	736	21.9	5 out of 300 ^a	1.7	0.4
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	984	16.5	9 out of 300 ^b	3	0.5
MICASE: 13 transcripts of the highly monologic discourse mode (American NSs)	134,096	284	21.2	15	5.3	1.1
MICASE: 48 transcripts of the highly interactive discourse mode (American NSs)	577,996	8,513	147.3	172 out of 300 ^c	57.3	84.4
ICE-GB: 70 unscripted monologues (British NSs)	153,646	235	15.3	6	2.6	0.4
ICE-GB: 90 private direct conversations (British NSs)	185,000	913	49.4	47 out of 300 ^d	15.7	7.7

* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B *like* per 10,000 words are based on an extrapolation of the percentages of Type B.

a, b, c and d in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

For the cases of random sampling, indicated with *a*, *b*, *c* and *d* in superscript in the table, the normalised frequencies of Type B *like* per 10,000 words were based on an extrapolation of the percentages of Type B. Take Case *a* for example. 1.7 per cent of the 736 instances of *like* in the Chinese NNSs' monologues in SECCL were Type B. On the basis of this information, an approximate number of instances of Type B *like* were obtained from 736 multiplied by 1.7 per cent. This number, 13, was normed on the basis of 10,000 words ($13/336,303 \times 10,000 = 0.4$). Therefore, Type B *like* might occur 0.4 times per 10,000 words.

From the normalised frequencies of Type B *like* in Table 4.1 above, it is self-evident that in both the speech of the Chinese NNSs and NSs, more instances of Type B *like* occurred in the dialogic genres than in the monologic genres. To support this finding, tests of statistical significance between the two types of genre and between speech of the Chinese NNSs and that of the NSs were carried out. The *z* test for two proportions was first done in *Minitab 15* after consultation with Dr Allan White of the Statistical Advisory Service of the University of Birmingham (for the results of the *z* test, see Appendix 7). The log-likelihood (LL) test, a common test of statistical significance in corpus studies, was also done (see Appendix 6 for the results of the LL test). These two tests revealed identical results when the level of significance was set at less than 0.01 (i.e. $p\text{-value} < 0.01$). In the discussion of frequency

comparisons in the analysis chapters, the LL test results are referred to, because the critical values in the LL test can be set at a higher value for the significance level of 0.0001 to increase reliability (Rayson, Damon and Brian 2004). This is further explained below.

Several significance tests are used to compare the distinctiveness between corpora; two of the most popular tests are the chi-square test and the LL test (McEnery, Xiao and Tono 2006: 55-56). The main disadvantage of the chi-square test is its limitation with small numbers (McEnery and Wilson 2001: 84). LL is generally preferred in the analysis of text, because it does not assume the word(s) for analysis will be normally distributed, which is often the case, given the nature of text (Dunning 1993). Accordingly, the LL test was used in this study.

The LL calculator created by Paul Rayson of Lancaster University (Rayson 2011) was used in order to have a robust indicator of the significance of the differences in frequency between the two broad types of genre and between the two groups of speakers. For ease in presenting the results, the sub-corpus of the Chinese NNSs' monologues in SECCL is referred to in Appendix 6 as Corpus A1, the sub-corpus of their dialogues as Corpus A2, the sub-corpus of the NSs' highly monologic discourse mode in MICASE as Corpus B1, the sub-corpus of the highly dialogic discourse mode as Corpus B2, the sub-corpus of the NSs' unscripted monologues in ICE-GB as Corpus C1 and the sub-corpus of the private direct conversations as Corpus C2. In the LL test, the critical values are 3.84, 6.63 and 10.83 for the levels of significance $p < 0.05$, $p < 0.01$ and $p < 0.001$ respectively. A p-value close to 0 indicates that there is a high statistically significant relationship, while a value close to 1 indicates that the difference is not significant and highly likely to be due to chance (Rayson *et al.* 2004: 7, McEnery *et al.* 2006: 55). To increase the reliability and applicability of the frequency comparisons of linguistic items, Rayson *et al.* (2004) recommend applying the critical value of 15.13 for significance at the $p < 0.0001$ level in corpus studies. In Appendix 6, LL scores displayed in strikethrough format are lower than the cut-off score, 15.13, which suggests that there is no statistically significant difference between the two sub-corpora being compared. A plus or minus symbol before the LL scores indicates over-representation or under-representation, respectively, in the first corpus relative to the second corpus.

In the analysis of *like*, the LL score -0.65 between Corpus A1 and Corpus A2 indicates that there is no statistically significant difference between the proportion of *like* as a DM in the Chinese NNSs' monologues and the proportion of this word in their dialogues. The scores

between Corpora B1 and B2 (-1882) and between Corpora C1 and C2 (-132.1) show a high level of significance and the under-representation in Corpora B1 and C1. This supports my hypothesis that DMs occur more often in the dialogic genres than in the monologic genres.

Between the speech of the Chinese NNSs and NSs, the differences are not significant, with -7.7 between Corpus A1 and Corpus B1 and 0 between A1 and C1, but the differences between the Chinese NNSs and NSs in the dialogic genres are highly significant. The LL scores between A2 and B2 and between A2 and C2 are -6593.41 and -268.75 respectively. This indicates that *like* as a DM is under-represented in the Chinese NNSs' dialogues when compared with the NSs' highly interactive discourse mode in MICASE and the private direct conversations in ICE-GB.

3.3.4 Collocates of the words and phrases

All the instances of each word/phrase and the instances of Type B word/phrase were processed separately in the *Concord* Tool in *WordSmith 4* (Scott 2004) to generate the pattern of the word/phrase. This listed the collocates (which should occur at least twice) of the word or phrase in order of frequency and provided an overall sense of the use of the word/phrase. The patterns are presented in the analysis chapters with highlighted collocates as indications of Type A use or Type B use, as well as possible co-occurring DMs. The highlighted collocates were used to cross-validate the findings showing whether the words/phrases were primarily used as DMs and to cross-examine the collocation phenomena surrounding the words/phrases. The patterns of Type B words/phrases helped to identify co-occurring DMs and present the preference for the orders of DM collocations.

3.3.5 Identifying the positions of Type B words and phrases

After the process of distinguishing Types A and B (see Section 3.3.2), the instances of Type A were taken out of the concordance lists. The instances of Type B words/phrases were further examined to identify their positions in an utterance/turn. The instances were classified into two broad categories, extra- and intra-clausal positions. In the former category, three sub-categories, utterance/turn-initial, -medial and -final were discerned. The instances in intra-clausal position were described with the element units of the *LUG* analysis (Sinclair and Mauranen 2006), which are discussed in Section 2.5 of Chapter 2 in detail.

The information on the positions of Type B words/phrases was added to the *Set* column

in the *Concord* in *WordSmith Tools* (Scott 2004). The instances of Type B words/phrases in extra-clausal utterance/turn-initial position were coded as *Ei*, those in extra-clausal utterance/turn-medial position as *Em* and those in extra-clausal utterance/turn-final position as *Ef*. The instances in intra-clausal position were coded as *I* following the *LUG* labels. These categories were re-sorted and counted and used for a further investigation of the collocation phenomena surrounding DMs.

3.3.6 Grammatical and discursal aspects of Type A and Type B words and phrases

Following the discussion of frequency information and the identification of the positions in an utterance/turn, the grammatical aspects, such as word classes and syntactical structures, of each word/phrase were examined to draw a distinction between Types A and B. Next, the word/phrase was analysed in the *LUG* analysis to illustrate how Types A and B fitted into this new model.

Co-occurrence of Type B words/phrases (DMs) was identified on the basis of linguistic evidence (e.g. hesitation markers, emphatic lexis, reported speech etc.), not an existing analytical framework or schema. In the literature, DMs have been typically described in terms of their functions, but the use of the term *function* was problematic because the researchers could read the linguistic evidence only and could never read the speaker's mind. Some of the researchers in the literature seemed to be unable to establish empirically the functions of DMs but instead interpret them intuitively. In my study, types of co-occurrence of the DMs were first identified and then taken as evidence for determining the categories for discussion, with function being secondary in interpretation. This procedure made the logic of identification of the functions clearer. The final results were shown in tables with types of co-occurrence in relation to the position of the word/phrase in question.

3.3.7 Random sampling procedure

The Type A and Type B words and phrases and the identification of the positions in an utterance/turn of Type B words and phrases had to be classified manually. In cases where the instances were too numerous (more than 400 instances) for complete manual inspection, three sets of 100-line concordance samples were used. The use of three random 100-line concordance samples has been proved to be a sufficient basis in Groom's PhD thesis (2007).

A set of 100-line concordance samples was manually grouped. In order to obtain more

reliable data of distribution and test the certainty of the manual classification of Types A and B, another two sets of the same size were classified for cross validation. The three sets of 100-line concordance samples were obtained by using the *Concord* Tool in *WordSmith 4* (Scott 2004). It was set to randomly select one in every designated line (calculated from the total instances of the word/phrase divided by 100). This would be more likely to produce a random sample of 100 lines equally distributed across each of the texts in the corpus under investigation. In the analysis of *like*, the number of times that four of the six sub-corpora occurred (noted with *a*, *b*, *c* and *d* in superscript in Table 4.2 above) were 736, 984, 8513 and 913, respectively. Accordingly, to obtain a 100-line concordance sample, the *Concord* Tool was set to return one instance in every 7, 9, 85 and 9 lines from these four sub-corpora respectively. The instances in each of the three sets of 100-line concordance samples were manually classified into Types A and B (see Section 3.3.2), assigned their positions in an utterance/turn (see Section 3.3.5) and the collocation phenomena (see Section 3.3.6) were identified.

It can be seen in Table 3.4 below, across the three sets of random samples, that the proportion of Types A and B *like* and the distribution of positions in an utterance/turn (extra-clausal or intra-clausal) were found to be similar. It was very clear that the three sets in these two cases revealed similar distributions, which indicates that the random sampling procedure was reliable.

Table 3.4: Distribution of *like* in the three sets of 100-line concordance samples

a. <i>Like</i> in the Chinese NNSs' monologues in SECCL	1st set of 100-line concordance samples	2nd set of 100-line concordance samples	3rd set of 100-line concordance samples	Total
Type A (Non-discourse use)	99	98	98	295
Type B (Discourse use: Extra-clausal)	0	1	1	2
Type B (Discourse use: Intra-clausal)	1	1	1	3
b. <i>Like</i> in the Chinese NNSs' dialogues in SECCL	1st set of 100-line concordance samples	2nd set of 100-line concordance samples	3rd set of 100-line concordance samples	Total
Type A (Non-discourse use)	97	98	96	291
Type B (Discourse use: extra-clausal position)	2	1	1	4
Type B (Discourse use: intra-clausal position)	1	1	3	5
c. <i>Like</i> in the American NSs' highly interactive discourse mode in MICASE	1st set of 100-line concordance samples	2nd set of 100-line concordance samples	3rd set of 100-line concordance samples	Total
Type A (Non-discourse use)	48	39	41	128
Type B (Discourse use: extra-clausal position)	22	32	23	77
Type B (Discourse use: intra-clausal position)	30	29	36	95
d. <i>Like</i> in the British private direct conversations in ICE-GB	1st set of 100-line concordance samples	2nd set of 100-line concordance samples	3rd set of 100-line concordance samples	Total
Type A (Non-discourse use)	82	89	82	253
Type B (Discourse use: extra-clausal position)	11	4	5	20
Type B (Discourse use: intra-clausal position)	7	7	13	27

The instances of Type B *like* in the three sets of 100-line concordance samples were then further analysed to identify co-occurrence. The distribution of their co-occurrence was discussed in relation to the positions in an utterance/turn. The analyses of the random samples of *like* in the sub-corpus of the NSs' highly interactive discourse mode (Case *c* in Table 3.4) are shown as an example in Tables 3.5 (the 1st set of 100-line concordance samples), 3.6 (the 2nd set), 3.7 (the 3rd set) and 3.8 (the three sets combined).

Unlike the similar proportion of Types A and B *like* and the similar distribution of positions in an utterance/turn across the three sets of 100-line concordance samples, the distribution of co-occurrence of Type B *like* is rather different. In the 1st set of random samples (Table 3.5), the most frequent co-occurrence is *explanations* (23.1%) and the least is *reported speech* (1.9%). In the 2nd set (Table 3.6), *expressions of uncertainty* (29.5%) accounts for most and *reported speech* (3.3%) least. In the 3rd set (Table 3.7), *expressions of uncertainty* (28.8%) is the most frequent co-occurrence and *reported speech* (0%) least. The figures in Table 3.8 combine those in Tables 3.5, 3.6 and 3.7 and show the distribution of the co-occurrence of 300 instances of Type B *like*.

Given that the three sets of 100-line concordance samples reveal their similar distribution in the classification of Types A and B and in the extra- and intra-clausal positions, it would

seem reasonable to conclude that three sets of 100-line concordance samples are a sufficient basis for drawing conclusions about the frequency, the proportion of Types A and B and their positions, but the figures for the distribution of co-occurrence in the three sets of random samples are sometimes not similar to each other. If one set of random samples alone had been used, the distributions of co-occurrence might have been misleading.

Table 3.5: *Like* as a discourse marker in MICASE: Highly interactive discourse mode (The 1st set of 100-line concordance samples)

Co-occurrence	Freq.		Extra-clausal						Intra-clausal							
			Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	10	19.2	1	1.9	3	5.8			3	5.8	1	1.9	2	3.8		
2. Numerical expressions and locations	6	11.5			1	1.9			4	7.7					1	1.9
3. Reported speech	1	1.9												1	1.9	
4. Expressions of uncertainty	4	7.7	1	1.9					1	1.9			1	1.9	1	1.9
5. Expressions of certainty/ key point	8	15.4							7	13.5			1	1.9		
6. Exemplifications	2	3.8			2	3.8										
7. Explanations	12	23.1	1	1.9	5	9.6			6	11.5						
Unclassified	9	17.3	1	1.9	2	3.8	5	9.6	1	1.9						
Occurrences: 52 out of 100 (random samples)	52	100.0	4	7.7	13	25.0	5	9.6	22	42.3	1	1.9	4	7.7	3	5.8

Table 3.6: *Like* as a discourse marker in MICASE: Highly interactive discourse mode (The 2nd set of 100-line concordance samples)

Co-occurrence	Freq.		Extra-clausal						Intra-clausal							
			Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	8	13.1	1	1.6	5	8.2							2	3.3		
2. Numerical expressions and locations	3	4.9							2	3.3					1	1.6
3. Reported speech	2	3.3			2	3.3										
4. Expressions of uncertainty	18	29.5			10	16.4			6	9.8					2	3.3
5. Expressions of certainty/ key point	12	19.7			4	6.6			8	13.1						
6. Exemplifications	8	13.1	1	1.6	4	6.6			3	4.9						
7. Explanations	5	8.2			1	1.6			4	6.6						
Unclassified	5	8.2	1	1.6	1	1.6	2	3.3	1	1.6						
Occurrences: 61 out of 100 (random samples)	61	100.0	3	4.9	27	44.3	2	3.3	24	39.3			2	3.3	3	4.9

Table 3.7: *Like* as a discourse marker in MICASE: Highly interactive discourse mode (The 3rd set of 100-line concordance samples)

Co-occurrence	Freq.	%	Extra-clausal						Intra-clausal							
			Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	8	13.6	1	1.7	1	1.7			2	3.4			3	5.1	1	1.7
2. Numerical expressions and locations	4	6.8			2	3.4			2	3.4						
3. Reported speech	0	0														
4. Expressions of uncertainty	17	28.8	2	3.4	2	3.4	1	1.7	9	15.3					3	5.1
5. Expressions of certainty/ key point	9	15.3			2	3.4			7	11.9						
6. Exemplifications	9	15.3			4	6.8			3	5.1			1	1.7	1	1.7
7. Explanations	4	6.8			3	5.1			1	1.7						
Unclassified	8	13.6	2	3.4	1	1.7	2	3.4	3	5.1						
Occurrences: 59 out of 100 (random samples)	59	100.0	5	8.5	15	25.4	3	5.1	27	45.8			4	6.8	5	8.5

Table 3.8: *Like* as a discourse marker in MICASE: Highly interactive discourse mode (The three sets of 100-line concordance samples)

Co-occurrence	Freq.	%	Extra-clausal positions						Intra-clausal positions							
			Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	26	15.1	3	1.7	9	5.2			5	2.9	1	0.6	7	4.1	1	0.6
2. Numerical expressions and locations	13	7.6			3	1.7			8	4.7					2	1.2
3. Reported speech	3	1.7			2	1.2								1	0.6	
4. Expressions of uncertainty	39	22.7	3	1.7	12	7.0	1	0.6	16	9.3			1	0.6	6	3.5
5. Expressions of certainty/ key point	29	16.9			6	3.5			22	12.8			1	0.6		
6. Exemplifications	19	11.0	1	0.6	10	5.8			6	3.5			1	0.6	1	0.6
7. Explanations	21	12.2	1	0.6	9	5.2			11	6.4						
Unclassified	22	12.8	4	2.3	4	2.3	9	5.2	5	2.9						
Occurrences: 172 out of 300 (random samples)	172	100.0	12	7.0	55	32.0	10	5.8	73	42.4	1	0.6	10	5.8	11	6.4

3.4 Qualitative text-based analysis

The corpus analyses helped to bring evidence of typical co-occurrence and distribution of DMs for analysis in the six sub-corpora, making it possible to empirically establish the functions of DMs rather than interpret them intuitively. Nevertheless, this bottom-up approach to the data revealed that genre seems to be a key factor in using DMs, but it was unable to give an adequate explanation for the observed under- or over-representation of the DMs. The qualitative text-based analyses were employed to examine the data more closely and test some hypotheses, which could not be done with the corpus methodologies.

The corpus investigation helped direct me to selective texts for qualitative text-based analyses. Some texts were analysed to expose broader contexts; all the DMs occurring in one text were studied together, in the hope of seeing a whole picture, and some aspects which could not be revealed by using corpus methodologies. Chapters 10 and 11 report the text-based analyses of NSs' and Chinese NNSs' speech respectively.

3.5 Key terms used in the study

This section presents the definitions of the key terms used in this thesis. Some of the terms are arbitrary choices and some are used as other researchers have used them.

Discourse marker (DM): DMs have been defined in different ways (see the review of the literature in Section 2.3). For the purpose of my thesis, I give five characteristics for determining a DM: 1) optionality, 2) flexibility of position, 3) prosodic independence, 4) connectivity and 5) multi-grammaticality.

Type A: Type A refers to the non-discourse use of the words and phrases in question. That is the words and phrases are used with their word classes, i.e. verbs, adverbs, adjectives, preposition, nouns, etc.

Type B: Type B refers to the discourse use of the words and phrases in question. This is the use of words/phrases as DMs.

Utterance: An utterance is a stretch of speech, which may contain pauses and silence. This definition applied equally to a one-word response token.

Turn: A turn is an utterance preceded or followed by a change of speaker.

A number of terms are used for the differences between monologues and dialogues and

between lectures and conversations. There is some confusion between these terms, partly because their meanings depend on the researchers that these terms come from. For example, Biber (see Biber (1988) and Biber *et al.* (1999)) uses *genre* and *register* interchangeably, which is very different from the term *register* in Halliday and Hasan's work (1985). *Speech event* comes from the tradition of the academic field of ethnography, so researchers outside this area tend not to use *speech event*. The selection of term is, to some extent, arbitrary. I chose to use *type of activity*, which is theory-neutral, to refer to what other researchers might call *genre*, *speech event* or *register*.

Type of activity: Types of activity in this thesis refer to speech events. In the NNS corpus under investigation, the type of activity is exams. In the sub-corpus of the highly monologic discourse mode in MICASE, the types of activity include lectures to large classes (more than 40 students) and to small classes (40 or fewer students) and colloquia. The sub-corpus of the highly interactive discourse mode in MICASE includes office hours, meetings, seminars, advising sessions, lab sessions, lectures to small classes, study group discussions, student presentations and interviews. In the subset of the unscripted monologues in ICE-GB, there are demonstrations, legal presentations, spontaneous commentaries and unscripted speeches. In the subset of the dialogues, the type of activity is private direct conversation.

Genre: The term *genre* is used to broadly refer to a particular type of texts in this study, i.e. monologic and dialogic. One of the aims in this thesis is to compare the use of DMs across the monologic genres and dialogic genres extracted from the SECCL corpus, the MICASE corpus and the ICE-GB corpus.

3.6 Codes and conventions

Table 3.9 below lists the codes and conventions used in the examples cited in this thesis. Some of them (Items 1 to 7) are created for this study and some are from the markup conventions in the corpora under investigation.

CHAPTER 4: ANALYSIS OF *LIKE*

4.1 Introduction

This chapter discusses the word *like*. Although the word *like* is not found in use as a DM in the selection process reported in Section 3.2 above, it is chosen for analysis in particular because the increasing number of studies of *like* in the NS speech (e.g. Dailey-O’Cain (2000), Cukor-Avila (2002), Siegel (2002), Fuller (2003b), Fox Tree (2006, 2007), Fox Tree and Tomlinson Jr. (2008), Barbieri (2009) and Jones and Schieffelin (2009)).

My hypotheses about the use of *like* and research questions are given first in this section, followed by a survey of the literature. As stated in the preceding chapter on research methods, a bottom-up approach is employed. The analysis begins with frequency data and patterns of all the instances of *like* and those of its discourse use (Type B) in the six sub-corpora under investigation. The major part of the analysis is the discourse aspects of Type B *like*, looking at its positions in an utterances/turn and the collocation phenomena surrounding *like*. The identification of co-occurrence leads to the interpretations of the functions of Type B *like*.

NNSs, in general, learn *like* as a verb and a preposition in the early stages of their learning. Moreover, some uses of Type B *like* in the speech of NS are fairly recent. I therefore hypothesise that in the Chinese NNSs’ speech under investigation, there are many fewer occurrences of Type B *like* than in the NSs’ speech. If this is so, the proportion of Type B *like* is different in the NNS data and the NS data and the uses of Type B *like* in the NNSs’ speech are not as varied as those in the NSs’ speech.

I test my hypotheses within the framework of the core research questions addressed in this thesis (see Section 1.1.2). Question 1 below asks proportion of Type A *like* and Type B *like* in the NNSs’ and NSs’ speech and ascertains whether the Chinese NNSs seldom use Type B *like*. Questions 2 to 4 concern the overall use of *like* and their answers support the manual classification of *like* between Type A and Type B (answering Question 1) while validating some claims about the uses of Type B *like*, which are based on types of co-occurrence and contextual information (answering Questions 5 and 6). The answers to the following six questions help to indicate the use made by the NNSs and NSs of Type B *like*.

1. What is the distribution of the word *like* in the NNSs’ and NSs’ speech?
2. What do the collocates of *like* reveal about its use?

3. How do the NNSs and NSs use Type A *like* and Type B *like*?
4. What other DMs does Type B *like* co-occur with?
5. Where does Type B *like* appear in an utterance/turn?
6. With what types of co-occurrence or in what contexts does Type B *like* tend to occur?

The distinction between Type A *like* and Type B *like* is clear-cut and in most cases, can be drawn without difficulty. As mentioned in Section 3.3.2, Type A *like* is straightforwardly identified by referring to its word classes and co-occurring syntactical structure. This is reviewed in the next section.

As specified in Chapter 3, since the tagging of DM is available in the ICE-GB corpus, it is interesting to compare my manual tagging with that in ICE-GB. In Appendix 5, the frequencies of Type B *like* in the manual classification mostly resemble the numbers retrieved by the accompanying software *ICECUP*. This adds confidence to both my manual examination process and the tagging in ICE-GB.

4.2 Previous studies of *like*

The use of Type A *like* is discussed in Sections 4.2.1 and 4.2.2. Section 4.2.3 demonstrates how Types A and B *like* are examined in *LUG*. Last, previous studies of Type B *like* are reported.

4.2.1 Word classes of *like*

Type A *like* can be categorised as a verb, preposition, noun, adverb, conjunction or adjective. In addition to these word classes, the pattern *BE + like* has been generally accepted as used for marking reported speech. However, Type B *like* does not belong to any word classes, but rather to a generally-accepted category, *discourse marker*. As noted in Chapter 2, the category *discourse marker* is currently popular in the literature.

4.2.2 Syntactical aspect of *like*

Type A *like* in the speech of the NNSs and NSs occurs mainly in two word classes, a verb and a preposition. As a verb, *like* expresses being fond of someone or something and, when preceded by *would*, wanting to do something. As a preposition, *like* expresses a similar comparison or examples.

The syntactical structures of *like* as a verb and as a preposition are straightforward. The patterns in Tables 4.4 to 4.9 below show that *like* as a verb tends to be preceded by the pronouns, *I*, *you*, *we* and *they*, as in Excerpt (4.2.1) below. It can also be part of the verb phrase *would like*, as in Excerpt (4.2.2). It tends to be followed by a gerund, *playing*, *to* + infinitive, as in Excerpt (4.2.3) and pronouns *her*, *him*, *me* and *you*.

(4.2.1)

.....I will teach my students as he as her way, and **I like** to be a teacher as, as her.

(SECCL: B01-67-19)

(4.2.2)

..... today **i'd like** to talk about the solution to this problem in sphere packings the Kepler Conjecture

(MICASE: COL385MU054)

(4.2.3)

B: **I like playing**<play> badminton.

A: Badminton?

B: Yeah! My... this is my favorite hobby.

(SECCL: C00-74-24)

Like as a preposition tends to be preceded by such verbs as *look* and *BE*, as in Excerpt (4.2.4) and such nouns as *something* and *things* and followed by demonstrative pronouns *this* and *that*, as in Excerpt (4.2.5).

(4.2.4)

.....if you look at the figures over here at first glance this could **look like** a mirrored reflection of her back

(MICASE: LEL320JU143)

(4.2.5)

.....Now, as the university student, some... some of them study grammar, sentence and sentence structure and **things like that** all in the classroom. And they are very tired and I did not do that, I did not only do that. Of course, I do that too, but I also do some movie.

(SECCL: B01-08-21)

Similar to other DMs, Type B *like*, in general, occurs flexibly in the clause and is syntactically optional (see Section 2.3.2 for a detailed discussion of the characteristics of DMs). It seems not to follow any syntactical rules. In Excerpt (4.2.6) below, *like* occurs between a prepositional phrase and a clause. In Excerpt (4.2.7), *like* is placed between an adverb and an adjective, which is not a major grammatical junction. In Excerpt (4.2.8), *like* is inserted into an infinitive, *to go over*.

(4.2.6)

S1: mhm and i i would say that it for the next hundred years it'll probably always mean stops stop, but **with American political culture |like| it's always changing** and like during campaigning the way you do it it's always different so

(MICASE: DIS495JU119)

(4.2.7)

S3: mhm you know what i mean like, **very |like| early** life forms where like,
S4: yeah, cuz it's says they're moving

(MICASE: SGR175MU126)

(4.2.8)

S1: okay. so i probably_ it probably would be more helpful **to |like|, go over** main concepts and like lecture notes and things like that you

S3: <OVERLAP2> yeah </OVERLAP2>

S1: <OVERLAP1> think? </OVERLAP1>

(MICASE: SGR175MU126)

The use of Type B *like* in the last three examples above follows no syntactical rules. Moreover, using traditional labels to describe the position of Type B *like* seems to imply ungrammatical use and such instance as *like* in Excerpt (4.2.8) cannot be described in grammars based on written English. Therefore, the newly-established device, *Linear Unit Grammar (LUG)* (Sinclair and Mauranen 2006), is used in this thesis to describe where DMs occur in an utterance/turn (see Section 2.5 in Chapter 2 for a detailed discussion of *LUG*).

4.2.3 *Linear Unit Grammar analysis of like*

Like as a verb in Excerpt (4.2.9) below (the same as Excerpt (4.2.1) above) is part of an M-element in the *Linear Unit Grammar (LUG)* analysis (Sinclair and Mauranen 2006) (see Appendix 4 for a list of the labels in *LUG*). In this case, *like* makes a propositional contribution. By contrast, *like* in Excerpt (4.2.10) (the same as Excerpt (4.2.8) above) is an O element, which does not augment knowledge but make the discourse flow. It is further categorised as an OI element rather than an OT element, because *like* in this example does not create cohesion at the textual level.

(4.2.9)

.....I will teach my students as he as her way, and I like to be a teacher as, as her.
M- +M MS OT M- +M MA MS

(SECCL: B01-67-19)

(4.2.10)

S1: okay. so i probably it probably would be more helpful to like, go over main concepts and like
OI OI MF M- +M- OI +M OT OI
lecture notes and things like that you
M OT M M-
S3: <OVERLAP2> yeah </OVERLAP2>
OI
S1: <OVERLAP1> think? </OVERLAP1>
+M

(MICASE: SGR175MU126)

As discussed in Section 2.5 of Chapter 5, *LUG* is used to facilitate the classification of Types A and B and to assign units in speech for the description of the position of DMs. The above two examples demonstrate that the *LUG* analysis is able to accommodate *like* as both Type A and Type B. The position of *like* in Excerpt (4.2.9) is between an OT and a +M and that in Excerpt (4.2.10) is between a +M- and a +M.

4.2.4 Previous studies of *like* as a discourse marker

There are three major uses of *like*: 1) *like* as a verb, preposition, noun, adverb, conjunction and adjective; 2) *like*, preceded by forms of *BE*, for quoting; and 3) *like* as a DM (Müller 2005: 197, Fox Tree 2006: 724). Most research on *like* has been done on the second and third uses. *Like* for quoting has been documented and there has been a marked increase in some varieties of English, such as African American (e.g. Cukor-Avila (2002: 212)) and Canadian (e.g. Tagliamonte (2005)) and American English (e.g. Fox Tree and Tomlinson Jr. (2008) and Jones and Schieffelin (2009)). Studies of *like* as a DM include Schourup (1985), Jucker and Smith (1998), Andersen (1998), Siegel (2002), D'Arcy (2005), Fox Tree (2006, 2007), etc.

In this thesis, the first and second uses are treated as Type A, non-discourse use. The use of *like* for quoting is still in the process of being grammatically recognised. The construction *BE + like* has been documented in the *Oxford Advanced Learner's Dictionary* (2000) and the *Longman Dictionary of Contemporary English* (2009: 1014), but it is listed in *like* as an adverb in the *Cambridge Grammar of English* (Carter and McCarthy 2006: 102) and the *Longman Grammar of Spoken and Written English* (Biber *et al.* 1999: 1120). The use of *like* for quoting is classified differently. It is recognised that *like* is used with *BE*, however. *BE + like* possesses idiomatic status. Although *BE* can be conjugated, it cannot be used in negation with reported speech (Andersen 1998: 162-163). Furthermore, on the basis of the five characteristics of DMs for this thesis (see discussion in Section 2.3.2), Type B *like* is semantically and syntactically optional, whereas the pattern *BE + like* for quoting makes

neither *BE* nor *like* optional and it is therefore treated here as Type A.

In the literature, researchers agree on the main functions of Type B *like* but fail to reach a consensus on their frequencies. It has been pointed out that *like* co-occurs with numerical expressions in order to withhold strong commitment from the exactness and correctness of the utterance (Schourup (1985) and Andersen (2001) cited in Müller (2005: 210) and Jucker and Smith (1998)). *Like* co-occurring with numerical expressions is one of the frequent uses found in Schourup's study (1985), but is least frequent in Müller's (2005), in which a quarter of the instances are produced by NNSs and the remainder by NSs.

Schourup ((1985: 42) cited in Müller (2005: 198)) claims that *like* is in most cases used as "a device available to speakers to provide for a loose fit between their chosen words and the conceptual material their words are meant to reflect". This can explain speakers' weak commitment to utterances which they follow by *like*.

In contrast to Schourup's argument (1985: 42), Underhill (1988: 241-245) offers instances where *like* does not mean *approximately*, *sort of* and *similar to* and he claims that "*like* functions as an approximator much less often than one might think"; instead, *like*, as a "focuser", marks new and significant information.

One of the meanings of *like* as a preposition is *for example*. Similarly, Type B *like* is followed by exemplifications. The former has its syntactical role and must not be omitted. The latter is syntactically optional. Type B *like* occurring in the context of exemplifications has been documented in the literature, such as Schourup ((1985: 48) cited in (Müller 2005: 213)) and Jucker and Smith (1998: 184, 188).

Other functions of Type B *like* identified in the literature include *like* for hedging, for indicating a search for words and for restarting. *Like* preceding an explanation has rarely been discussed, except in Müller's study (2005: 215). It occurs more frequently in her American NNSs' speech than in the German NNSs'.

4.3 Frequency information in the speech of the non-native speakers and native speakers

4.3.1 Overall frequency of *like*

As listed in Table 4.1 below, there are 736 and 984 occurrences of *like* in the NNSs' monologues and dialogues respectively. Less than 3% of these occurrences are Type B *like*.

The raw frequencies of Type B *like* are normed on a basis of 10,000 words. The normalised counts show that *like* occurs with similar frequencies across the two types of genre in the NNSs' speech. In the monologues, there are 0.4 instances of *like* per 10,000 words and in the dialogues, 0.5 instances.

Considering the number of speakers, in the NNSs' data, it can be concluded that the NNSs rarely use Type B *like*. 1,143 speakers produce only 42 instances⁶ of Type B *like* in their monologues and dialogues.

Table 4.1 below shows that there are variations in frequency between the monologic and dialogic genres and between the speech of the NNSs and NSs. In terms of genre, there are more instances of *like* in the dialogic genre than in the monologic genre. In the NNSs' monologues and dialogues, the subset of the highly monologic discourse mode in MICASE and two subsets in ICE-GB, a small proportion is Type B *like*. However, in the subset of the highly interactive discourse mode in MICASE, *like* is primarily used as a DM, accounting for 57.3% of all the occurrences.

⁶ The count of *like* as a discourse marker in the Chinese NNS speech under investigation is based on an extrapolation of the percentage of random samples (see Table 4.1). 42 instances are from the raw frequency of *like* multiplied by the percentage of Type B. In the monologues, 736 multiplied by 1.7% equals 12.5, and in the dialogues, 984 times 3% is 29.5.

Table 4.1: Frequency information of *like* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percent-age (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	736	21.9	5 out of 300 ^a	1.7	0.4
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	984	16.5	9 out of 300 ^b	3	0.5
MICASE: 13 transcripts of the highly monologic discourse mode (American NSs)	134,096	284	21.2	15	5.3	1.1
MICASE: 48 transcripts of the highly interactive discourse mode (American NSs)	577,996	8,513	147.3	172 out of 300 ^c	57.3	84.4
ICE-GB: 70 unscripted monologues (British NSs)	153,646	235	15.3	6	2.6	0.4
ICE-GB: 90 private direct conversations (British NSs)	185,000	913	49.4	47 out of 300 ^d	15.7	7.7

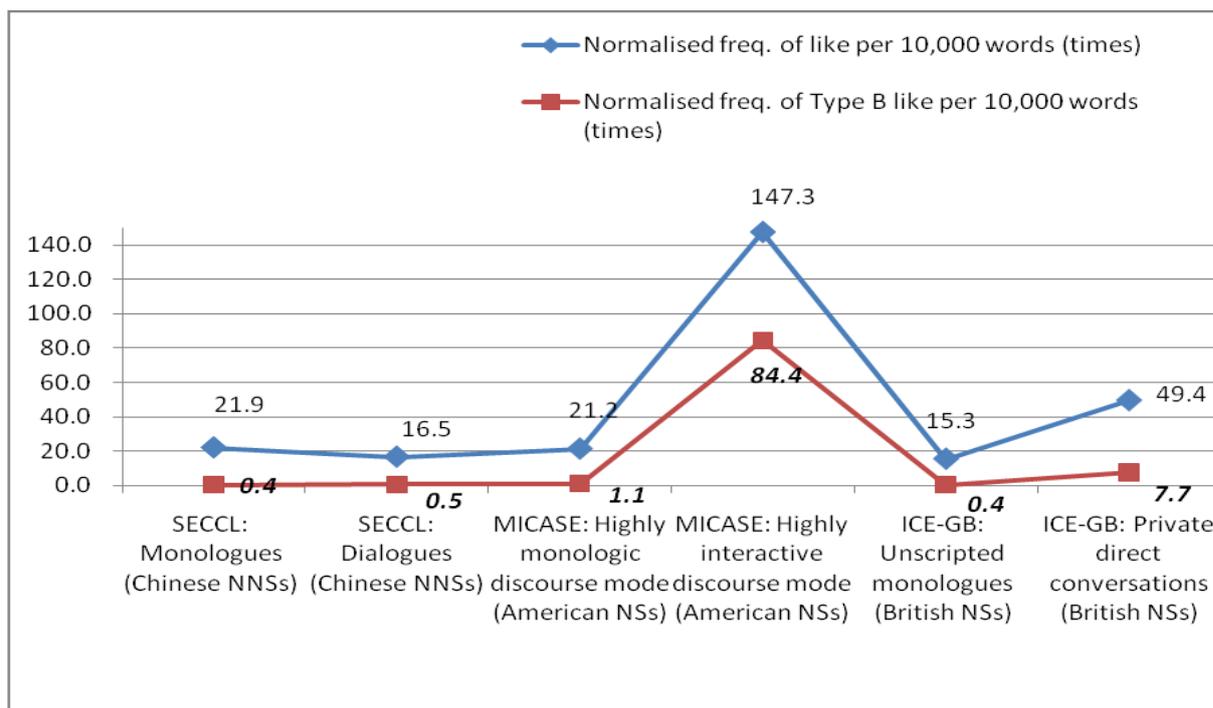
* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B *like* per 10,000 words are based on an extrapolation of the percentages of Type B.

a, b, c and d in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

The raw counts of *like* are normed on a basis of 10,000 words and the normalised counts, ranging from 15.3 to 147.3 times across the six sub-corpora, are shown in Table 4.1 above. The same normalisation is used on the frequencies of Type B *like*, which range from 1.7 to 57.3 times. Figure 4.1 below shows the comparison of normalised frequencies of *like*. It can be clearly seen that, in the NS sub-corpora, there are more instances of *like* in the dialogic genres than in the monologic genres. This supports my hypothesis that the more interactive the genres or types of activity are, the more DMs occur. Interestingly, there are considerably more instances of Type B *like* in the subset of the highly interactive discourse mode in MICASE with 84.4 instances per 10,000 words, as opposed to fewer than 7.7 instances in the five other sub-corpora.

Figure 4.1: Comparison of normalised frequencies of *like* across sub-corpora



As explained in Section 3.3.3 of Chapter 3, to test if the frequencies differ from each other significantly, the log-likelihood (LL) test was used to carry out a test of statistical significance between the two types of genre and between the speech of the Chinese NNSs and that of the NSs. The results of the LL test are presented in Appendix 6. It is found that there is a statistically significant difference between the two types of genre in MICASE (LL:-1882, p-value: < 0.0001) and ICE-GB (LL:-132.1, p-value: < 0.0001). The negative LL scores indicate that Type B *like* is under-represented in the monologic genres. In the speech of Chinese NNSs, there is no statistically significant difference between the two types of genre. As mentioned above, Chinese NNSs seldom use *like* as a DM. Neither is the difference in the monologic genres between the Chinese NNSs and NSs significant (LL: -7.7 between Corpora A1 and B1 and LL: 0 between Corpora A1 and C1). However, in the dialogic genres, the difference between the two groups of speakers is highly significant (LL: -6593.41 between Corpora A2 and B2 and LL: -268.75 between Corpora A2 and C2). It can be concluded that DM *like* appears much more frequently in the NSs' dialogic genres, in particular in MICASE.

4.3.2 Collocates of *like*

The patterns of *like* (Tables 4.2 to 4.7) in the subsets of SECCL, MICASE and ICE-GB show

the overall use of *like*. The two patterns of the NNSs' monologues and dialogues (see Tables 4.2 and 4.3) reveal similar collocates of *like*. The collocates to the left, *I, you, we, they* and *would* and to the right, *to, her, him, me, you* and *playing*, highlighted in boldface, suggest that *like* is used as a verb. The collocates to the left, *just, things* and *looks* and to the right, *that* and *this*, highlighted in both boldface and italics, suggest that *like* is used as a preposition. The manual classification identifies that more than half the 300 occurrences of *like* in the monologues and also in the dialogues are used as a verb. The concordancer, which returns all instances of *like*, is set to search for the word *like* rather than the lemma *LIKE*, which includes such conjugated forms as *likes* and *liked*. Therefore, it is very possible that there are more instances of *LIKE* as a verb in the NNSs' speech than the identified percentage, 50%. *Like* as a preposition is also frequently used by the NNSs. Further investigation finds that 38.6% (116) in the monologues and 32.6% (98) in the dialogues of the 300 instances of random sampling function as a preposition. It is concluded that *like* as a verb and as a preposition are its two main uses in the Chinese NNSs' speech under investigation.

Table 4.2: Pattern of *like* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (16)	like (19)	and (24)	i (44)	like (300)	to (62)	very (15)	like (15)	the (13)
2	very (10)	i (12)	i (23)	just (37)		a (28)	i (11)	i (14)	he (11)
3	a (10)	and (8)	don (11)	t (18)		the (21)	talk (10)	much (12)	i (9)
4	the (9)	you (7)	is (8)	you (16)		that (16)	and (9)	and (10)	and (9)
5	she (9)	she (7)	the (8)	d (10)		this (12)	to (8)	she (9)	us (8)
6	and (9)	of (6)	didn (7)	things (9)		it (11)	she (8)	he (9)	so (7)
7	my (6)	the (6)	he (7)	we (8)		her (9)	friend (6)	the (8)	my (7)
8	he (6)	said (6)	eh (6)	not (8)		him (8)	eh (6)	with (8)	a (6)
9	2 (6)	he (5)	um (6)	um (8)		me (7)	we (6)	about (5)	very (6)
10	will (5)	is (5)	to (6)	looks (7)		you (6)	tell (5)	um (5)	she (6)

Table 4.3: Pattern of *like* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (18)	you (16)	i (33)	you (48)	like (298)	to (68)	a (21)	i (21)	i (19)
2	a (16)	i (12)	you (12)	i (47)		that (31)	i (14)	a (16)	the (16)
3	to (10)	a (11)	would (11)	just (32)		the (23)	but (13)	in (12)	a (13)
4	b (10)	eh (10)	b (10)	d (19)		you (23)	and (11)	like (9)	you (13)
5	you (10)	and (10)	and (10)	something (16)		a (13)	eh (10)	you (9)	think (11)
6	eh (10)	b (9)	or (9)	would (14)		this (11)	b (9)	but (8)	b (9)
7	the (9)	the (9)	do (9)	t (8)		it (10)	like (7)	eh (8)	and (9)
8	and (8)	do (8)	if (8)	they (6)		some (6)	you (7)	b (8)	like (7)
9	some (8)	like (7)	don (8)	think (6)		eh (6)	have (7)	the (8)	so (6)
10	yes (7)	some (6)	a (7)	not (5)		playing (5)	very (7)	and (7)	can (6)

In the pattern of the NSs' highly monologic discourse mode (see Tables 4.4), the collocates immediately to the left, *something, things, just, look, looked, looks* and *it's*, highlighted in both boldface and italics, seem to suggest that *like* is used as a preposition. Fewer collocates, such as *I'd, to, it* and *you*, suggest *like* may be used as a verb. This finding correlates with the manual classification of Types A and B. In the NSs' highly monologic discourse mode, 79% (225) of the 284 instances of *like* are used as prepositions and 13% (36) as verbs. In the highly interactive discourse mode, in 30% (89) of the 300 instances of random sampling, it is used as a preposition and 6% (18) as a verb.

Table 4.4: Pattern of *like* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (14)	of (10)	of (13)	<i>something</i> (23)	like (283)	<i>this</i> (49)	and (14)	the (13)	the (19)
2	it (10)	to (10)	a (13)	<i>things</i> (16)		the (27)	one (9)	that (11)	and (8)
3	that (8)	that (9)	it (12)	<i>just</i> (9)		<i>that</i> (26)	s (8)	and (10)	that (8)
4	in (8)	you (8)	that (7)	<i>look</i> (9)		a (19)	you (7)	d (8)	to (7)
5	a (7)	um (6)	you (6)	<i>looked</i> (9)		<i>to</i> (15)	so (6)	to (8)	in (7)
6	and (7)	what (6)	to (6)	<i>looks</i> (8)		l (8)	on (5)	you (7)	you (7)
7	to (7)	uh (5)	or (6)	<i>it's</i> (7)		uh (7)	a (5)	what (7)	i (6)
8	you (6)	so (5)	and (5)	of (6)		i (6)	the (5)	or (6)	a (6)
9	of (6)	is (5)	have (5)	institutions (5)		<i>it</i> (5)	it (4)	uh (5)	if (4)
10	is (4)	but (4)	sort (5)	<i>i'd</i> (5)		<i>you</i> (4)	is (4)	in (5)	an (4)

In Table 4.5 below, the highlighted collocates related to the use of Type A *like* are not as strong as those in the NNSs' data in terms of the frequency. This leads one to suppose that *like* in the NSs' highly interactive discourse mode are more likely to be Type B. As shown in Table 4.1 above, 57.3% of the instances of *like* in the NSs' highly interactive discourse mode are Type B.

Table 4.5: Pattern of *like* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	like (12)	like (13)	i (19)	and (13)	like (291)	the (18)	like (9)	like (11)	the (13)
2	i (9)	i (11)	it (14)	<i>just</i> (11)		a (13)	you (9)	and (7)	to (11)
3	know (7)	the (9)	like (10)	<i>it's</i> (8)		i (12)	the (9)	you (7)	like (9)
4	the (7)	you (9)	the (8)	<i>looks</i> (8)		<i>that</i> (12)	i (8)	the (7)	a (8)
5	and (7)	and (7)	you (8)	mean (7)		<i>this</i> (10)	and (7)	or (6)	and (8)
6	you (7)	that (7)	is (5)	to (7)		<i>you</i> (9)	a (6)	that (5)	you (6)
7	that (5)	what (6)	this (5)	but (5)		<i>it</i> (6)	know (6)	it (5)	i (6)
8	of (5)	to (6)	and (5)	know (5)		they (6)	just (5)	one (5)	but (6)
9	we (4)	in (5)	that (5)	that (5)		in (6)	don't (4)	of (5)	is (5)
10	a (4)	of (4)	it's (4)	so (5)		if (6)	that (4)	if (5)	i'm (5)

In the patterns (Tables 4.6 and 4.7) of the NSs' unscripted monologues and private direct conversations in ICE-GB, the first ten immediate left/right collocates are close to those in the subsets of SECCL and MICASE. The highlighted collocates in the tables below seem to suggest that *like* is used as a verb and a preposition. The manual classification confirms that most of the instances of *like* in ICE-GB use it as a preposition, with 65.5% (154) of the 235 instances in the unscripted monologues and 55% (165) of the 300 instances in the private direct conversations.

Table 4.6: Pattern of *like* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	of (11)	of (9)	i (26)	d (22)	like (235)	to (35)	and (10)	and (12)	the (10)
2	a (11)	and (8)	if (17)	you (22)		the (23)	you (8)	of (9)	a (8)
3	the (10)	the (7)	a (7)	something (12)		this (18)	uh (8)	to (8)	of (8)
4	you (7)	a (7)	it (7)	looks (11)		that (15)	the (8)	i (8)	s (6)
5	to (6)	what (6)	you (7)	would (10)		a (15)	to (6)	it (7)	it (5)
6	and (6)	it (6)	that (6)	things (8)		it (7)	this (4)	a (7)	to (5)
7	uh (4)	this (6)	of (5)	i (6)		you (7)	see (4)	the (6)	and (4)
8	s (4)	if (6)	with (5)	look (6)		and (6)	or (3)	uh (5)	in (4)
9	they (4)	that (5)	and (5)	rather (6)		uhm (5)	talk (3)	on (5)	like (4)
10	like (3)	up (5)	is (4)	anything (5)		uh (4)	thank (3)	in (4)	on (3)

Table 4.7: Pattern of *like* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (17)	it (19)	i (32)	things (20)	like (300)	that (44)	and (11)	i (20)	the (12)
2	you (12)	i (16)	and (22)	i (17)		a (23)	i (10)	you (9)	and (11)
3	it (8)	but (7)	it (18)	something (13)		i (19)	mm (8)	but (8)	you (11)
4	to (8)	and (7)	you (15)	s (13)		to (18)	a (8)	it (8)	it (10)
5	and (7)	like (6)	s (15)	don't (12)		this (14)	you (8)	s (7)	i (9)
6	of (7)	the (5)	a (9)	just (12)		the (13)	s (7)	and (7)	a (7)
7	the (6)	to (5)	sort (8)	you (12)		it (12)	to (6)	that (6)	of (7)
8	mean (6)	think (5)	would (7)	of (10)		uhm (9)	that (6)	like (6)	s (6)
9	a (6)	what (5)	that (6)	d (6)		you (6)	but (6)	just (6)	uhm (6)
10	if (6)	a (5)	he (6)	it (6)		oh (4)	know (5)	know (5)	is (5)

In the NNSs' speech, *like* is more often used as a verb than a preposition; in the NSs' speech, the reverse is true. One reason may be that the topics for the NNSs (see Appendix 1), such as describing a person or an event, offer more opportunities for the NNSs to use *like* as a verb, which indicates subjectivity. The NNSs were more likely to get responses with a high level of subjectivity in their discussion than would typically be found in lectures in the NS corpus. Most of the NS data in MICASE are composed of lectures, in which the speakers may frequently use *like* with such demonstrative pronouns as *this* and *that* in order to explain or

give examples.

To look more closely at the use of Type B, the patterns of Type B *like* in the NNSs' and NSs' speech are shown in Tables 4.8 to 4.13. In the NNSs' speech, *like* tends to co-occur with the hesitation marker *eh*, while in the NSs' speech, *like* in MICASE tends to follow *and*, *I mean* and *but*, which could be DMs and in the conversations in ICE-GB it tends to collocate with such vague language as *sort of*. (The pattern of Type B *like* in the unscripted monologues in ICE-GB (Table 4.12) produces no collocates of the six instances of Type B *like*.)

Table 4.8: Pattern of Type B *like* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	which (2)	the (2)	problem (2)	is (2)	like (5)	eh (3)	eh (2)	how (2)	many (2)

Table 4.9: Pattern of Type B *like* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (2)	college (2)	just (2)	eh (3)	like (9)	eh (3)	eh (2)	for (2)	a (2)
2	at (2)			just (2)					

Table 4.10: Pattern of Type B *like* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1		and (2)	they (2)	to (2)	like (15)	this (2)	you (2)	up (2)	and (3)
2				have (2)				to (2)	

Table 4.11: Pattern of Type B *like* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (6)	like (8)	<i>i</i> (10)	and (11)	like (165)	i (9)	you (8)	if (5)	the (8)
2	know (6)	the (7)	like (8)	mean (7)		the (8)	and (7)	of (5)	and (6)
3	like (4)	what (5)	the (8)	to (6)		if (5)	the (6)	you (5)	i (5)
4	that (4)	you (5)	it (6)	that (5)		in (5)	a (5)	it (4)	to (4)
5	so (4)	and (5)	you (5)	have (4)		they (4)	like (4)	like (4)	is (4)
6	just (4)	in (4)	and (4)	but (4)		there's (4)	just (4)	the (4)	you (4)
7	about (3)	to (4)	they (3)	so (4)		you (4)	it (3)	this (3)	a (4)
8	the (3)	of (3)	is (3)	uh (3)		this (3)	of (3)	with (3)	that (3)
9	think (3)	i (3)	bed (3)	know (3)		one (3)	think (3)	or (3)	this (3)
10	and (3)	were (2)	to (3)	for (3)		what's (3)	don't (3)	and (3)	were (3)

Table 4.12: Pattern of Type B *like* in the native-speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1					like (6)				

Table 4.13: Pattern of Type B *like* in the native-speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (3)	i (4)	<i>sort</i> (7)	<i>of</i> (10)	like (47)	a (5)	a (4)	you (4)	the (5)
2	a (3)	the (4)	you (4)	this (2)		uhm (4)	s (3)	in (3)	and (4)
3	up (2)	girls (2)	kind (3)	thought (2)		oh (3)	on (2)	shop (2)	they (2)
4	within (2)	and (2)	i (3)	we (2)		i (3)	you (2)	up (2)	re (2)
5	what (2)	at (2)	they (2)	uhm (2)		uh (2)	up (2)	a (2)	know (2)
6	hard (2)		and (2)	have (2)		hold (2)	i (2)	don't (2)	of (2)
7	oh (2)		school (2)	and (2)		the (2)	course (2)	be (2)	
8	realise (2)		being (2)	s (2)		it (2)	must (2)	s (2)	
9				know (2)		shut (2)	it (2)		

4.4 Discourse aspects of Type B *like*

In this section, the positions in an utterance/turn where Type B *like* occurs are first described and then the linguistic items which Type B *like* tends to co-occur with are discussed. This investigation of the positions and collocation phenomena leads to my interpretations of the use of Type B *like* in the speech of the Chinese NNSs and the American and British NSs.

4.4.1 Positions in an utterance/turn

The distribution and percentages of *like* in an utterance/turn in the six sub-corpora under investigation are shown in Table 4.14 below. Intra-clausal positions are preferred in the NNSs' monologues and dialogues, the two subsets of MICASE and the private direct conversations of ICE-GB. In contrast, extra-clausal positions are more often used in the unscripted monologues of ICE-GB.

Like in intra-clausal positions tends to occur after an M- element. All occurrences in extra-clausal positions are utterance-medial in the three sub-corpora of the monologic genres. This can be attributed to the categories of the classification. In the NNSs' monologues, each monologue is taken as an utterance. There are 1,143 utterances, while in the 1,143 dialogues, there are 29,542 turns⁷. In the NS data, this piece of information is not supplied.

⁷ The two speakers in the NNSs' dialogues are identified as *a* and *b*. Searching for *a:* and *b:* reveals 15,110 instances of *a:* and 14,432 of *b:*. From these figures, it can be inferred that there are probably 29,542 turns in the NNSs' 1,143 dialogues.

Table 4.14: Distribution of the positions of Type B *like* in an utterance/turn

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues	
	Random samples (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)
Positions in an utterance of <i>like</i>	5	100	15	100	6	100
Extra-clausal: utterance-initial	0	0.0	0	0.0	0	0.0
Extra-clausal: utterance-medial	2	40.0	3	20.0	4	66.7
Extra-clausal: utterance-final	0	0.0	0	0.0	0	0.0
Intra-clausal: after an M-	2	40.0	11	73.3	0	0.0
Intra-clausal: after an MA	0	0.0	0	0.0	0	0.0
Intra-clausal: after an MF	1	20.0	0	0.0	0	0.0
Intra-clausal: others	0	0.0	1	6.7	2	33.3
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations	
	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)
Positions in an utterance/turn of <i>like</i>	9	100	172	100	47	100
Extra-clausal: turn-initial	2	22.2	12	7.0	4	8.5
Extra-clausal: turn-medial	2	22.2	55	32.0	15	31.9
Extra-clausal: turn-final	0	0.0	10	5.8	0	0.0
Intra-clausal: after an M-	1	11.1	76	44.2	19	40.4
Intra-clausal: after an MA	1	11.1	1	0.6	1	2.1
Intra-clausal: after an MF	0	0.0	7	4.1	2	4.3
Intra-clausal: others	3	33.3	11	6.4	6	12.8

4.4.1.1 *Like* in extra-clausal position

In the NNSs' and NSs' speech, *like* is often placed in extra-clausal utterance/turn-medial position, as in Excerpts (4.4.1) and (4.4.2).

(4.4.1)

P: Utterance-medial

.....But now although he had, we still keep on writing to each other um, we still talk about a lot of things um |like| we just sat sitting face to face um.

(SECCL: B98-21-12)

(4.4.2)

P: Utterance-medial

.....in this experiment, they took some some mouse melanoma cells were taken and injected into the tail vein of the mice, just (male) mice just |like|, the experiment on twenty-one, on twenty-one we focused on the lung.

(MICASE: LEL175SU106)

A small proportion of the instances of *like* in the NNSs' dialogues (22.2%), the NSs' highly interactive discourse mode (7%) and the direct conversations (8.5%) are used as turn

openers, as exemplified in Excerpt (4.4.3).

(4.4.3)

P: Turn-initial S3: what do you mean higher?
S2: um **|like|** there's an example in the book. like s- for example it, that one that's got more carbons attached it is higher priority than this one.
(MICASE: SGR200JU125)

Utterance/turn-final position is least used, as in Excerpt (4.4.4). In most cases, the speakers are interrupted by others rather than stopping voluntarily after *like*. This use is not found in the NNSs' dialogues.

(4.4.4)

P: Turn-final S12: hey we're no we're talking about like, this whole <OVERLAP1> article and it it <OVERLAP1> affects people who support **|like|**
S8: <OVERLAP2> that's a good one <OVERLAP2>
S2: well i'm just saying that not <OVERLAP1> everyone's <OVERLAP1>
S12: <OVERLAP2> and it's <OVERLAP2> gonna be against his decision right here
(MICASE: LES220SU140)

4.4.1.2 *Like* in intra-clausal position

In the NNSs' speech, over half (60% in the monologues and 55.6% in the dialogues) the instances of Type B *like* are placed in intra-clausal position, as in Excerpts (4.4.5) and (4.4.6).

(4.4.5)

P: M- + *like* + OI Task 2
I remember when I was in high school, the teacher gave <give> us assignment on geography problem, and to solve out a, a, a difficult task... eh... ... which involves... eh... ... which the problem is **|like|**... eh... ... eh... ... how many satellites is used to cover the whole, the whole globe.....
(SECCL: B99-25-29)

(4.4.6)

P: MA + *like* + M A: So... you can pay your time let... um... you can do some other things rather than eh... just reading the text book, then you won't be so bored.
B: Yes, I suppose. I should go out and do some jokin **|like|** jogging like you. Maybe next time we'll have a... time, pleasant time, playing basketball or something else.
(SECCL: C98-17-26)

In the NSs' speech, a slightly larger proportion, 80% in the NSs' highly monologic discourse mode, 55.2% in the highly interactive discourse mode and 59.6% in the private direct conversations, occur in intra-clausal position, as Excerpt (4.4.7) exemplifies:

(4.4.7)

P: M- + *like* + +M

S3: let's, yeah, let's </OVERLAP1> just talk about |**like**| endergonic and
<EVENT DESC="DRAWING ON BOARD"></EVENT> exergonic so, that's four then
right?

(MICASE: SGR175SU123)

4.4.2 Contexts where Type B *like* tends to occur

The positions of Type B *like* in an utterance/turn are described in the preceding section and the results are referred to in the present discussion of the contexts where *like* tends to occur. The contexts are identified on the basis of collocation phenomena. *Like* is found to co-occur with 1) hesitation markers, pauses and restarts, 2) numerical expressions and locations, 3) reported speech, 4) expressions of uncertainty, 5) expressions of certainty and key points, 6) exemplifications and 7) explanations. Most of these types of co-occurrence are found in the NNSs' speech, because the Chinese NNSs seldom use Type B *like*.

The instances in ambiguous contexts, with no linguistic evidence and insufficient contextual information remain unclassified in my analysis. The tables at the end of this section (Tables 4.15 to 4.20) show the distribution of the types of co-occurrence of Type B *like* in relation to its positions in an utterance/turn in the six sub-corpora under investigation.

4.4.2.1 *Like* co-occurring with hesitation markers, pauses and restarts

In the NNSs' speech, *like* co-occurring with the hesitation markers *eh* and *um* and pauses, as in Excerpt (4.4.8), is one of the frequent contexts where *like* tends to occur. In the patterns of Type B *like* in the NNSs' monologues and dialogues (see Tables 4.8 and 4.9 above), *eh* and *um* are two of the most frequent collocates.

(4.4.8)

P: M- + *like* + OI

E: Hesitation marker,
eh; pauses

F: To suggest a
search for contents
or lexis

I remember when I was in high school, the teacher gave <give> us assignment
on geography problem, and to solve out a, a, a difficult task... eh... ... which
involves... eh... ... which the problem is |**like**|... **eh**... ... **eh**... ... how many
satellites is used to cover the whole, the whole globe.

(SECCL: B99-08-08)

In Excerpt (4.4.9), *like* co-occurs with hesitation marker *um*. Although it is not possible to draw direct conclusions about cognitive processes from the evidence of utterances, it is arguable that when *like* co-occurs with hesitation markers and pauses, this is because the speaker is searching for content information or appropriate lexical expressions. In addition, a

possible interpretation from the other supporting evidence is that Item (1) *I think I understand what you're saying*, is that the speaker has just made an attempt to understand Speaker 1.

(4.4.9)

<p>P: Turn-initial</p> <p>E: Hesitation marker, <i>um</i>; pauses; Item (1) indication of cognitive process</p> <p>F: To suggest a search for contents or lexis</p>	<p>S2: <OVERLAP2> yeah or like um, </OVERLAP2></p> <p>S1: <OVERLAP1> do you follow what i'm saying? like, </OVERLAP1> i don't know if that makes any sense.</p> <p><PAUSE DUR=":06" ></PAUSE></p> <p>S2: ⁽¹⁾ i think i understand what you're saying... um</p> <p style="text-align: right;">(MICASE: SGR565SU144)</p>
--	---

In Excerpt (4.4.10), Item (2) rephrases the latter part of Item (1) and *like* occurs between them, marking a restart.

(4.4.10)

<p>P: <u>MA</u> + <i>like</i> + <u>M</u></p> <p>E: Item (2) rephrases Item (1)</p> <p>F: To suggest a search for contents or lexis</p>	<p>S5: okay so therefore the D-N-A couldn't reproduce and ⁽¹⁾ couldn't ex- like ⁽²⁾ be encapsulated in the protein and like explode out. do you know what i mean cuz like, do you see what I remember like, um the step before this is like this this and then like this is the picture i remember it was like some like weird shape with</p> <p style="text-align: right;">(MICASE: SGR175MU126)</p>
---	---

In the NNSs' monologues, 4 out of 5 instances of Type B *like* and in the dialogues, 5 out of 9 instances co-occur with hesitation markers, pauses and restarts. This category shows the strong co-occurrence of Type B *like* in the NNSs' speech. It can reasonably be argued that the NNSs need more time to search for lexical items or content information when speaking in English and use *like* to maintain the floor. There is only one instance in this category in the NSs' highly monologic discourse mode, probably because in this sub-corpus most of the speakers, primarily lecturers, have authority and prepare well for speaking in the classroom.

4.4.2.2 *Like* co-occurring with numerical expressions and indications of location

As mentioned earlier, *like* co-occurring with numerical expressions is one of the frequent uses in Schourup's study (1985), but is the least frequent use in Müller's (2005), in which a quarter of the instances are produced by the German NNSs and the remainder from the NSs. In my data, this type of co-occurrence is only found in the sub-corpus of the NSs' highly interactive discourse mode in MICASE, accounting for 7.6%.

In the previous literature (Jucker and Smith (1998), Schourup (1985: 38) and Andersen

(2001) cited in (Müller 2005: 210)), it has been pointed out that *like* co-occurs with numerical expressions in order not to give a strong commitment to the exactness and correctness of the utterance. *Like* in Excerpt (4.4.11) below seems to belong to this category, as it is more likely that the speaker gives the approximate time, *two*, for going to bed. This interpretation can also be applied to the second instance of *like* in the same turn.

(4.4.11)

<p>P: <u>M-</u> + <i>like</i> + <u>+M</u></p> <p>E: Numerical expressions</p> <p>F: To make an approximation</p>	<p>S2: you're right you're right. you win a prize.</p> <p>S3: s- so tired right now. i slept at two o'clock and woke up at, like eight o'clock.</p> <p>S2: why did you like oh</p> <p>S3: prayers</p> <p>S2: <u>i went to bed at like two</u> and woke up at <u> like </u>, ten forty-five. (MICASE: SGR385SU057)</p>
---	---

Like in Excerpt (4.4.12) also co-occurs with a numerical expression, *three and five*, but it does not suggest that the two numbers are approximate, because Item (1) confirms that Speaker S4's *three and five* refer to specific items. This use demonstrates, as Underhill (1988: 241-245) argues, *like* being used as a “focuser”, marking new and significant information.

(4.4.12)

<p>P: <u>+M-</u> + <i>like</i> + <u>+M</u></p> <p>E: Numerical expressions; Item (1)</p> <p>F: To focus the coming information</p>	<p>S4: yeah like i don't, i don't get like, <u>between like three and five</u>, i don't (agree with) that</p> <p>S6: well ⁽¹⁾he used an example with three, about like, i think it was like something about watching a movie, (MICASE: SEM475JU084)</p>
---	---

The two instances of *like* in Excerpts (4.4.13) and (4.4.14) co-occur with a description of locations. In Excerpt (4.4.13) *like* is followed *over there somewhere*, a description of a rough indication of an area; therefore, *like* could suggest approximation. In Excerpt (4.4.14), *like* is followed by *right there*, which is a specific location, so that Speaker 1 realises that he is referring to *Justin*. In the NSs' highly interactive discourse mode, 4 out of 172 instances are found to co-occur with descriptions of place.

(4.4.13)

P: M- + *like* + +M
E: Item (1) refers to location
F: To make an approximation

S3: and uh, you'll get my paper, and hopefully it'll be good.
S2: i'm
S3: <OVERLAP2> oh </OVERLAP2>
S2: <OVERLAP1> sure </OVERLAP1> it'll be good.
S3: (there you go.) (xx)
S2: (put) the little |like|⁽¹⁾ over there somewhere and i (trust i- your judgment)
S3: i put it in the
S2: okay

(MICASE: OFC115SU060)

(4.4.14)

P: M- + *like* + +M
E: Item (1) refers to location
F: To focus the coming information

S1: <OVERLAP1> what one guy? </OVERLAP1>
S4: yeah he was hilarious.
<EVENT DESC="LAUGH"></EVENT>
S2: <OVERLAP1> oh yeah. </OVERLAP1>
<EVENT DESC="LAUGH"></EVENT>
S3: <OVERLAP1> the one who always </OVERLAP1> said the really random stuff.
S4: he sat |like|⁽¹⁾ right there.
S2: the one that was really like, gung ho
<EVENT WHO="SS" DESC="LAUGH"></EVENT>
S1: oh Justin
S2: yeah.

(MICASE: SEM300MU100)

4.4.2.3 *Like* co-occurring with reported speech

Similar to the previous type of co-occurrence, reported speech⁸ is less frequently used. No instance is found in the NNSs' speech and the NSs' monologic genres. 3 (1.7%) out of 172 instances of *like* co-occurring with reported speech in the sub-corpus of the NSs' highly interactive discourse mode and 8 (17%) out of 47 in the sub-corpus of the NSs' private direct conversations are identified. In Excerpt (4.4.15) below, *like* precedes the reported speech, *why most people*.

(4.4.15)

P: OT + *like* + +M
E: Item (1) is a reported speech
F: To mark reported speech

S6: yeah i asked my dad that |like|⁽¹⁾ why most people but he said something that um the U-S was looked at as a better place to go to that it was harder to get here, something like that but

(MICASE: OFC115SU060)

4.4.2.4 *Like* co-occurring with expressions of uncertainty

Like co-occurs with expressions of uncertainty, such as *I don't know*, *I guess*, *I suppose*, *probably*, *maybe*, *could*, *right?*, *sort of* and *kind of*. This type of co-occurrence seems to indicate that *like* can be interpreted as a hedging device, as argued by Schourup ((1985: 42)

⁸ The use of *like* co-occurring with reported speech is different from that of *BE + like* for quoting, which is treated as non-discourse use. See the discussion in Section 4.2.4.

cited in Müller (2005: 198)).

In Excerpt (4.4.16) below, Item (2) suggests that the speaker is not very certain about Item (1), which is preceded by *like*. Therefore, it seems to be reasonable to assume that *like* is used as a hedging device.

(4.4.16)

P: M- + *like* + +M

E: Item (2) shows uncertainty about Item (1)

F: To express uncertainty

S1: well they ha- you know the tru- they have the trucks the trucks are specialized for ethylene production.

S3: and they have **|like|**⁽¹⁾ little compartments that, spray out ⁽²⁾ **i suppose it's** **like** a little gassing truck or something.

(MICASE: LES405JG078)

In Excerpt (4.4.17) *like* co-occurs with vague language *kind of* to express uncertainty or imprecision.

(4.4.17)

P: M- + *like* + +M-

E: Vague language, *kind of*

F: To express uncertainty

B: I just just realised that it was actually the st study of architecture I really enjoyed <,>

B: And uh you just **kind of** **|like|** get a a few hints at what actually working in the profession's like

(ICE-GB: S1A-034)

This use is only found in the NSs' speech, with 2 (13.3%) out of 15 instances in the NSs' highly monologic discourse mode, 39 (22.7%) out of 172 in the NSs' highly interactive discourse mode in MICASE and 10 (21.3%) out of 47 in the NSs' private direct conversations in ICE-GB.

4.4.2.5 *Like* co-occurring with expressions of certainty and key points

In contrast to the previous type of co-occurrence, *like* co-occurs with expressions of certainty, such as *gonna* and *going to*. In Excerpt (4.4.18), Item (1) indicates that the speaker is quite sure about his action of paying for the ticket.

(4.4.18)

P: M- + *like* + +M

E: Item (1) shows the speaker's decision

F: To express certainty

S2: what if you have a ticket?

S1: what if i do?

S2: uhuh.

S3: it's twenty dollars.

S1: i, mail in a, check with the ticket and, pay it off. it's my girlfriend's car so ⁽¹⁾ **i'm not gonna** **|like|** make her pay it.

S2: yeah. i'll just pay it off, you know. what else am i gonna do?

(MICASE: SGR385SU057)

Like is found to co-occur with a key point. In Excerpt (4.4.19), Item (1) seems to be of importance, since such emphatic lexis as *especially* and *very* is used. In Excerpt (4.4.20), Item (1) is the response to the key information in Speaker 1's utterance.

(4.4.19)

P: Utterance-medial
E: Emphatic lexis,
especially and *very*
in Item (1)
F: To signal a key
point

B: Oh, it's very, very dangerous, but in China, there's also bad things.
A: But I think um it is safer than go, than go abroad. Because, because you know in... in... some other countries **|like|**... ⁽¹⁾ **especially you say some discriminations are very... eh... .. very bad.** So I think our Chinese people will be treated bad.

(SECCL: C01-99-15)

(4.4.20)

P: M- + *like* + +M
E: Item (1) as a
response to the key
point
F: To signal a key
point

S1: <OVERLAP1> so </OVERLAP1> basically all of this area is kinda like upper middle-class, west- this is our |like| west suburbs
S2: ⁽¹⁾ **i see so it's just showing the west suburbs**
S1: pretty mu- yeah

(MICASE: OFC115SU060)

Like co-occurring with expressions of certainty and key points can be seen as a “focuser”, orientating the listener (Underhill 1988: 241-245). This use occurs only once in the NNSs' speech, but it is one of the frequent types of co-occurrence in the NSs' speech, with 2 (13.3%) out of 15 instances in the highly monologic discourse mode, 29 (16.9%) out of 172 in the highly interactive discourse mode, 1 (16.7%) out of 6 in the unscripted monologues and 5 (10.6%) out of 47 in the private direct conversations.

4.4.2.6 *Like* co-occurring with exemplifications

Like is found to co-occur with exemplifications in both the NNSs' speech, as in Excerpt (4.4.21) and the NSs' speech, as in Excerpt (4.4.22). Exemplifications can be broadly divided into two variants; one of which is exemplifications preceded by a general term or description, as illustrated in Excerpt (4.4.21), while the other is not being preceded by a general term or description, as shown in Excerpts (4.4.22) and (4.4.23).

(4.4.21)

P: Utterance-medial
E: Items (2)-(4)
exemplify Item (1)
F: To introduce
exemplifications

..... She usually let us give her ⁽¹⁾ **some performances**, in the class, **|like|** with sometimes give her ⁽²⁾ **a play**, ⁽³⁾ **a speech** or ⁽⁴⁾ **a debate** and things like that. She stressed the group work and we found that you may ask more active in the class.

(SECCL: B01-08-15)

(4.4.22)
P: Utterance-medial
E: Item (2)
 exemplifies Item
 (1)
F: To introduce
 exemplifications

.....they set off on a five year voyage. um, his life on the Beagle itself wasn't that great. it was, this cramped little boat and he had about as ⁽¹⁾ **much room** as, i don't know maybe, **|like|** ⁽²⁾ **from here to here.** and he was always seasick

(MICASE: LEL175JU154)

(4.4.23)
P: Turn-initial
E: Item (2)
 exemplifies Item
 (1)
F: To introduce
 exemplifications

A: You send pictures of you
 B: No I don't send pictures of me Other people send pictures of me I
 <unclear-word> ⁽¹⁾ **I send pictures** I like
Like| like ⁽²⁾ **this friend of mine Andy** I always send a picture of uhm Andy
 Warhol

(ICE-GB: S1A-015)

Compared with other types of co-occurrence, *like* occurring with exemplifications is more frequent, with 1 (20%) out of 5 instances in the NNSs' monologues and 2 (22.2%) out of 9 in the dialogues. In the NSs' speech, 7 (46.7%) out of 15 instances are found in the highly monologic discourse mode in MICASE and 19 (11%) out of 172 instances in the highly interactive discourse mode. In ICE-GB, there are 2 (33.3%) out of 6 instances in the unscripted monologues and 4 (8.5%) out of 47 in the private direct conversations.

4.4.2.7 *Like* co-occurring with explanations

Like is found in the context where an explanation is provided. In Excerpts (4.4.24) and (4.4.25) below, Item (2) is an explanation of Item (1) and *like* occurs between these two items.

(4.4.24)
P: Turn-medial
E: Item (2) explains
 Item (1)
F: To introduce an
 explanation

A: Yes, as a... a so young student he can deal with these difficulties in the outside world.
 B: Yes.... Of course, we can... I think I think ⁽¹⁾ **the different culture is a good thing to us not bad |like|**... ⁽²⁾ **everybody has its own feature. I think every culture has own characteristics. We can learn much from that.**

(SECCL: B01-01-05)

(4.4.25)
P: Utterance-medial
E: Item (2) explains
 Item (1)
F: To introduce an
 explanation

S1: <OVERLAP1> i didn't have, </OVERLAP1> i just didn't have, as much of an idea i don't think i liked it as well as Buddhism. and, then, also i j- i don't know what it was i ⁽¹⁾ **the papers were different in length |like|** the, ⁽²⁾ **Buddhism papers were only two to three pages, whereas my Chinese philosophy were like eight, seven or eight,** which might have something to do with it but, not a D minus worth.

(MICASE: SEM300MU100)

As mentioned earlier in Section 4.2.4, this context has rarely been discussed in the literature, except in Müller's study (2005: 215). In the NNSs' and NSs' speech under investigation, *like* co-occurring with an explanation is found. As Müller (2005) notes, more instances of this are observed in the speech of NSs than the NNSs. In the present study, there are no instances of it in the NNSs' monologues and only 1 (11.1%) out of 9 instances in the dialogues. In the NSs' speech, there are 2 (13.3%) out of 15 instances in the highly monologic discourse mode, 21 (12.2%) out of 172 in the highly interactive discourse mode, 2 (33.3%) out of 6 in the unscripted monologues and 9 (19.1%) out of 47 in the private direct conversations.

4.4.2.8 Unclassified instances of *like*

In the NNSs' speech, Type B *like* is rarely used. All the instances found in the random samples are classified and discussed in the previous section. In the NSs' highly interactive discourse mode, 22 (12.8%) out of 172 instances of Type B *like* are found impossible to classify. 12 of these, mostly in turn-final position, remain unclassified due to the interruption by another speaker, as exemplified in Excerpt (4.4.26).

(4.4.26)

P: Turn-final	SU-f: American Redstart that was the variations on sweet wasn't it?
E: Interrupted by	SU-f: mhm
Speaker SU-m	SU-f: and they have like
F: Unidentified	SU-m: on what? on sweet sw-

(MICASE: LAB175SU026)

In addition to the 12 unclassified instances of *like* in the NSs' highly interactive discourse mode, some instances, as exemplified in Excerpts (4.4.27) and (4.4.28), seem idiosyncratic to the speakers, since in their utterances, they frequently use Type B *like*.

If taking the roles of the speakers into consideration, between two senior undergraduates in Excerpt (4.4.27) and between a junior undergraduate and a non-teaching university employee in Excerpt (4.4.28), it could be speculated that the instances of Type B *like* function as such hesitation markers as *eh* and *um*, or function as solidarity markers to create vagueness, hedging or uncertainty in order to avoid sounding assertive.

(4.4.27)

- P:** Turn-medial S3: <OVERLAP1> yeah i can meet the </OVERLAP1> workers from there fine i think. i'm a little bit i mean it's the owners
- E:** overly use of Type B *like* S2: <OVERLAP2> yeah </OVERLAP2>
S3: <OVERLAP1> i'm </OVERLAP1> more nervous about. **like** i mean they're really, when they get really busy and **like**,
- F:** Speaker's idiosyncrasy S2: yeah i don't know. we'll see. mhm. i can see why that would be nerve-racking.
S3: right. so i'm trying to finish up **like** i only have **like** a chapter left in, um this buis- this book and then |**like**| there's one chapter in this one that's about my topic and **like** i've read five books, <OVERLAP1> and </OVERLAP1>
S2: <OVERLAP2> that's </OVERLAP2> a lot. yeah like i really spent a- and they were they're all pretty u- they were all pretty useful?
(MICASE: SGR999SU146)

(4.4.28)

- P:** M- + *like* + +M S3: i know it_ kind of like the corny way **like** a lot of kids might learn about it my father is a doctor,
- E:** overly use of Type B *like* S1: uhuh
S3: and so i mean **like** early on i, understood some things about the medical profession **like**, my parents' friends were all doctors i'd always like interacted with them
- F:** Speaker's idiosyncrasy S1: <OVERLAP2> mhm </OVERLAP2>
S3: <OVERLAP1> and, </OVERLAP1> i really, i think i i really would enjoy that lifestyle
S1: mhm
S3: **like** i'm starting to volunteer at our local hospital and,
S1: mhm
S3: and stuff um, i've always leaned i'm more of a math science person,
S1: mhm
S3: definitely i've alwa- i've always leaned towards the math and science, even though, **like** i took a creative writing class my senior year. my senior year though, i really like, i just as soon as college hit, (it) was like as soon as college questions hit it seemed that i went from **like** less specific as what i always thought i wanted to be like i always thought i wanted to be a doctor but then **like**, my senior year, when college |like| confronted me it seemed like oh no but there are all these other options i want to explore,
S1: yeah
S3: like creative writing and, and such, but i've always leaned toward math and science. and, i've always really i took anatomy and biology and really enjoyed both,
(MICASE: ADV700JU023)

The remaining instances are unclassified due to the lack of linguistic evidence and insufficient contextual information.

4.4.2.9 Summary of the contexts where Type B *like* tends to occur

Tables 4.15 to 4.20 show the distribution of the identified types of co-occurrence in relation to the positions in an utterance/turn of Type B *like*. It is clear that Type B *like* occurs much less often in the NNSs' speech and most of the instances tend to co-occur with hesitation markers, pauses and restarts, while the instances of Type B *like* in the NSs' speech are found to have a

variety of types of co-occurrence.

In the literature, *like* co-occurring with numerical expressions and expressions of uncertainty is widely discussed; however, this is found only in the sub-corpus of the NSs' highly interactive discourse mode.

Like co-occurring with an explanation is rarely discussed in the literature, except in Müller's study (2005); however, this co-occurrence is found in the speech of both the NNSs and NSs under investigation and it is one of the frequent types of co-occurrence of Type B *like*.

It is found that a larger proportion of Type B *like* in the NNSs' speech than in the NSs' co-occur with hesitation markers, pauses and restarts. This could indicate a search for content information or lexical words. In Müller's study (2005: 209-210), more than half the American NSs use *like* for the function of indicating a search for lexical words, as opposed to one fifth of the German NNSs and it is four times as frequently used in the NSs' speech as in the NNSs'. In this study, although the instances of Type B *like* co-occurring with hesitation markers, pauses and restarts in the NSs' speech are not as frequent as those in the NNSs' speech, the identification of this co-occurrence shows that *like* may also be used by the NSs to indicate a search for contents or lexis.

Table 4.15: Distribution of co-occurrence of *like* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions					
		Utterance- initial	Utterance- medial	Utterance- final		after an M-	after an MA	after an MF	others		
1. Hesitation markers; pauses; restarts	80.0		1 20.0			2 40.0		1 20.0			
2. Numerical expressions and locations	0										
3. Reported speech	0										
4. Expressions of uncertainty	0										
5. Expressions of certainty/ key point	0										
6. Exemplifications	20.0		1 20.0								
7. Explanations	0										
Unclassified	0										
Occurrences: 5 out of 300 (random samples)	100.0		2 40.0			2 40.0		1 20.0			

Table 4.16: Distribution of co-occurrence of *like* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions					
		Turn- initial	Turn- medial	Turn- final		after an M-	after an MA	after an MF	others		
1. Hesitation markers; pauses; restarts	55.6	2 22.2					1 11.1			2 22.2	
2. Numerical expressions and locations	0										
3. Reported speech	0										
4. Expressions of uncertainty	0										
5. Expressions of certainty/ key point	11.1		1 11.1								
6. Exemplifications	22.2					1 11.1				1 11.1	
7. Explanations	11.1		1 11.1								
Unclassified	0										
Occurrences: 9 out of 300 (random samples)	100.0	2 22.2	2 22.2			1 11.1	1 11.1			3 33.3	

Table 4.17: Distribution of co-occurrence of *like* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	6.7							1	6.7						
2. Numerical expressions and locations	0														
3. Reported speech	0														
4. Expressions of uncertainty	13.3							2	13.3						
5. Expressions of certainty/ key point	13.3							2	13.3						
6. Exemplifications	46.7			1	6.7			5	33.3					1	6.7
7. Explanations	13.3			1	6.7			1	6.7						
Unclassified	6.7			1	6.7										
Occurrences: 15 out of 284	100.0			3	20.0			10	73.3					1	6.7

Table 4.18: Distribution of co-occurrence of *like* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	15.1	3	1.7	9	5.2			5	2.9	1	0.6	7	4.1	1	0.6
2. Numerical expressions and locations	7.6			3	1.7			8	4.7					2	1.2
3. Reported speech	1.7			2	1.2									1	0.6
4. Expressions of uncertainty	22.7	3	1.7	12	7.0	1	0.6	16	9.3			1	0.6	6	3.5
5. Expressions of certainty/ key point	16.9			6	3.5			22	12.8			1	0.6		
6. Exemplifications	11.0	1	0.6	10	5.8			6	3.5			1	0.6	1	0.6
7. Explanations	12.2	1	0.6	9	5.2			11	6.4						
Unclassified	12.8	4	2.3	4	2.3	9	5.2	5	2.9						
Occurrences: 172 out of 300 (random samples)	100.0	12	7.0	55	32.0	10	5.8	73	42.4	1	0.6	10	5.8	11	6.4

Table 4.19: Distribution of co-occurrence of *like* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%
		Utterance- initial	Utterance- medial	Utterance- final		after an M-	after an MA	after an MF			
1. Hesitation markers; pauses; restarts	16.7									1	16.7
2. Numerical expressions and locations	0										
3. Reported speech	0										
4. Expressions of uncertainty	0										
5. Expressions of certainty/ key point	16.7									1	16.7
6. Exemplifications	33.3		2	33.3							
7. Explanations	33.3		2	33.3							
Unclassified	0										
Occurrences: 6 out of 235	100.0		4	66.7						2	33.3

Table 4.20: Distribution of co-occurrence of *like* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%		
		Turn- initial	Turn- medial	Turn- final		after an M-	after an MA	after an MF					
1. Hesitation markers; pauses; restarts	19.1		5	10.6		1	2.1		2	4.3	1	2.1	
2. Numerical expressions and locations	0												
3. Reported speech	17.0	1	2.1	1	2.1	5	10.6				1	2.1	
4. Expressions of uncertainty	21.3					6	12.8				4	8.5	
5. Expressions of certainty/ key point	10.6		1	2.1		4	8.5						
6. Exemplifications	8.5	1	2.1	2	4.3	1	2.1						
7. Explanations	19.1		6	12.8		2	4.3	1	2.1				
Unclassified	4.3	2	4.3										
Occurrences: 47 out of 300 (random samples)	100.0	4	8.5	15	31.9	19	40.4	1	2.1	2	4.3	6	12.8

4.5 Further investigation

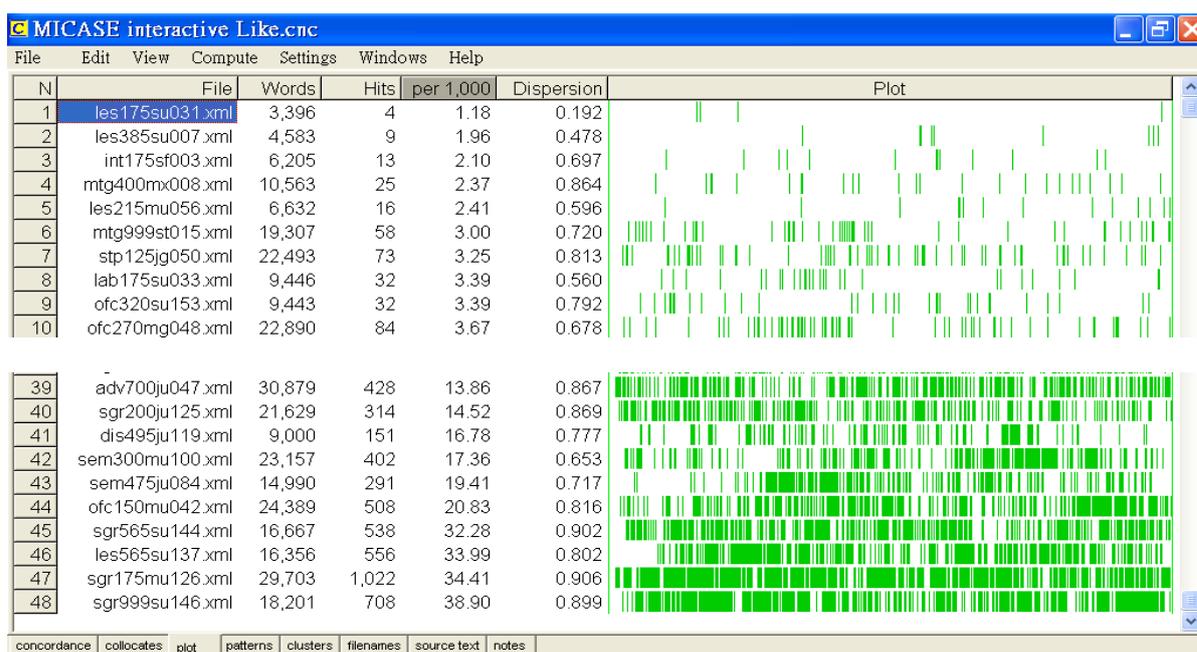
4.5.1 Variations in the use of like across texts and among speakers: Revisiting the sub-corpus of the American NSs' highly interactive discourse mode in MICASE

The instances of *like* in the NNSs' and NSs' speech are investigated with corpus methodologies. The co-occurrence of Type B *like* is identified in its context. Of the six sub-corpora, *like* is primarily used as a DM only in the sub-corpus of the NSs' highly interactive discourse mode in MICASE; therefore, this sub-corpus is chosen for further examination. (See Appendix 2 for the fact sheet of the sub-corpus of the highly interactive discourse mode in MICASE.)

The plot below produced by the *Concord of WordSmith 4* (Scott 2004) shows the raw counts (see *Hits* in Figure 4.2 below) of *like* in the 48 texts in this sub-corpus. The *Hits* range from 1.18 to 38.9 times per 1,000 words. The types of activity in the first ten texts (in ascending order of normalised frequency) with the fewest occurrences of *like* tend to be formal, such as lectures (abbreviated to *les* in the filename in Figure 4.2), interviews (*int*), meetings (*mtg*), student presentations (*stp*) and office hours (*ofc*), whereas those in the last ten texts tend to be less formal, such as study groups (*sgr*) and seminars (*sem*).

Figure 4.2: Plot of *like* in the sub-corpus of the NSs' highly interactive discourse mode in MICASE

(omitting the texts listed in the middle in order to show the wide differences between the first ten texts and the last ten texts)



The first text, LES175SU031, with the fewest occurrences of *like* and the last text, SGR999SU146, were manually examined. The four instances in the first text are used as a verb. The primary speaker in this text is a senior faculty member, who does not use Type B *like* in this event. However, it is found that he uses other DMs, such as *okay* and *now* as boundary markers, indicating a change of topic, *oh* prefacing a question and *well* prefacing a response to a question. Two possible interpretations of the speaker's non-use of Type B *like* are: first, that this lecture to a small class (40 students or fewer) is highly interactive by other means, such as frequent turn-changing between the speaker and students and, second, that Type B *like* has been found to be a popular DM among young Americans and therefore, this speaker, as a British NS in the group of 50-year-olds and older, does not use it in his speech.

By contrast, as presented in Figure 4.2 above, in the last text, SGR999SU146, *like* occurs 38.9 times per 1,000 words, which is almost 33 times as frequent in this text as in the first text, LES175SU031. The three senior undergraduates in the last text frequently use Type A *like* (highlighted in bold in Excerpt (4.5.1)) and Type B *like* (highlighted in bold and italics). They are using different kinds of *like*, in particular, *BE + like* for quoting (9 instances). Apparently, this use increases the total counts of *like*.

The conversation begins with Speaker 3's account of trying to get an extension for a paper from her advisor. It is reasonable to assume that the three speakers are speculating about what might happen; therefore, they may be using Type B *like* as a hedging device. It is also possible that the use of Type B *like* is one of the speech characteristics of young Americans, and could be a way of expressing solidarity among group members.

The use of DMs, discussed in Chapter 2, should facilitate the understanding of propositional meaning or interactional interpretations. However, in this excerpt, the highlighted stretches of utterance by the three speakers seem to hinder the progression of understanding, as the use of some instances of *like* cannot be objectively identified. It seems that the constant use of *like* is to construct a persona for a particular context. Speakers' construction of identities will be further explored in the text-based analyses in Chapter 11.

(4.5.1)

.....

S3: to see if she'll be around this summer and if she's not then, i don't know whether i can ask Frank is he gonna be around?

S2: well he said he's gonna be in Germany in May but

S1: <OVERLAP2> yeah </OVERLAP2>

S2: <OVERLAP1> he'll </OVERLAP1> be here in June and <OVERLAP1> i would </OVERLAP1>

S3: <OVERLAP2> okay </OVERLAP2>

S2: cuz all you need to do is just hand it in to him. he's just gonna <OVERLAP1>read it or whatever.</OVERLAP1>

S3: <OVERLAP1> well, i mean, </OVERLAP1> do you think it's okay if i wa- t- were to do a paper that didn't have an advisor? *like* she didn't end up reading it or
<EVENT WHO="SS" DESC="LAUGH" />

S2: well i don't_ i mean he asks me he asked me again last week if he was my advisor for my I-C-P.

S1: oh dear but i know when he had me do the senior thes- you know i mean he's technically should be my advisor for that too,

S3: mhm but, i think he was under the impression that my professor on North Campus was gonna read it too and i might have her read it_ well actually, my big news is, involves that she probably will read it but i don't think he would care.

S3: yeah mkay cuz *like*

S2: could y- could you email it to her or something?

S3: i was thinking about that. i'm_ the thing i'm most worried about is that i want a letter of recommendation from her and i feel *like* she's gonna *be like*, you know she didn't really turn in a very good paper for me last semester and now she's not even doing this one i only kn- i know she's only taking a couple classes *like*

S2: <OVERLAP2> mhm mhm </OVERLAP2>

S3: <OVERLAP1> why doesn't she have </OVERLAP1> *like*

S1: i don't think so i think that *like* she might feel that you're going *like* out of your way to do this special project

S3: <OVERLAP2> yeah </OVERLAP2>

S1: <OVERLAP1> or something. </OVERLAP1> and she might *be like*,

S3: yeah

S1: this <OVERLAP1>girl is motivated.</OVERLAP1>

S3: <OVERLAP1>i need to have,</OVERLAP1> i need to have Frank email her and just *like* i think she's too *like* she thinks it's more of a thing than it is *like* that. you know cuz i *was like* oh it's *like* twenty-five or thirty pages and she's *like* well how much *like* ah you know she wants *like*,

S2: <OVERLAP2>mm</OVERLAP2>

S3: <OVERLAP1>i don't know.</OVERLAP1>

S1: um she's not used to the R-C. i, i thought that yeah *like* i thought that Mabelle my advisor was gonna she was gonna read my paper and write my <OVERLAP1>evaluation. that's what i think.</OVERLAP1>

S3: <OVERLAP1>yeah, that's the other thing.</OVERLAP1> is i'm afraid that,

S2: oh <OVERLAP1> um </OVERLAP1>

S1: <OVERLAP1> that's what </OVERLAP1> Fra- that's actually what Frank told me.

S3: mkay <OVERLAP1> so </OVERLAP1> <OVERLAP1> yeah </OVERLAP1> see that's my other problem so, if <OVERLAP1> she needs </OVERLAP1>

S2: maybe you won't be able to yeah i might not be able to or i need to see if *like* someone else could *like* kinda just take it on right now or, if Frank if i could *be like like* oh Frank will you just write it? or something *like* that you know. uhuh mhm,

S3: yeah. well so i got the extension from him and then i finally just ended up talking to my professor on North Campus about it. and i told her what i was interested in doing, and then i *was like* could i use the research? and she *was like*_ cuz there's several different parcels of research *like*, there's *like* five different communities i think in Detroit that they did interviews with and *like* s- *like* (peer g-) the diff- there're different characteristics for all of 'em. some of 'em are *like* industrial areas some of 'em are retail areas, and so some of them a- um all of them are gonna be going into the research that she does at the end of the year and stuff but also *like* she went to *like* different neighborhood organizations *like*, maybe *like* um, employment organizations and things in the area and *was like* do you guys want some of the information that we get from this stuff? *like* cuz it would probably be

S3: <OVERLAP2>mhm</OVERLAP2>

S1: <OVERLAP2>yeah</OVERLAP2>

S3: <OVERLAP1>helpful to</OVERLAP1> them. so that's what i've been working on a lot for her last year was doing that for a couple um *like* uh, industrial organizations around the area and some of them haven't been done at all. so she wanted me to do some of those um for, my project. so then <EVENT WHO="SS" DESC="LAUGH" /> so then i would do them and i would put that in my project. *like* i would have a part that *was like* this is empowerment zones in Detroit.

S1: yeah this is um, *like* what local labor market theory is and then *like* this is one specific thing so i would on- also get i would paid for it. <EVENT DESC="LAUGH" /> <OVERLAP1>which is kind of

cool.</OVERLAP1>
 S3: <OVERLAP1>that's nice.</OVERLAP1>
 S1: <OVERLAP1>i think that that,</OVERLAP1> that would be good because *like*, we're supposed to have
like a other aspect besides
 S2: <OVERLAP2>mhm</OVERLAP2>
 S1: <OVERLAP1>*like*</OVERLAP1> the research in our project. <OVERLAP1>(we'll *like*) do something
 else</OVERLAP1>
 S3: <OVERLAP1>mhm yeah</OVERLAP1>
 S2: <OVERLAP1>yeah</OVERLAP1> that's what he's been saying a lot is
 S3: <OVERLAP2>right</OVERLAP2>
 S2: <OVERLAP1>that *like*</OVERLAP1> it's gotta be *like* h- kinda hands-on or something
 S3: <OVERLAP2>right</OVERLAP2>
 S2: <OVERLAP1>*like* that.</OVERLAP1> so,

(MICASE: SGR999SU146)

4.5.2 Analysis of *just like*

In Section 4.3.2 above, the patterns of the six subsets of SECCL, MICASE and ICE-GB indicate that *just* is one of the most frequent collocates immediately to the left. When *just like* is preceded by the verb *BE*, *like* is very likely to be Type A. In some cases, the subject and verb before *just like* are omitted, as in Excerpts (4.5.2) and (4.5.3). This kind of ellipsis is one of the main features of spoken English.

(4.5.2)

.....In a word, Miss Li is unusual teacher in my eyes, for she is not just a teacher, but also a friend
 <friends> of mine,... eh... ... |**just like**| this word a friend in need is a friend indeed.....

(SECCL: B01-50-12)

(4.5.3)

now consistent with that, argument, are just a few things, first is that, if you look at the facial
 neuro-muscular mechanisms, fun to say that three times they show continuity from higher primates to
 man. all that means is that again |**just like**| you wanna see continuity, in the, in the uh the b- evolution of
 the brain,

(MICASE: LEL500JU034)

To investigate if *like* co-occurring with *just* tends to belong to Type A, the two spoken sections, *brspok* and *usspok*, of the Bank of English⁹ are used to search for *just + like*. In British spoken English, there are 1,970 instances of *just like* and in American spoken English 99 instances. In Table 4.21, the highlighted collocates show that *just like* is Type A. The highlighted collocates immediately to the left are paradigmatic inflections of verb *BE*, combining with the collocates immediately to the right, *that*, *it* and *this*, could suggest that

⁹ “The data comes from the Bank of English corpus jointly owned by HarperCollins Publishers and the University of Birmingham. In 2008 the corpus stands at 450 million words.”
[\(http://www.titania.bham.ac.uk/docs/\)](http://www.titania.bham.ac.uk/docs/)

like is used as a preposition. The collocates to the right, *to* and *being*, indicate that *like* is being used as a verb.

Table 4.21: Concordance for *just like* in the Bank of English (Accessed on 10 August 2009)

and	it	s	NODE	like	to	say
it	i	d	NODE		a	know
well	you	was	NODE		that	you
i	and	i	NODE		you	and
but	they	would	NODE		the	a
er	that	re	NODE		i	i
erm	he	they	NODE		it	the
know	we	it	NODE		er	it
so	yeah	is	NODE		this	to
like	the	you	NODE		erm	of
that	she	and	NODE		in	that
the	just	were	NODE		we	in
just	like	know	NODE		they	s
mean	know	we	NODE		any	just
you	s	be	NODE		being	out
yeah	a	that	NODE		if	make
they	erm	he	NODE		an	like
to	but	she	NODE		your	yeah
or	was	like	NODE		oh	on
now	is	t	NODE		one	erm
a	to	are	NODE		me	have
of	so	yeah	NODE		when	little
t	mm	not	NODE		get	other

"s". Tot freq:361827. Freq as coll:337. t-sc:16.5125. MI:3.3150. '?' for help

The patterns of the NNSs' speech (see Tables 4.2 and 4.3 above) show that *just* is a strong immediately left collocate. In some of the instances of *just like*, it is difficult to distinguish between Types A and B. The instances of *just like* in Excerpts (4.5.2) and (4.5.3) above are classified into Type A *like*. The alternative is defensible. They can also be taken as Type B *like*. Based on the investigation of the NSs' use of *just like* in the Bank of English, it is probable that the NNSs under investigation use *just like* not preceding the subject and verb. Such use also shows that ellipsis is used in the NNSs' speech.

4.6 Chapter summary and conclusions

The above analysis of *like* supports my hypothesis that the use made of DMs by the Chinese NNSs is different from the NSs in terms of frequency and function. This chapter reveals that, in the NNSs' speech under investigation, most of the instances of *like* show it used as a verb and a preposition. Very few instances (5 and 9 out of 300 instances of random samples in the NNSs' monologues and dialogues respectively) of Type B *like* are used. Most of these instances co-occur with hesitation markers, pauses and restarts. Due to the low frequency

counts of Type B *like* and the collocation phenomena, it is probably the case that the NNSs do not know how to use *like* as a DM. Another possibility is that the NNSs know how to use it but choose not use it in the test-taking context. Since the NNSs do use other DMs, such as *oh*, *well* and *you know*, in their speech, the first possibility is more likely to be true.

To confirm whether Chinese NNSs do not use *like* as a DM, two corpora are used. One is the second version of the SECCL corpus (Wen, Liang and Yen 2008), which was published after I embarked on my project. This corpus consists of more recent data collected between 2003 and 2007. The other corpus is the College Learners' Spoken English Corpus (COLSEC) (Yang and Wei 2005). A random sample of 100 concordance lines of *like* was extracted from the Chinese NNSs' monologues in SECCL, version 2. Of the 100 instances, 48 were used as a verb, 44 as a preposition, 1 as a conjunction and 7 as a DM. Another random sample of the same size was taken from COLSEC. 81 out of 100 instances of *like* were used as a verb, 15 as a preposition, 3 as a DM and 1 was unclassified. These figures show a similar pattern to my analysis of *like* in SECCL, version 1. *Like* as a DM is seldom used by the Chinese NNSs. The quick searches in another two corpora of the speech of Chinese NNSs help to increase the reliability of my findings.

In speech by the NNSs, the instances of Type B *like* are too few to support any conclusions. The percentage information in Tables 4.15 and 4.16 above is of little importance, because it is calculated from the incidence, 5 and 9, found in the random samples. Therefore, the analyses of the positions of Type B *like* in an utterance/turn and the co-occurrence of *like* contribute more to the use of *like* made by the NSs in MICASE and ICE-GB.

The speech by NSs shows clearly that Type B *like* occurs more often in the dialogic genres than in the monologic genres. It can reasonably be argued that *like* as a DM is more frequently used in the contexts of higher interactivity or with more speakers involved. In the sub-corpus of the highly interactive discourse mode in MICASE, *like* is primarily used as a DM, accounting for 57.3% of the instances.

From the immediately left collocates displayed in the patterns of the node word *like*, it is found that Type B *like* in MICASE may co-occur with such DMs as *and*, *I mean* and *but* and that in ICE-GB it tends to co-occur with vague language *sort of*.

Generally, across the four NS sub-corpora, Type B *like* occurs more often in intra-clausal positions, usually after an M- element. The percentage information is highlighted in bold in Tables 4.17 to 4.20 above. Seven types of co-occurrence of Type B *like* are 1) hesitation

markers, pauses and restarts, 2) numerical expressions and locations, 3) reported speech, 4) expressions of uncertainty, 5) expressions of certainty/key points, 6) exemplifications and 7) explanations. The functions of *like* may be adumbrated on the basis of co-occurring linguistic evidence and it is argued that varying interpretations of functions appear. For example, when *like* precedes numerical expressions, some researchers may take it as an approximator but other may see it as a focuser. In the present study, the collocation phenomena are used to decide the categories for discussion and functions of *like* are secondary interpretations.

Further investigation of *like* in the sub-corpus of the highly interactive discourse mode in MICASE shows variations in frequency across texts. Similarly, Müller's comparative study (2005) reveals variations in the frequency of functions among speakers. These indicate that more detailed analysis should accompany frequency information.

The analysis of *just like* has shown that when *like* follows *just*, it is most probable that *like* is functioning as a preposition. Some of the instances of *just like* in the NNSs' speech are not preceded by a subject and a verb. It can be concluded that the NNSs are prone to omit subject and verb, which is one of the features of spoken English.

This chapter has identified the Chinese NNSs' lack of familiarity with NS usages of Type B *like*. This raises two wider questions: 1) whether it matters that the NNSs do not know how to use Type B *like*; and 2) whether the NS speech should be taken as an appropriate target norm for the use of DMs. These questions are further discussed in Chapter 12.

CHAPTER 5: ANALYSIS OF *OH*

5.1 Introduction

The hypotheses in the use of *oh* are set out first, followed by a review of the literature. As stated in Chapter 3, a bottom-up approach is employed. The analysis presents the frequency information and patterns of *oh*, showing the overall use of *oh*. The major analysis is the discourse aspects of *oh*, looking at its positions in utterances/turns and the collocation phenomena surrounding *oh*. The identification of co-occurrence leads to the interpretations of the functions of *oh*.

In my analysis of *like* in Chapter 4, the Chinese NNSs tend to use Type A *like* and do not use Type B *like* as the NSs do. In this chapter, I investigate the word *oh*, which is different from the word *like*. *Oh* does not have any semantic meaning and does not belong to any grammatical word class. From the perspective of spoken interaction, *oh* is categorised as “only interactional”, referring to “lexical items that cannot be described as clause elements” (Stenström 1994: 208). In other words, *oh* is used as a DM (Type B in this thesis) only and therefore there is no need to distinguish between Types A and B.

Furthermore, it is unlikely that *oh* has formed part of the curriculum in the Chinese NNSs’ English language learning. I hypothesise that *oh* is not used by the NNSs as frequently as it is by the NSs. If this is not so, then it is hypothesised that the use of *oh* by the NNSs is different from that by the NSs. I aim to find how similar or different in the speech of the NNSs and NSs under investigation is the use of *oh*.

I tested my hypotheses within the framework of the core research questions addressed in this thesis (see Section 1.1.2). The answer to Question 1 below provides the frequencies of *oh* in the speech of the NNSs and NSs. The answers to Questions 2 to 4 show the overall use of *oh* while validating some claims about the use of *oh*, which are based on types of co-occurrence and contextual information (answering Question 5).

1. What are the frequencies of *oh* in the speech of the NNSs and NSs?
2. What do collocates of *oh* reveal about its use?
3. What other DMs does *oh* co-occur with?
4. Where does *oh* appear in an utterance/turn?
5. With what co-occurrence or in what contexts does *oh* tend to occur?

5.2 Previous studies of *oh*

The use of *oh* is first reviewed from its grammatical aspect and then described in the *Linear Unit Grammar (LUG)* analysis (Sinclair and Mauranen 2006). The literature survey reveals that it has been extensively investigated and taken as a marker of change-of-state.

5.2.1 Word class of *oh*

The word *oh* does not fit into any conventional word class. Grammar books and dictionaries assign *oh* to the category of *interjection*, which shows speakers' emotions (Biber *et al.* 1999: 1083, *Longman Dictionary of Contemporary English* 2009: 1212). *Cambridge Grammar of English* (Carter and McCarthy 2006: 115) draws a distinction between *interjection* and DM. *Oh* as an *interjection* expresses emotions, such as surprise, disappointment and pain, while *oh* as a DM is used to show that new or surprising information has been received. The *Collins COBUILD* dictionary (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 994) assigns *oh* to the category of *convention*, which is particularly demonstrated in conversations. In the above grammar books and dictionaries, the use of *oh* is covered, but the semantic meaning of *oh* is left undefined.

5.2.2 Syntactical structure of *oh*

The syntactical status of *oh* is ambiguous. It can be seen as either a single unit or part of the following unit. The prosodic information and punctuation in transcriptions may be used to help segmentation (Biber *et al.* 1999: 1076). In Excerpt (5.2.1) below, *oh* is a separate unit because there is a brief pause between *oh* and *really* in the recording and this affects the transcribers' use of the comma.

(5.2.1)

Task 3

A: <Chinese> Have you ever seen Peter recently? He is going to leave high school this year and he is preparing to go abroad for further education.

B: **Oh**, really? I don't think it's a good idea for him to go abroad <abroads> when he is just high school graduate. He should finish college education first, I think.

(SECCL: C01-123-29)

As discussed above, *oh* does not belong to any conventional word class. In any case, it is problematic to describe spoken language with syntactical structures based on written English.

In Excerpt (5.2.2), it is difficult to describe the syntactical structure in which *oh* occurs. It seems to be acceptable to say that *oh* occurs between an incomplete and a complete clause.

(5.2.2)

<OVERLAP1> yeah </OVERLAP1> like i'd like to spend um, a good portion of the paper talking about, like why i should interview them and like what makes them similar. because i mean it is_ you wanna make sure that there's like enough to be like **oh** you're not just picking random businesses.

(MICASE: SGR999SU146)

The existing grammars seem not to be able to describe spoken English satisfactorily. Therefore, the *LUG* analysis (Sinclair and Mauranen 2006) is adopted to assign units and describe where *oh* occurs in an utterance. In the same example, *oh* occurs after a message revision (MR) element, *enough to be like*, which is the revision of the partial completion of message unit (+M-), *there's like*, and is followed by a completion of message unit (+M), *you're not just picking random businesses* (see Appendix 4 for the labels used in *LUG*).

<OVERLAP1> yeah </OVERLAP1> like i'd like to spend um, a good portion of the paper talking about,
 OI MF M- +M- OI +M- +M-
like why i should interview them and like what makes them similar. because i mean it is_
 OI +M OT OI M OT OI MF
you wanna make sure that there's like enough to be like **oh** you're not just picking random businesses.
 +M- OT +M- MR OI +M

(MICASE: SGR999SU146)

5.2.3 The use of *oh* in spoken English

The use of DMs is a typical feature in informal conversations (McCarthy 1998: 59). *Oh* is one of the words which occurs significantly more frequently in the spoken mode of English than the written. In previous studies of conversations in English (e.g. Jucker and Smith (1998: 176) and Aijmer (2002: 105)), *oh* is identified as among the most frequent words.

It is traditionally accepted that *oh* is an interjection or exclamation used by speakers to express emotions, such as surprise, pain and disappointment (Schiffrin 1987: 73, Biber *et al.* 1999: 1083, 1096, Carter and McCarthy 2006: 115). Aijmer (2002: 103-104), however, maintains that the use of *oh* is relevant to but not restricted to the meanings (the expressions of feelings such as surprise, disappointment or happiness) of *oh* as an interjection. She lists two core functions of *oh* as a backwards-looking particle: 1) as a reaction on the textual level and 2) as an intensifier on the interpersonal level. On the basis of the collocations, positions, the prosody of *oh* and text types in which *oh* occurs, Aijmer (2002) concludes that *oh*, in most cases, functions as an “attention-getting or intensifying” device for emphasising what follows.

Additionally, *oh* has the property of multifunctionality, functioning as a marker of response, stance and intensification.

As a DM, it is noted for introducing utterances and for responding to new information (Biber *et al.* 1999: 1083, 1096, Carter and McCarthy 2006: 115). *Oh* often co-occurs with expletives, such as *oh my goodness* and *oh my god* and with other DMs, such as *oh well* and *oh yes* (Carter and McCarthy 2006: 115-116). These DM collocations are found also in the NNSs' and NSs' speech under investigation. More details are given in Section 5.3.2.

5.2.4 *Oh* as a marker of change-of-state

In general, researchers in the previous studies have struggled to generalise the meanings or functions of this particle from an apparently heterogeneous range of uses. Heritage (1984: 300) argues that *oh* is used as a marker of change-of-state which can be found in two environments: “(counter)informing” (e.g. Example (1) in Table 5.1) and “repairs” (e.g. Example (2) in Table 5.1). He identifies *oh* as used as a receipt token and in the sequences of “question-answer-*oh* receipt” and “repair-initiation-repair-*oh* receipt”.

Similarly, Schiffrin (1987: 74) argues that *oh* is a marker of information management and is used when speakers shift their orientation to information. She maintains that *oh* “marks a focus of speaker’s attention” (1987: 99) as well as listener’s attention and therefore, *oh* is a marker of information state transitions. Schiffrin (1987) analyses *oh* in her theoretical framework and claims that *oh* is a marker of “information management tasks”, such as the replacement of information (including different types of repairs) (e.g. Example (3) in Table 5.1), the completion of the proposition in question-answer-acknowledgement sequences (e.g. Example (4) in Table 5.1) and the recognition of the relevance of old information and the receipt of new information (e.g. Example (5) in Table 5.1).

Table 5.1: Examples of the use of *oh* in the literature

Researcher(s)	Example	Classification
1. Heritage (1984: 302)	A: ... Well lately in the morning Rosemary's been picking me up.-Yihknow so I (haven' been) even takin' a train in [(the morning) B: [hh Oh that's great.	Speaker B's <i>oh</i> as a receipt token + assessment to Speaker A's informing
2. Heritage (1984: 316)	A: Well who'r you workin' for. B: 'hhh Well I'm working through the Amfat Corporation. A: The who? B: Amfah Corpora[tion. T's a holding company. A: [Oh A: Yeah	Speaker A's <i>oh</i> as a receipt token to Speaker B's repair by repetition
3. Schiffrin (1987: 75)	Jack: I think it was in seventeen: fifteen, or seventeen fifty five, I'm not sure when. Eh: oh I'm wrong. Seventeen seventeen.	Replacement of information (<i>oh</i> indicates self-initiated repair)
4. Schiffrin (1987: 86)	Irene: How can I get an appointment t'go down there t'bring my son on a tour? Debby: Oh I didn't even know they gave tours! I'm not the one t'ask about it.	Completion of proposition (<i>oh</i> indicates Debby's receipt of new information and re-orientates Irene's assumption of Debby's knowledge)
5. Schiffrin (1987: 92)	Freda: Sometimes he got a notice for staying out past curfew. Recently. In August, that was. Val: Oh curfew? What's curfew? Freda: A certain time that children have to be in. Val: Oh your children. Oh I see. Oh it's personal. Oh I-- I thought there might be police or something.	Marking the receipt of new information

Fox Tree and Schrock (1999: 282) point out that because of the function of marking change-of-state, *oh* tends to co-occur with repairs, answers to questions and reported speech. When repairing a previous utterance, the speaker may use *oh* to foreshadow a change-of-state. When receiving answers to questions for updating knowledge, the inquirer, rather than using *yes*, may use *oh* to show the receipt of new information, which can express listenership and maintain the floor. When reporting others people's speech, the speaker may use *oh* to indicate the change of the utterance to the state of the person being quoted. This last context is further discussed in the next section.

The function of *oh* as a marker of change-of-state is mainly discussed on the basis of researchers' intuitive interpretation of the contexts in which *oh* occurs. These interpretations of the use of *oh* are more likely to be true as our communication experience and the command of the language increase. Nevertheless, whether or not the speaker has undergone a mental process is unknown to others.

In contrast to Heritage (1984) and Schiffrin (1987), Fox Tree and Schrock (1999) conducted stimulus-response experiments, which indicate that listeners more quickly recognise a given word after they hear *oh* faster than when they do not hear it. The researchers inferred from the results that the change of state implied by *oh* helped listeners to respond

quicker. Their investigation of *oh* concludes that *oh* may facilitate the integration process of discourse in that it marks the non-relationship between the preceding and following utterances and by so doing signals a change of state.

5.2.5 *Oh* for marking reported speech

Oh for marking reported speech is a particular case of change-of-state in that it is particularly frequent and is therefore worthy of specific comment in a separate section.

It seems to be a transcription convention that *oh* is placed within quotation marks. However, it is unknown that whether *oh* is produced by the speaker or the person being quoted by the speaker. Differing views are taken in the relevant literature. Biber *et al.* (1999: 1118) take *oh* as the beginning of quoted speech. Nevertheless, some writers find it open to doubt. Fox Tree and Schrock (1999) point out that in the literature (e.g. Schourup (1985) and Wade and Clark (1993)) it is accepted that the direct quotations in speech rarely copy exactly the speech being quoted. To argue this point, researchers in this area use different terms, for example, “constructed dialogue” in Tannen’s investigation (1989, 2007) of casual conversations and “speech, writing and thought presentation” by Semino and Short (2004).

Tannen (1989, 2007: 103) believes that an utterance is seldom repeated in exactly identical wording and manner by another speaker in a different context. Therefore, she finds it appropriate to use the term “constructed dialogue” to refer to direct quotation and reported speech, which is created by the speaker, rather than the person being quoted.

Semino and Short (2004) conduct a corpus study of reported speech in written English and discuss “speech, writing and thought presentation”. Their interest is mainly in the discourse being *presented* rather than its being *reported* or *represented* (which is often claimed by grammarians and critical discourse analysts), since they believe that it is unlikely that someone else’s speech can be faithfully (re)presented.

Similarly, Trester (2009) claims that *oh* functions as a signal of the shift in “footing”¹⁰ (Goffman 1981) (marking the boundary between the speaker’s own utterance and speech being quoted) as well as conveying the speaker’s stance towards the quoted speech. She argues that *oh* contributes to two levels of interaction: interaction between the speaker and other participants in the interview and interaction between the speaker and the contexts and

¹⁰ The word *footing* in speech was coined by Goffman in 1981. It was defined as “significant shifts in alignment of speaker to hearers” (1981: 127). Shifts of footing are common in spontaneous speech. Shifting from reporting our current self, i.e. “addressing self”, to others can be marked para-linguistically or by code-switching.

characters from which the speaker refers to. Another interesting finding by Trester (2009: 154) is that there is great variation among the seven speakers whom she studied in their use of reported speech and the use of *oh* prefacing reported speech. She argues that *oh* is used to signal the speaker's negative stance towards the reported speech. She gives examples of this use and points out that this is probably because the topics and contexts of her data. She interviews performers in improvisational theatre. When using reported speech, the speaker could be distant from the speaker being quoted and this is particularly useful when evaluating negatively.

In my data, I have not found this use. The NNSs under investigation frequently use *oh* for marking reported speech, in particular in their narrative accounts. Since the reported speech is originally spoken in Chinese, it is likely that *oh* is added by the speaker rather than being spoken by the speaker being quoted. If the speaker's stance is attached, it refers to the speaker's interpretation of the attitudes or emotions conveyed by the reported speech.

5.3 Frequency information in the speech of the non-native speakers and native speakers

5.3.1 Overall frequency of *oh*

The overall frequency of *oh* is shown in Table 5.2 below. There are 397 and 2,882 occurrences of *oh* in the Chinese NNSs' monologues and dialogues respectively. The raw counts are normed on a basis of 10,000 words. The normalised frequencies show that *oh* is used more than four times as often in the dialogues as in the monologues. In the monologues, there are 11.8 instances of *oh* per 10,000 words and in the dialogues, 48.3 instances. This supports my hypothesis that the more interactive the genres or types of activity are, the more often DMs occur. The presence of interlocutors and the nature of interactivity in the dialogues are likely to be the reasons why there are more instances of *oh* in the dialogues than in the monologues.

Table 5.2 below presents the frequencies of *oh* across the types of genre and between the speech of the NNSs and NSs. It can be clearly seen in Figure 5.1 that in terms of genre, there are apparently more instances of *oh* in the dialogic genres than in the monologic genres. In terms of the speaker, the NNSs use many more instances of *oh* in their monologues than the NSs do. This challenges my assumption, made at the beginning of this chapter, that the NNSs do not use DMs as frequently as the NSs do.

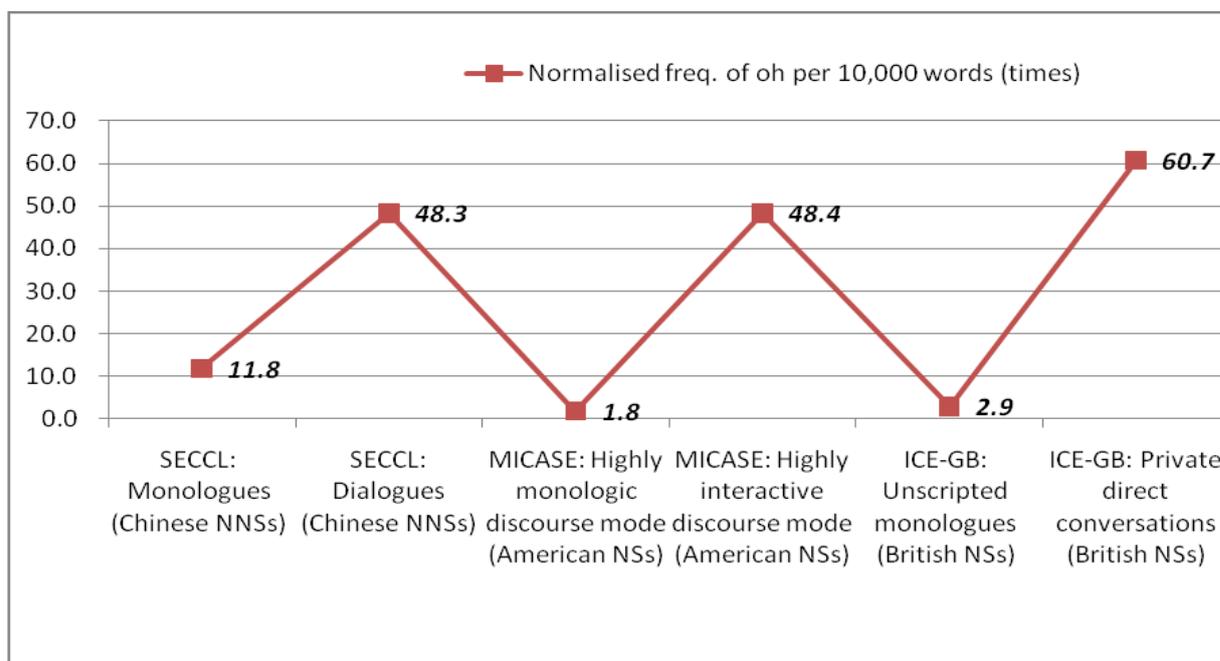
Interestingly, comparing the NNSs' dialogues with the NSs' highly interactive speech in MICASE, *oh* occurs with similar frequencies. By contrast, *oh* occurs more frequently in the private direct conversations in ICE-GB than in the highly interactive texts in MICASE. It seems that genre is a crucial factor in the frequency of *oh*.

Table 5.2: Frequency information of *oh* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)
SECCL: 1,143 monologues (Chinese NNSs)	336,303	397	11.8
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	2,882	48.3
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	24	1.8
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	2,795	48.4
ICE-GB: 70 unscripted monologues (British NSs)	153,646	45	2.9
ICE-GB: 90 private direct conversations (British NSs)	185,000	1,123	60.7

* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

Figure 5.1: Comparison of normalised frequencies of *oh* across sub-corpora



Although it is clearly evident in Figure 5.1 that *oh* occurs more frequently in the dialogic genres than in the monologic genres and it occurs more frequently in the sub-corpus of the Chinese NNSs' monologues than in the other two sub-corpora of the NSs' monologic genres, the tests of statistical significance discussed in Section 3.3.3 of Chapter 3 were undertaken to

support these findings.

The results of the LL test are presented in Appendix 6. The LL scores show that there is a statistically significant difference between the two types of genre in SECCL (LL: -966.54, p-value: < 0.0001), MICASE (LL:-1882, p-value: < 0.0001) and ICE-GB (LL:-132.1, p-value: < 0.0001). The negative LL scores indicate that Type B *like* is under-represented in the monologic genres. Between the two groups of speakers, the difference in the monologic genres between the Chinese NNSs and NSs is significant (LL: +142.58, p-value: < 0.0001 between Corpora A1 and B1 and LL: +112.27, p-value: < 0.0001 between Corpora A1 and C1). In the dialogic genres, the difference is not statistically significant between SECCL and MICASE (LL: 0), but it is significant between SECCL and ICE-GB (LL: -40.74, p-value: < 0.0001).

5.3.2 Collocates of *oh*

In Tables 5.3 and 5.4 patterns of *oh* in the NNSs' monologues and dialogues, it can be seen that there is a clear distinction between the two types of genre. In Table 5.3, the collocates to the left of the node word *oh* constitute patterns relating to reported speech, such as *SAY + oh*, *SAY to + PRONOUN + oh*, *SAY/ASK (VERB) + that + oh* and *TELL + PRONOUN + oh*. To the right of the node word *oh*, a higher proportion of the collocates seems to form expressions of emotions, such as *oh my god*, *oh my dear*, *oh dear* and *oh no*¹¹.

¹¹ Some researchers, such as Trester (2009), exclude the instances of *oh* in such constructions as *oh yeah*, *oh god* and *oh well*. It is argued in this thesis that they can be used separately.

Table 5.3: Pattern of *oh* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (18)	i (16)	to (25)	said (68)	oh	i (73)	i (36)	i (31)	i (18)
2	and (17)	said (15)	i (24)	that (21)	(397)	it (28)	was (18)	you (18)	you (16)
3	to (15)	and (15)	and (24)	me (20)		my (28)	god (17)	t (11)	my (13)
4	a (13)	the (12)	said (19)	it (13)		you (17)	is (13)	that (11)	the (12)
5	the (12)	to (11)	he (18)	him (8)		it's (10)	s (13)	a (11)	was (10)
6	very (10)	he (9)	the (16)	oh (8)		no (9)	you (11)	so (10)	it (8)
7	my (9)	me (9)	my (12)	you (7)		the (8)	my (10)	very (10)	and (7)
8	said (8)	she (9)	a (12)	say (6)		oh (8)	dear (9)	to (9)	he (7)
9	was (8)	in (8)	you (8)	eh (5)		how (7)	the (9)	was (9)	a (7)
10	me (7)	my (8)	of (7)	birthday (5)		sorry (7)	a (7)	my (8)	are (6)
11	at (7)	of (7)	told (7)	2 (5)		he (7)	don (6)	it (7)	very (6)
12	he (6)	a (6)	me (6)	think (4)		she (6)	eh (6)	and (6)	to (6)
13	um (5)	very (6)	she (5)	her (4)		dear (6)	t (5)	the (6)	sorry (5)
14	his (4)	us (5)	at (5)	thought (4)		we (5)	know (5)	sorry (5)	is (5)
15	in (4)	just (4)	his (5)	um (4)		when (5)	m (5)	not (5)	that (5)
16	that (4)	one (4)	very (4)	and (4)		what (5)	have (5)	me (5)	do (5)
17	one (4)	um (4)	was (4)	time (4)		eh (5)	are (5)	do (5)	in (4)
18	no (4)	birthday (4)	with (4)	know (3)		mr (4)	it (4)	is (4)	me (4)
19	her (4)	you (4)	asked (4)	school (3)		um (4)	we (4)	it's (4)	know (4)
20	of (3)	have (4)	were (4)	shouted (3)		is (4)	did (4)	come (4)	said (4)

In Table 5.4, it is clear that 85% (2,452) of the 2,878 instances of *oh* in the NNSs' dialogues are used to open a turn. The two speakers in the dialogues are referred to as *a* (1,204 turns) and *b* (1,248 turns) in Table 5.4. In addition, the collocates to the right include such words and phrases as *yes*, *yeah*, *really*, *well* and *I think*, which seem to be used together as DM collocations. These DM collocations are similar to those used by NSs, such as *oh yeah*, *oh yes*, *oh no*, *oh aye*, *oh well*, *oh God*, *oh I see* and *oh right* (Biber *et al.* 1999: 1083).

Interestingly, *a* (82) and *b* (65) are the two immediate collocates to the right, which means that the node word *oh* is likely to be used alone as a single utterance. This represents 5% (147) of the 2,878 instances of *oh*.

Table 5.4: Pattern of *oh* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (138)	the (171)	you (140)	b (1248)	oh	<i>i</i> (611)	<i>i</i> (338)	<i>i</i> (248)	i (212)
2	you (123)	a (132)	it (96)	a (1204)	(2878)	yes (275)	you (229)	you (180)	you (172)
3	to (121)	you (105)	a (64)	oh (40)		you (181)	think (206)	think (151)	a (123)
4	the (109)	to (96)	yes (60)	eh (21)		yeah (119)	a (126)	a (132)	think (111)
5	in (102)	b (95)	b (52)	yes (16)		thank (92)	know (96)	the (67)	it (67)
6	b (58)	about (76)	yeah (50)	think (15)		really (88)	s (94)	know (62)	the (66)
7	is (51)	your (73)	abroad (45)	um (13)		a (82)	b (88)	t (57)	to (62)
8	i (47)	of (65)	so (43)	yeah (12)		it (78)	but (68)	it (55)	b (60)
9	of (46)	i (58)	know (42)	you (11)		b (65)	is (52)	eh (55)	eh (54)
10	for (42)	very (55)	advice (39)	so (10)		eh (63)	see (50)	b (53)	so (53)
11	think (41)	my (52)	venture (36)	know (8)		that (59)	don (47)	that (48)	that (51)
12	and (39)	for (49)	me (36)	that (8)		but (56)	eh (43)	have (48)	is (42)
13	about (38)	some (47)	study (35)	but (7)		what (50)	don't (43)	yeah (46)	have (38)
14	do (33)	think (43)	3 (35)	l (7)		i'm (40)	and (41)	but (43)	are (36)
15	your (30)	joint (41)	future (32)	it (6)		oh (40)	are (38)	um (42)	but (36)
16	are (30)	in (38)	university (32)	and (6)		um (37)	have (36)	so (40)	um (33)
17	my (29)	good (37)	think (31)	i (5)		no (35)	the (35)	very (38)	very (31)
18	me (29)	with (37)	life (30)	study (4)		maybe (33)	it (34)	and (31)	know (31)
19	what (28)	go (36)	them (29)	job (4)		it's (31)	m (32)	yes (30)	s (29)
20	have (26)	is (32)	do (29)	see (4)		well (28)	um (31)	is (30)	about (28)

Among the NS corpora under investigation, the sub-corpora of the highly monologic discourse mode in MICASE and the unscripted monologues in ICE-GB, shown in Tables 5.5 and 5.6, reveal a small number of occurrences of *oh*. The collocates immediately to the left, *say* in Table 5.5 and *said* in Table 5.6, are of importance, which are also the prominent collocates immediately to the left in the NNSs' monologues in SECCL.

Table 5.5: Pattern of *oh* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	if (2)	you (2)	you (3)	say (4)	oh	that's (3)	this (2)	they (2)	to (2)
2		and (2)			(22)	yeah (2)	know (2)		i (2)
3		a (2)				you (2)	and (2)		
4						my (2)			
5						the (2)			

Table 5.6: Pattern of *oh* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (4)	to (5)	i (3)	said (4)	oh	that (8)	s (12)	a (7)	a (3)
2	good (3)	s (2)	to (3)	and (3)	(44)	i (4)	was (2)	i (3)	to (2)
3	the (2)	and (2)	for (2)	about (3)		he (4)	a (2)	to (2)	superb (2)
4	that (2)	from (2)	the (2)	uhm (2)		no (3)	i (2)	that (2)	
5	m (2)		forward (2)	uh (2)		so (2)			
6				gregory (2)		a (2)			
7						it (2)			
8						about (2)			

In Table 5.7 below, it can be seen that the most frequent collocates to the left of *oh* are *was*, *be* and *like*, suggesting the use of *BE + like* for quoting to mark the boundary of reported speech. In MICASE (Table 5.7) and ICE-GB (Table 5.8) the collocates immediately to the right include answers or responses to prior questions or statements, such as *okay*, *yes*, *yeah* and *no*. *Oh* also co-occurs with DMs *well* and *right*. In ICE-GB, *oh* co-occurs with *dear* and *god*, probably to express emotions.

Table 5.7: Pattern of *oh* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (20)	the (23)	was (23)	like (75)	oh	okay (400)	i (112)	i (78)	i (72)
2	you (14)	i (22)	be (18)	oh (27)	(1471)	yeah (255)	you (81)	the (69)	the (61)
3	the (11)	and (20)	the (16)	xx (25)		i (193)	the (65)	you (52)	you (35)
4	it (9)	it (12)	you (15)	uh (21)		no (93)	so (62)	that (46)	a (32)
5	i (8)	you (10)	and (15)	um (17)		you (89)	okay (59)	okay (42)	is (31)
6	this (8)	is (9)	it (12)	that (17)		that's (85)	that's (58)	a (40)	it (31)
7	have (7)	a (8)	i (11)	it (15)		so (78)	and (50)	it (36)	this (31)
8	that (7)	so (8)	i'm (10)	okay (14)		my (71)	is (47)	is (33)	that (28)
9	one (7)	for (7)	oh (10)	so (13)		oh (59)	yeah (46)	know (33)	have (24)
10	we (6)	to (7)	is (8)	yeah (13)		it's (57)	see (45)	yeah (29)	okay (24)
11	is (6)	oh (7)	this (7)	say (10)		well (51)	oh (45)	to (28)	to (24)
12	oh (5)	like (7)	a (7)	one (8)		really (48)	god (44)	this (27)	of (23)
13	of (5)	it's (7)	to (7)	go (7)		that (48)	don't (43)	it's (26)	it's (21)
14	a (5)	was (6)	like (6)	the (6)		and (42)	it's (37)	so (26)	yeah (20)
15	what (5)	that (5)	just (6)	there (6)		the (36)	it (36)	xx (25)	in (19)
16	there (4)	of (5)	that (6)	is (6)		it (36)	that (33)	have (25)	was (18)
17	like (4)	two (5)	so (6)	you (6)		this (33)	a (31)	be (25)	xx (17)
18	okay (4)	then (5)	one (5)	that's (5)		wait (26)	this (30)	what (24)	um (17)
19	so (4)	do (4)	right (5)	here (5)		i'm (25)	was (29)	just (24)	so (17)
20	just (4)	um (4)	of (5)	but (5)		we (24)	right (26)	don't (22)	just (17)

Table 5.8: Pattern of *oh* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (38)	s (32)	i (39)	yes (40)	oh	i (136)	i (94)	i (70)	i (48)
2	the (33)	the (28)	the (31)	it (38)	(1120)	yes (112)	s (81)	s (58)	it (39)
3	and (27)	you (27)	oh (30)	yeah (36)		right (63)	you (42)	you (43)	s (30)
4	s (26)	i (25)	it (28)	uhm (35)		yeah (62)	it (38)	it (30)	a (25)
5	i (23)	a (25)	a (27)	said (29)		no (53)	that (36)	oh (29)	that (24)
6	to (23)	and (24)	you (26)	mm (26)		well (51)	see (34)	the (25)	you (24)
7	that (20)	it (23)	of (26)	you (21)		that (51)	oh (30)	that (21)	the (21)
8	you (19)	oh (22)	s (23)	right (18)		dear (39)	yes (29)	no (21)	know (20)
9	in (18)	that (21)	to (18)	oh (18)		god (39)	yeah (28)	so (21)	to (20)
10	a (17)	in (20)	and (16)	no (17)		it (34)	and (20)	ve (18)	oh (17)
11	of (13)	to (19)	he (15)	that (15)		you (32)	well (19)	a (18)	yes (17)
12	was (12)	of (14)	yeah (13)	know (11)		oh (18)	no (18)	yes (17)	yeah (15)
13	oh (11)	was (14)	this (12)	s (11)		my (15)	is (18)	know (17)	no (15)
14	uh (11)	or (12)	that (12)	uh (10)		he (15)	know (17)	yeah (17)	have (13)
15	is (10)	is (12)	mm (12)	thought (10)		good (15)	god (17)	and (15)	not (13)
16	no (10)	so (11)	was (12)	think (10)		really (14)	was (17)	right (14)	is (13)
17	she (9)	just (10)	don't (12)	really (9)		we (12)	don't (16)	they (14)	right (12)
18	about (9)	what (10)	is (12)	is (9)		she (12)	right (14)	was (13)	of (11)
19	for (9)	uhm (9)	in (12)	to (8)		the (12)	uhm (13)	what (12)	so (11)
20	think (8)	yes (8)	as (11)	well (8)		what (12)	the (13)	don't (12)	good (11)

The frequency information and collocates of *oh* are used as starting points. The frequencies of *oh* in the six sub-corpora reveal that the NNSs do use *oh* as frequently as the NSs do. This contradicts my hypothesis that the NNSs seldom use DMs in their speech. The above analyses present the ways in which the NNSs and NSs commonly use *oh*. In the speech of the two groups of speakers, the salient collocates immediately to the left of *oh* are *SAY* (see Tables 5.3, 5.5 and 5.6) and *BE + like* (see Table 5.6), which suggest that *oh* is used to mark the boundary of reported speech. In the NNSs' dialogues, the speaker identities, *a* and *b*, are shown in the pattern, which reveal that the NNSs tend to use *oh* as a turn opener. This piece of information is annotated in MICASE and ICE-GB but cannot be seen in the patterns generated from the concordance lines of *oh*. It can be seen in the patterns of *oh* in MICASE and ICE-GB that *oh* frequently co-occurs with other DMs, such as *okay*, *yes*, *yeah*, *no* and *well*. The following sections further examine the use of *oh* in the NNSs' and NSs' speech.

5.4 Discourse aspects of *oh*

In this section, the positions where *oh* occurs in an utterance/turn are first described and then the linguistic items with which *oh* tends to co-occur are discussed. All the occurrences of *oh* in the NNSs' monologues (397), the NSs' highly monologic discourse mode in MICASE (24)

and the unscripted monologues in ICE-GB (45) were manually analysed, but the sheer number of occurrences in the NNSs' dialogues (2,882), the NSs' highly interactive discourse mode in MICASE (2,795) and the private direct conversations in ICE-GB (1,123) was unmanageable for manual analysis; instead, three sets of 100-line concordance samples from each sub-corpus were manually examined.

5.4.1 Positions in an utterance/turn

In this section the positions of *oh* in an utterance/turn are discussed. Its distribution and percentages in the six sub-corpora under investigation are shown in Table 5.9 below. There is a marked difference in the positions in an utterance/turn of *oh* across two types of genre, but not between the two groups of speakers. In the NNSs' monologues, the occurrences of *oh* are almost equally distributed in extra- (52.9%) and intra-clausal (47.1) positions. In the monologic genres in MICASE and ICE-GB, there is a slightly larger proportion of the occurrences of *oh* in the extra-clausal position, accounting for about two-thirds of the occurrences. In the three dialogic genres, over 94% of the occurrences are in the extra-clausal position and most of them are placed initially in each turn.

Table 5.9 Distribution of the positions of *oh* in an utterance/turn

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues		
	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	
Positions in an utterance of <i>oh</i>	397	100	24	100	45	100	
Extra-clausal: utterance-initial	7	1.8	9	37.5	3	6.7	
Extra-clausal: utterance-medial	203	51.1	7	29.2	26	57.8	64.4
Extra-clausal: utterance-final	0	0.0	0	0.0	0	0.0	
Intra-clausal: after an M-	132	33.2	8	33.3	11	24.4	
Intra-clausal: after an MA	10	2.5	0	0.0	0	0.0	
Intra-clausal: after an MF	6	1.5	0	0.0	3	6.7	
Intra-clausal: others	39	9.8	0	0.0	2	4.4	
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations		
	Random samples' (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples' (times)	Percentage (%)	
Positions in a turn of <i>oh</i>	300	100	278*	100	300	100	
Extra-clausal: turn-initial	264	88.0	219	78.8	263	87.7	
Extra-clausal: turn-medial	24	8.0	45	16.2	19	6.3	94.3
Extra-clausal: turn-final	1	0.3	1	0.4	1	0.3	
Intra-clausal: after an M-	5	1.7	9	3.2	11	3.7	
Intra-clausal: after an MA	1	0.3	1	0.4	3	1.0	
Intra-clausal: after an MF	1	0.3	2	0.7	2	0.7	
Intra-clausal: others	4	1.3	1	0.4	1	0.3	

*22 instances of *oh* in the 300 random samples are the number 0.

5.4.1.1 *Oh* in extra-clausal position

In the NNSs' monologues, *oh* is fairly evenly distributed in extra- and intra-clausal positions. In the NSs' monologic genres, about two-thirds of the instances are placed in extra-clausal positions. The extra-clausal utterance-medial position, as in Excerpt (5.4.1), is the most frequent position.

(5.4.1)

P: Utterance-medial And we also have a very good dinner. My mother prepared... a very... big cake for me. **[Oh]**, it's... ve... it was very delicious.

(SECCL: B00-11-30)

In the NNSs' monologues under investigation, each text is seen as a single utterance. 7 instances (1.8%) of *oh* are found to be utterance launchers, as shown in Excerpt (5.4.2).

(5.4.2)

P: Utterance-initial

Task 2

... **[Oh]**... When I was ten years old, I had<have> a wonderful... birthday. You know in China, the ten years is an important age in one's life.

(SECCL: B00-58-10)

Unlike *oh* in the monologues, it is clear in Table 5.9 above that in the three sub-corpora of the dialogic genres, a large proportion (about 95% on average) of the instances of *oh* are in extra-clausal position. 84% on average of all instances are used as turn openers, as exemplified in Excerpt (5.4.3) below.

(5.4.3)

P: Turn-initial

B: I I think... um... gradua... go abroad for graduating... after graduating... for study is good to you.

A: **Oh**, I think it <it's> can do a lot of good for us... eh... ... in English studying.

(SECCL: C01-123-09)

As in the NNSs' monologues, one monologue is taken as one utterance and in the dialogues, a turn is taken as a unit of utterance, it can be argued that due to the sheer number of turns in the dialogues, *oh* is more likely to occur turn-initially in the dialogues. The following calculation can also be made. In the monologues, 7 out of 1,143 utterances begin with *oh*, representing 1.8%. In the dialogues, 2,452¹² out of 29,542¹³ turns in the 1,143 texts begin with *oh*, accounting for 8.3%. Therefore, it is reasonable to assume that *oh* is used as a turn opener more often in the NNSs' dialogues than their monologues. The use of *oh* as a turn opener in the NNSs' dialogues is further investigated below.

5.4.1.2 *Oh* in intra-clausal position

It is likely that *oh* in intra-clausal positions occurs more often in the three monologic genres (about 38% on average) than in the dialogic genres (less than 5% on average). Of the intra-clausal positions, *oh* occurs more often after an incomplete message (M-) element, as exemplified in Excerpt (5.4.4).

¹² In Table 5.4 above, *a* and *b* refer to the identification of the two speakers in the NNSs' dialogues. 2,452 instances of *a* (1,204) and *b* (1,248) precede the node word *oh*.

¹³ The two speakers in the NNSs' dialogues are identified as *a* and *b*. Searching for *a*: and *b*: reveals 15,110 instances of *a*: and 14,432 of *b*:. From these figures, it can be inferred that there are probably 29,542 turns in the NNSs' 1,143 dialogues.

(5.4.4)

P: M- + *oh* + +M

It's beyond my expectation and my parents said to me |oh|, my dear, this day is <was> your birthday. Happy birthday."

(SECCL: B00-58-15)

It is difficult to be precise, but the positions where *oh* occurs seem to affect its functions in discourse. For example, *oh* following an M- element is likely to mark the boundary of reported speech (e.g. Excerpt (5.4.4) above) and to co-occur with a repair. The types of co-occurrence of *oh* are discussed in the following section.

5.4.2 Contexts where *oh* tends to occur

The positions of *oh* in an utterance/turn described in the preceding section are referred to in the present discussion of the contexts where *oh* tends to occur. The types of co-occurrence of *oh* in both the NNSs' and NSs' speech under investigation are discussed in order of the strength of the evidence, from stronger linguistic evidence to intuitive interpretation. As mentioned earlier, although linguistic evidence is used to determine the categories for discussion, sometimes intuition is unavoidable in interpreting the use of DMs.

Oh is found to co-occur with 1) reported speech, 2) hesitation markers, pauses or repetitive words, 3) repairs and rephrasing, 4) opening and changing a topic, 5) as a (preface to a) response to a question or new information, 6) for showing emotions and 7) implying that a cognitive process has been completed.

The instances in ambiguous contexts, with no linguistic evidence and insufficient contextual information remain unclassified in my analysis. Tables 5.10 to 5.15 at the end of this section show the distribution of the types of co-occurrence of *oh* in relation to its positions in an utterance/turn in the six sub-corpora under investigation.

5.4.2.1 *Oh* co-occurring with reported speech

In the NNSs' monologues, the most frequent collocates to the left of *oh* are *said* and *say* (see Table 5.3 above). These collocates imply that a high proportion of the 397 instances of *oh* co-occurs with reported speech, as in Excerpt (5.4.5) below.

(5.4.5)

P: M- + *oh* + +M

E: Quoting verb *said*

F: To signal reported speech

We played very happy. And my grandmother **said**: "|Oh|, this is the most happy time I have "

(SECCL: B00-11-01)

Table 5.10 below shows that *oh* marks reported speech accounts in 51.4% of cases. This is much higher than the pattern of *oh* revealed in Table 5.3, because in some instances of reported speech, the reporting verbs, such as *SAY* and *TELL*, follow quotations and occur out of a 4:4 span of the node word *oh*. In some instances, the reported speech does not co-occur with reporting verbs. Manual investigation reveals that more than half (204) of the 397 occurrences of *oh* mark the boundary between the mode of the speaker and reported speech, as in Excerpt (5.4.5) above. In the three sets of 100-line concordance samples from the NNSs' dialogues, only one instance is found of *oh* signalling reported speech.

In the NS speech, *oh* is one of the utterance openers used as a device to mark the boundary between the mode of the speaker and reported speech (Biber *et al.* 1999: 1118). In the NSs' monologic genres under investigation, *oh* co-occurring with reported speech represents 33.3% and 17.8% in MICASE and ICE-GB respectively. The figures are not as high as they are in the NNSs' monologues. In the NSs' highly interactive discourse mode in MICASE and the direct conversations in ICE-GB, the proportion is slightly higher than in the NNSs' dialogues, accounting 2.5% and 3.3%, as opposed to 0.3%.

In the NNSs' monologues, the high proportion of *oh* co-occurring with reported speech may be attributed to the topics. The speakers were asked to talk about previous experience and events, in which they might have more chances to quote from others. In the dialogues, they were asked to exchange opinions and here very few narratives were produced (see Appendix 1 for the topics for discussion in the NNSs' monologues and dialogues).

As discussed earlier in Section 5.2.5, *oh* could be part of a direct quotation or could have been added by the present speaker. In the NNS data for analysis, the latter is quite possible, given the fact that the quotations reported by the NNSs would have been translated from their L1, Chinese.

5.4.2.2 *Oh* co-occurring with hesitation markers, pauses and repetitive words

Oh is found to co-occur with hesitation markers, pauses or repetitive words. In Excerpt (5.4.6) below, *oh* is accompanied by the vocative hesitation marker *eh* and pauses which are likely to have been used to delay speech while the speaker searches for an appropriate description. (This is similar to one of the uses of *like*, see Section 4.4.2.1.)

(5.4.6)

P: M + *oh* + MR

E: Vocative hesitation
marker *eh* and
pauses

F: To search for
contents or lexis

I met a lots of teachers who... *eh*... make me unforgettable <unforgettable>
but... **eh**... Most of them seems quite... |**oh**|... quite a... a responsible and
follow the traditional teaching way <wayed>

(SECCL: B01-99-20)

This type of co-occurrence accounts for 10.1% and 3.7% in the NNSs' monologues and dialogues respectively. In addition, the patterns of *oh* (see Tables 5.3 and 5.4 above) indicate that *eh* is one of the salient collocates immediately to the left and to the right. It is concluded that *oh* may indicate a search for content information or lexical words.

In the NSs' speech, this is one of the less-used types of co-occurrence, with 4.2% and 4.4% in the monologic genres and no occurrence in the dialogic genres in MICASE and ICE-GB. Two possible reasons are that NSs use something else to signal hesitation and that NSs are more fluent in English than NNSs.

5.4.2.3 *Oh* co-occurring with repairs and rephrasing

Oh is found to be followed by a correction or rephrasing of the term or statement before *oh*. In Excerpt (5.4.7) below, *oh* prefaces the correction of Item (1) and in Excerpt (5.4.8), Item (2) replaces Item (1). In these cases, *oh* seems to suggest that a correction is made and serves to mark a self-repair. This is one of the less frequent co-occurrence in the NNSs' speech, with only 7.1% and 2% in the monologues and dialogues respectively. In the NSs' speech, few instances are found (4 instances in MICASE and 6 in ICE-GB), as in Excerpt (5.4.9).

(5.4.7)

P: MA + *oh* + +M

E: Item (2) revises
Item (1)

F: To signal a repair

we admire him very much. Ah, ⁽¹⁾Thomas was, |**oh**|, ⁽²⁾is very kind-hearted.

(SECCL: B01-08-01)

(5.4.8)

P: M + *oh* + MR

E: Item (2) replaces
Item (1)

F: To signal a repair

A: Yeah, independence! And I think we always depend on our, our parents or ⁽¹⁾our students... |**oh**|... ⁽²⁾our teachers, right?
B: Um.

(SECCL: C96-13-12)

(5.4.9)
P: MF + OI + *oh* +
M- uh what he's using, and, so what we do is, uh, we compare_ so the gene string is
just the, uh just the phase at each, each pixel, and we compare pixel by pixel to
E: Item (2) revises what the phase would be for the transform-limited pulse, and, uh, and so we add
Item (1) up all those differences and that's actually squared right? ⁽¹⁾that's just the, uh |oh|
F: To signal a repair ⁽²⁾it's either squared or

(MICASE: MTG485SG142)

As mentioned in Section 5.4.1, the positions where *oh* occurs seem to correlate with its co-occurrence in discourse. Most of the instances of *oh* signalling a repair are placed in intra-clausal positions.

5.4.2.4 *Oh* occurring at the opening and changing of a topic

In the NNSs' monologues under investigation, each text is seen as a single utterance. 7 instances (1.8%) of *oh* are found to be utterance launchers, as in Excerpt (5.4.10). It seems to be rare for an utterance to begin with *oh*. However, in cases where the speakers were recorded after listening to the test questions, my interpretation is that *oh* is used to respond to the given instruction. In Excerpt (5.4.11), *oh* occurs before Item (1), which is a new topic. It seems to be reasonable, then, to suggest that *oh* prefaces a new topic.

(5.4.10)
P: Utterance-initial Task2
E: Occurring utterance **Oh**|... oh... a teacher I found... he's unusual. Oh... he;s called Mr zhang...
initially and beginning ah... He is our... ah... he is our English teacher.
a topic (SECCL: B01-123-22)
F: As a topic opener

(5.4.11)
P: Turn-initial B: I won't bother to pronounce the Latin names because I'm only just uh
starting to <,> to learn them
E: Item (1) is a new A: Michelle on Sunday was sort of pronouncing some of the Latin names
topic B: Yes <laugh> |Oh| ⁽¹⁾**you have to go and enter a competition about that**
F: As a topic opener B: Oh look there's a <,> a betula betula pendulosa or something <laugh>
A: This week she's doing uh she's into doing dairy and pigs <,> so if she
does come tonight she'll probably be stinking of pigs
(ICE-GB: S1A-036)

Oh does not often occur at the opening and changing of a topic, appearing 2.3% and 6.7% of the time in the NNSs' monologues and dialogues respectively. In the NSs' speech, this type of co-occurrence is also rare, accounting for 4.2% and 2.2% in the two subsets of the monologic genres and 4.3% and 5.7% in the dialogic genres in MICASE and ICE-GB.

In this category, some instances of *oh* at the changing of a topic are found to preface

indicators of *misplacement*. In Excerpt (5.4.12), *oh* occurs between Item (1), which closes a topic, and Item (2), which begins a new topic; therefore, it seems to be reasonable to suggest that *oh* prefaces a new topic. In this case, *oh* is not used only for changing a topic, but might introduce a topic which the listeners do not expect. In Excerpt (5.4.13), *oh* precedes Item (1), which is a “misplacement marking” (Schegloff and Sacks 1973: 319). There is no such instance in the NNS and NS sub-corpora of the monologic genres. In the NNSs’ dialogues, 3 out of 20 instances of *oh* co-occur with indicators of *misplacement*. In the highly interactive discourse mode in MICASE, 3 out of 12 instances and in the private direct conversations in ICE-GB, 3 instances are found out of 17.

(5.4.12)

P: Turn-medial
E: Item (1) closes a topic
and Item (2) begins a
new topic
F: As a topic opener

A: Now, I see, and ⁽¹⁾**thank you for your help.** |**Oh** |, ⁽²⁾**I have another point.**
B: What?
A: Some, some students still, still feel very puzzled about how to, about social contact.

(SECCL: C01-01-01)

(5.4.13)

P: Utterance-medial
E: Item (1) begins a
relevant topic
F: As a topic opener

the regulations of these field releases tends to focus on human safety, they also um, look at, potential risks to the environment. so the factors that they usually consider, |**oh**| ⁽¹⁾**the other thing i wanted to just mention**, was the, regulatory commissions at this point.

(MICASE: LES405JG078)

The examples of *oh* co-occurring with indicators of *misplacement* show that the NNSs tend to use language which might be interpreted as too direct, for example, Item (2), *I have another point*, in Excerpt (5.4.12) above. In contrast, the NSs’ use of Item (1) in Excerpt (5.4.13), *the other thing I wanted to just mention*, is more hedged. The possible implications for pedagogy are discussed, in Chapter 12, with other less direct uses of language for softening speech.

5.4.2.5 *Oh* as a (preface to a) response to a question and new information

A very high proportion of the occurrences of *oh* in the NNSs’ dialogues (88%), the NSs’ highly interactive discourse mode (78.8%) and the private direct conversations (87.7%) show *oh* in the turn-initial position (see Table 5.9). Most of these instances of *oh* are a response token or a preface to an answer/response to a question and new information. This type of co-occurrence is more frequent in the NNSs’ dialogues, at 81.3% (see Table 5.11 below). In a

slightly lower proportion in the NSs' speech, it accounts for 76.3% in the NSs' highly monologic discourse mode (see Table 5.13) and 68.3% in the NSs' private direct conversations (see Table 5.15).

The instances in this category include three types of *oh* as a response. It is found that *oh* is used 1) as a response token, 2) in fixed expressions, such as *oh really* and *oh yes (yeah)* and 3) as a preface to a response, answer, comment or evaluation.

In Excerpt (5.4.14), *oh* is used as a standalone response token to acknowledge the receipt of information.

- (5.4.14)
- | | | |
|---|--|---------------------------|
| <p>P: Turn-initial
 E: Responding to Item (1) new information
 F: As a response token to new information</p> | <p>B: Oh I remember that I once read a book mm. ⁽¹⁾ It was written by a American famous mm American scholar.
 A: Oh.
 B: His name is Carnage.</p> | <p>(SECCL: C97-01-28)</p> |
|---|--|---------------------------|

Oh frequently occurs in such fixed expressions as *oh yes (yeah)*, *oh no*, *oh thank you* and *oh really* as a response, as exemplified in Excerpts (5.4.15).

- (5.4.15)
- | | | |
|---|---|---------------------------|
| <p>P: Turn-initial
 E: <i>Oh yes</i> as a fixed expression to answer Item (1)
 F: As a preface to a response in a fixed phrase</p> | <p>A: Um... it is easy for you to be fired. ⁽¹⁾ Do you think so?
 B: Oh, yes . That's a problem.</p> | <p>(SECCL: C99-25-32)</p> |
|---|---|---------------------------|

In Excerpt (5.4.16) below, *oh* prefaces Speaker B's answer to Speaker A's question.

- (5.4.16)
- | | | |
|---|--|---------------------------|
| <p>P: Turn-initial
 E: <i>Oh</i> prefaces Speaker B's answer, Item (2), to Speaker A's question, Item (1)
 F: As a preface to a response to a question</p> | <p>Task 3
 A: Um... can I ask you question, ⁽¹⁾ what do you think of your future?
 B: Oh , ⁽²⁾ I have <had> a lot of dreams. <A: Your?> Yeah... eh... .. first <firsty> I want to become... a... good interpreter.</p> | <p>(SECCL: C98-08-16)</p> |
|---|--|---------------------------|

In Excerpt (5.4.17), *oh*, prefacing the speaker's evaluation, acknowledges the receipt of new information.

- (5.4.17)
- | | |
|--|---|
| <p>P: Turn-initial</p> <p>E: Responding to Item (1) new information</p> <p>F: As a preface to a response to new information</p> | <p>B: Something we can't experience I think take a part-time job in short take part-time job is have has more advantages than disadvantages.</p> <p>A: Yes, ⁽¹⁾ this is also my opinion</p> <p>B: Oh, I like this.</p> |
|--|---|
- (SECCL: C96-05-08)

5.4.2.6 *Oh* to show emotions

As mentioned earlier, the general understanding of *oh* is that it is used as an interjection to express emotions. Likewise, one of the main functions of *oh* suggested in the literature is to express speakers' emotions, such as surprise and not expecting something (Aijmer 1987: 61, Biber *et al.* 1999: 1083). In Excerpt (5.4.18) below, *oh* seems to show the speaker's surprise and in Example (5.4.19) the speaker's happiness. These arguments are supported by Items (1) and (2), in which the speakers explicitly report their previous emotions.

- (5.4.18)
- | | |
|---|---|
| <p>P: Utterance-medial</p> <p>E: Item (1) explicitly describe emotion and Item (2) reiterates the same emotion</p> <p>F: To show emotions (surprise)</p> | <p>when the first time I saw him ⁽¹⁾ I was really astonished <atonised> by his appearances. Oh look... what he is... ver what he is veering he wear<wear> long hair and a shirt with combined colors and blue jeans and she even w... v... eh... ... and he even wore a glasses with pikuliar frim ⁽²⁾ I was really surprised.</p> |
|---|---|
- (SECCL: B01-01-26)

- (5.4.19)
- | | |
|--|--|
| <p>P: Turn-medial</p> <p>E: Item (1) and (2) explicitly state the emotions</p> <p>F: To show emotions (happiness)</p> | <p>And my, my parents, my parents, my parents give me, give me a computer as, as, as my, my birthday, birthday present. Oh , I was, I was, I was longing for it. I, so, I was. ⁽¹⁾ I was very happy, very very happy, very very happy. ⁽²⁾ I was so excited.</p> |
|--|--|
- (SECCL: C00-29-26)

In 21 out of 397 instances, *oh* is used with (*my*) *god* and *dear* to show emotions, as in Excerpt (5.4.20). *Oh* is often used with (*my*) *god* and *dear*, combining fixed expressions.

- (5.4.20)
- | | |
|--|---|
| <p>P: Utterance-medial</p> <p>E: Co-occurring with <i>my god</i></p> <p>F: To show emotions (upset; surprise)</p> | <p>I mounted to the bicycle and put my hands on the handles and just when I want started to move, oh my god! I just fall down on the ground with the bicycle, the poor bicycle.....</p> |
|--|---|
- (SECCL: B99-66-22)

Both *oh* as a single item and *oh* in fixed expressions are found to show emotions in the

NNSs' speech. This use occurs frequently in the monologues, representing 22.7%, but it accounts for only 4.3% in the dialogues.

Three of the four NS sub-corpora contain a similar proportion of *oh* showing emotions to that in the NNS sub-corpora. It represents 25% in the NSs' highly monologic discourse mode and 6.1% in the highly interactive discourse mode in MICASE and 11.7% in the private direct conversations in ICE-GB. By contrast, there is a much larger proportion (66.7%) in the NSs' unscripted monologues in ICE-GB, as in Excerpt (5.4.21) below. A further look reveals that 23 out of the 26 instances are from sports commentaries, in which commentators use *oh* to convey their emotions.

(5.4.21)

P: Utterance-medial
E: Item (1) shows the speaker's emotions
F: To show emotions (happiness)

Has he found touch |**Oh**| ⁽¹⁾it 's a superb kick , that really is into the wind
 That 's a long kick and takes play just inside the Irish twenty-two far side of
 the field,

(ICE-GB: S2A-002)

5.4.2.7 *Oh* to indicate a cognitive process has been completed

Based on the use of cognition-related verbs, such as *remember*, *know* and *realize*, *oh* can mark the speaker's cognitive process. In Excerpt (5.4.22) below, *oh* co-occurs with *remember* to indicate that a cognitive process has just taken place.

(5.4.22)

P: Utterance-medial
E: Co-occurring with *remember*
F: To indicate a cognitive process has just been done

she is a formal writing teacher... oh... she was quite activated <activity> in
 our class oh... sometimes... she was <were> asked us to answer <answering>
 the question which she posted very much in China... sometimes maybe shi...
 will ask us to be the teacher in turn, and she be... become the student. I think
 it's quite interesting for us. |**Oh**| **I remember** Nancy often said that the book
 didn't reflect now much you learn from this major.

(SECCL: B01-99-20)

In Excerpt (5.4.23), no cognition-related verbs are used, but from the context, it is clear that the speaker's use of *oh* is to suggest that enacting of a cognitive process.

(5.4.23)

P: Turn-medial

E: The speaker finds a misunderstanding

F: To indicate a cognitive process has just been done

A: I heard... you will go, you will go aboard for <of> study after the your, your graduation... this, this term. Is that true?

B: ... um... I will go abroad? By here you, by there that you want to go abroad. The one to go abroad?

A: Oh, hm... um.

B: <Silence> |oh|, **yeah... sorry. I made a mistake** and... eh... ... I think that... eh... ... the friend of us will go abroad I although think, I think... eh... ... it's too young for him to go abroad... eh... ... what's your opinion?

(SECCL: C01-99-34)

To identify the use of *oh* in this last category, an intuitive interpretation is sometimes needed. To test my interpretation, *I see* and *I remember* can be inserted after *oh* to test if the utterance still makes sense and if it contains new information. In Excerpt (5.4.23) above, if *I see* is added after *oh*, the utterance is actually clearer. *Oh I see* suggests the course of a cognitive process and a revised form of the information follows.

This type of co-occurrence is less frequent, with 5% and 1.3% in the NNSs' monologues and dialogues respectively, with 16.7% and 8.3% in the NSs' highly monologic and highly interactive discourse mode in MICASE and with 6.7% and 8.3% respectively in the unscripted monologues and the private direct dialogues in ICE-GB.

5.4.2.8 Problematic and unclassified instances of *oh*

It is possible that more than one type of evidence can be identified. The remaining cases of *oh* are classified as examples of stronger evidence. In Excerpt (5.4.24), *oh* occurs between the quoting verb *said* and the reported speech and it can also be interpreted as part of the fixed expression *oh dear* to show emotions. Though it is unknown whether *oh dear* is produced by the person being quoted or the person being recorded, the co-occurrence of reported speech seems to be stronger evidence and therefore this instance of *oh* is coded as in the category of reported speech.

(5.4.24)

P: M- + *oh* + OI

E: Quoting verb; fixed expression showing emotions

F: Primarily to signal reported speech

The... the absent-minded teacher said,... |oh| **dear**, I suppose you are right. I remember now, when I... eh..., when I... came out of the car<card>.

(SECCL: B01-50-05)

Excerpts (5.4.25) and (5.4.26) are observed to co-occur with emphatic lexis. *Oh* seems to preface emphatic lexis in order to emphasise what follows. However, another type of

co-occurrence is identified. In Excerpt (5.4.25), *oh* occurs between the correction (Item 2) and the corrected (Item 1). In Excerpt (5.4.26), *oh* co-occurs with the cognition-related verb, *realize*, suggesting the completion of a cognitive process. It is not easy to identify the primary co-occurrence and two instances are not enough to form a category. Therefore, Excerpt (5.4.25) is classified in the category of repair and Excerpt (5.4.26) into that of a cognitive process.

(5.4.25)

P: MA + *oh* + M + M

E: Emphatic lexis, *of course*; Item (2) revises Item (1)

F: To emphasise what follows; to signal a repair

I asked um... why why did you touch my bag for many times. He look at he looked at me surprised and said I didn't but I'm at that time, ⁽¹⁾ I'm |oh| of course ⁽²⁾ I didn't believe him.

(SECCL: B02-100-18)

(5.4.26)

P: Utterance-medial

E: Emphatic lexis, *certainly*; cognition-related verb, *realize*

F: To emphasise what follows; to indicate a cognitive process has just been completed

I opened the door, she came in with a bunch of flowers. T is <oli> <meithid> in <waieient> that clour I loved most, and said to me, "Happy birthday". |Oh|, **certainly**, I **realized** it was my birthday eh... quickly,

(SECCL: B00-58-25)

6 (1.5%) out of 397 instances in the NNSs' monologues and 1 (0.3%) out of 300 instances of random samples in the NNSs' dialogues were found impossible to classify due to the lack of linguistic evidence. 3 (1.1%) out of 278 instances in the NSs' highly interactive discourse mode in MICASE and 3 (1%) out of 300 instances of random samples in the private direct conversations in ICE-GB remain unclassified also.

5.4.2.9 Summary of the contexts where *oh* tends to occur

There seems to be no obvious distinction in the use of *oh* in the speech of the NNSs from that of the NSs. In terms of positions in an utterance/turn (see Table 5.9 above), in the monologic genres, the occurrences of *oh* is similarly distributed in extra- and intra-clausal positions, while in the dialogic genres, over 94% of the occurrences are in extra-clausal position and most of them are placed turn-initially.

Regarding the contexts where *oh* occurs, it can be seen in Tables 5.10 to 5.15 below that

there is no marked distinction between the speech of the NNSs and NSs in the distribution of the co-occurrence of *oh*, except that a larger proportion of *oh* in the NNSs' speech than in the NSs' co-occurs with 1) reported speech and 2) hesitation markers, pauses and repetitive words.

In the NNSs' monologues and the NSs' highly monologic discourse mode in MICASE, *oh* co-occurs most frequently with reported speech. Therefore, it is concluded that one of the major uses of *oh* is to mark reported speech. In the NSs' unscripted monologues in ICE-GB (see Table 5.14), two thirds of the instances of *oh* are used to show emotions and almost all of these instances are from sports commentaries. This is attributed to the nature of this activity, mostly sports commentaries, in this sub-corpus (see Appendix 3 for the fact sheet of the sub-corpus of the unscripted monologues in ICE-GB).

In the three sub-corpora of the dialogic genres, *oh* is most frequently used as a (preface to a) response to a question and new information. In particular, this use represents 81.3% (see Table 5.11) in the NNSs' dialogues, as opposed to 76.3% (see Table 5.13) and 68.3% (see Table 5.15) in the NSs' highly interactive discourse mode in MICASE and the private direct conversations in ICE-GB. With a closer look at these three sub-corpora, some differences in frequency are identified and these phenomena are further discussed in the next section.

Table 5.10: Distribution of co-occurrence of *oh* in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Reported speech	51.4			75	18.9			109	27.5	1	0.3			19	4.8
2. Hesitation markers; pauses; repetitive words	10.1			17	4.3			13	3.3			1	0.3	9	2.3
3. Repairs; rephrasing	7.1			5	1.3					9	2.3	5	1.3	9	2.3
4. Opening and changing a topic	2.3	7	1.8	2	0.5										
to a question	0														
5. As a (preface to a) response	0														
to a question (fixed phrase)	0														
to new information	0														
to new information (fixed phrase)	0														
6. Showing emotions	22.7			87	21.9			3	0.8						
7. Implying a cognitive process has been completed	5.0			12	3.0			6	1.5					2	0.5
Unclassified	1.5			5	1.3			1	0.3						
Occurrences:		397	100.0	7	1.8	203	51.1	132	33.2	10	2.5	6	1.5	39	9.8

Table 5.11: Distribution of co-occurrence of *oh* in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Reported speech	0.3							1	0.3						
2. Hesitation markers; pauses; repetitive words	3.7	2	0.7	5	1.7			3	1.0					1	0.3
3. Repairs; rephrasing	2.0			1	0.3					1	0.3	1	0.3	3	1.0
4. Opening and changing a topic	6.7	15	5.0	5	1.7										
to a question	15.3	42	14.0	4	1.3										
5. As a (preface to a) response (81.3%)	44.0	128	42.7	4	1.3										
to a question (fixed phrase)	2.7	8	2.7												
to new information	19.3	57	19.0	1	0.3										
to new information (fixed phrase)	19.3	57	19.0	1	0.3										
6. Showing emotions	4.3	11	3.7	1	0.3			1	0.3						
7. Implying a cognitive process has been completed	1.3	1	0.3	3	1.0										
Unclassified	0.3													1	0.3
Occurrences:		300 (random samples)	100.0	264	88.0	24	8.0	1	0.3	5	1.7	1	0.3	1	0.3

Table 5.12: Distribution of co-occurrence of *oh* in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%		
		Utterance-initial	%	Utterance-medial	%	Utterance-final	%	after an M-	%			after an MA	%
1. Reported speech	33.3			3	12.5			5	20.8				
2. Hesitation markers; pauses; repetitive words	4.2							1	4.2				
3. Repairs; rephrasing	0												
4. Opening and changing a topic	4.2			1	4.2								
to a question	12.5	3	12.5										
5. As a (preface to a) response	4.2	1	4.2										
to a question (fixed phrase)	0												
to new information	0												
to new information (fixed phrase)	0												
6. Showing emotions	25.0	3	12.5	2	8.3			1	4.2				
7. Implying a cognitive process has been completed	16.7	2	8.3	1	4.2			1	4.2				
Unclassified	0												
Occurrences:	24	9	37.5	7	29.2			8	33.3				

Table 5.13: Distribution of co-occurrence of *oh* in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%		
		Turn-initial	%	Turn-medial	%	Turn-final	%	after an M-	%			after an MA	%
1. Reported speech	2.5							7	2.5				
2. Hesitation markers; pauses; repetitive words	0												
3. Repairs; rephrasing	1.4			1	0.4					1	0.4	1	0.4
4. Opening and changing a topic	4.3	7	2.5	5	1.8								
to a question	5.0	14	5.0										
to a question (fixed phrase)	4.3	12	4.3										
5. As a (preface to a) response (76.3%)	37.8	97	34.9	8	2.9								
to new information	29.1	77	27.7	4	1.4								
to new information (fixed phrase)	6.1	6	2.2	11	4.0								
6. Showing emotions	6.1	6	2.2	11	4.0								
7. Implying a cognitive process has been completed	8.3	6	2.2	14	5.0			2	0.7			1	0.4
Unclassified	1.1			2	0.7	1	0.4						
Occurrences:	278 (random samples)	219	78.8	45	16.2	1	0.4	9	3.2	1	0.4	2	0.7

Table 5.14: Distribution of co-occurrence of *oh* in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%			after an MA	%	after an MF
1. Reported speech	17.8			1	2.2			7	15.6					
2. Hesitation markers; pauses; repetitive words	4.4							2	4.4					
3. Repairs; rephrasing	2.2			1	2.2									
4. Opening and changing a topic	2.2										1	2.2		
to a question	0													
5. As a (preface to a) response	0													
to a question (fixed phrase)	0													
to new information	0													
to new information (fixed phrase)	0													
6. Showing emotions	66.7	3	6.7	24	53.3			1	2.2		1	2.2	1	2.2
7. Implying a cognitive process has been completed	6.7							1	2.2		1	2.2	1	2.2
Unclassified	0.0													
Occurrences:	45	3	6.7	26	57.8			11	24.4		3	6.7	2	4.4

Table 5.15: Distribution of co-occurrence of *oh* in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%				
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%			after an MA	%	after an MF	%
1. Reported speech	3.3	1	0.3					9	3.0						
2. Hesitation markers; pauses; repetitive words	0														
3. Repairs; rephrasing	1.7			4	1.3					1	0.3				
4. Opening and changing a topic	5.7	13	4.3	2	0.7			1	0.3				1	0.3	
to a question	4.3	11	3.7	2	0.7										
5. As a (preface to a) response (68.3%)															
to a question (fixed phrase)	3.3	9	3.0	1	0.3										
to new information	27.7	81	27.0	2	0.7										
to new information (fixed phrase)	33.0	98	32.7	1	0.3										
6. Showing emotions	11.7	30	10.0	3	1.0	1	0.33				1	0.3			
7. Implying a cognitive process has been completed	8.3	17	5.7	4	1.3			1	0.3	2	0.7	1	0.3		
Unclassified	1.0	3	1.0												
Occurrences:	300 (random samples)	262	87.7	19	6.3	1	0.3	11	3.7	3	1.0	2	0.7	1	0.3

5.5 Further investigation

5.5.1 The over-representation of *oh* in extra-clausal turn-initial position in the sub-corpora of the dialogic genres

Based on the corpus analysis, in both the NNS and NS sub-corpora of the dialogic genres, the most popular position of *oh* is extra-clausal turn-initial and the most frequent use is as a (preface to a) response to a question and to new information. A full-length text (5.5.1) is now examined in a broader context to see if there is any subtle difference in the use of *oh* made by the NNSs and NSs.

All the six instances of *oh* in Text (5.5.1) are placed turn-initially. At a glance, the opening of this dialogue seems to be odd, but, bearing in mind that the two speakers are performing a role-play in a test-taking setting, their discourse is peculiar in some respects, for example, the formal greetings between the two fellow students and the use of *oh* to raise the first question, which is seen as a topic opener. In the NSs' speech, *oh* is also used to open a topic. As in Excerpt (5.5.2), *oh* prefaces a question in order to change the topic. This instance of *oh* also shows the speaker's surprise, which is different from the first instance of *oh* in Text (5.5.1).

The 2nd, 3rd and 5th instances of *oh* in Text (5.5.1) are taken as a preface to the answer to a question. A similar use is found in the NSs' speech, as in Excerpt (5.5.3). With a slight difference, the use of *oh* (2) in Text (5.5.1) is to preface the answer, *oh, my name is Wangyan*, to a simple question, *what's your name*. It sounds unnatural to answer a simple question with this word, which tends to be used to express emotions or indicate that a cognitive process has just ensued.

(5.5.1)

Task 3

- (1) opening a topic and prefacing a question
- (2), (3) & (5) as a preface to the response to a question
- (4) responding to new information in a fixed expression
- (6) responding to new information
- A: How do you do?
B: Fine, thank you. How do you do?
A: ⁽¹⁾ **Oh**, what's your name?
B: ⁽²⁾ **Oh**, my name is Wangyan. I think you are a freshman.
A: Yes, I'm glad to meet you. My name is Shengtianyun.
B: Me, too.
A: Um... I think you are a sophomore.
B: Yes, I'm a second year student in this university <uniwersity>.
A: Um... I want to know how do you feel about your university life? You know, I'm a freshman. I don't know how the university life like. Could you tell something about it?
B: ⁽³⁾ **Oh**, I think um... the university, the university life was colorful, but it also full of challenge.
A: ⁽⁴⁾ **Oh, really**?
B: Yes, um... if you have, um... stayed here for a long time, you'll discover it.
A: But, um... I have heard that um... the university life is not... har, hard, because um... the student doesn't need to study very hard.
B: But the condition in our university is not the same as the other university, because...
A: Is the competition fierce?
B: ⁽⁵⁾ **Oh**, I think if you prepared yourself, it is not very fearful, such as um... because we have to pass the Band Four examination in the, in the second year.
A: Um... is there important test in these four years?
B: Band Four and also Band Eight.
A: Band Eight?
B: Yes, so you must prepare yourself, but fully prepared yourself.
A: ⁽⁶⁾ **Oh**... And how about the relation between students. Sometimes I'm afraid that the relationship between students in the university is complicated. And I never lived in dormitory. I'm not sure whether I can get on well it or not. Could you give me some advice?
B: Um... I think most of the student are easy to get along with, such as, in my dormitory. And ah... the, the students are very simple and purity. They are easy to get along. I think um... This is my advice <advice> I think um... as your study, you should always go to the library, and broaden your ha, broaden your vision. The school has creat a good condition for us. Have you been to the library?
A: No, I haven't. Because I'm just come to this university and enroll my name in the university. And I didn't have time to um... go around the university. Um... maybe later, you can show me around.
B: Ah. OK, I'm very glad.
A: Um... I think um... maybe the study um... is very difficult, because we are major in English.
B: If you follow the teacher's advice, and study step by step, not to or... step by step, and then you will keep up with the other students. You can, you can buy some reference books, um... such as... for the Band Four examination.
A: Um... is Band Four examination is difficult?
B: Um. I think if you pre, got fully prepared, eh... .. it is not very difficult. Most of the student can pass it. Um... I have another advice to give you. I think um... the university life, the university not only put much emphasis on the study but also your ability. You should cultivate...

(SECCL: C00-11-01)

(5.5.2)

- P:** Turn-initial SU-m: there's a, flicker.
SU-f: is that a Blue Jay? <OVERLAP1> oh </OVERLAP1>
- E:** Item (1) begins a new topic SU-f: <OVERLAP1> yeah i heard, </OVERLAP1> Blue Jays.
SU-f: Blue Jays, a couple (of'em)
<PAUSE DUR=":10" ></PAUSE>
- F:** As a topic opener SU-f: oh there's something in um, in the field along the fence row there's like, three dead, trees and in the tallest
SU-f: <OVERLAP2> oh yeah </OVERLAP2>
SU-f: <OVERLAP1> one there's </OVERLAP1> a big tr- bird (in it.)
SU-m: is it a Northern Flicker?
SU-f: yeah it's got (a, black head) (xx)
SU-f: good I-D
<PAUSE DUR=":07" ></PAUSE>
SU-m: **oh**⁽¹⁾ **what's that?**
SU-m: oh, ooh
SU-f: they had a little white <OVERLAP1> on their </OVERLAP1>
SU-m: <OVERLAP1> that makes two </OVERLAP1>
SU-f: <OVERLAP1> another one. </OVERLAP1>
- (MICASE: LAB175SU026)

(5.5.3)

- P:** Turn-initial B: ⁽¹⁾ **Do you not think women have evolved to a higher uhm**
- E:** *Oh* prefaces Speaker A's answer, Item (2), to Speaker B's question, Item (1) A: **Oh**⁽²⁾ **I'm certain of it** <,>
B: In-style life form whatever you call it
A: Mm <,>
- F:** As a preface to a response to a question (ICE-GB: S1A-011)

The above text-based analysis seems to support the use of DM *oh* resulting from the corpus analysis. In general, the use made of *oh* by the NNSs is similar to that made by the NSs, but with some subtle differences. It can be concluded that the corpus-based analysis helps to understand the use of *oh* in the speech of the NNSs and NSs and a more detailed examination is needed to reveal the particular uses of *oh* in the NNSs' speech.

5.5.2 Revisiting *oh* as a (preface to a) response

Since *oh* is frequently used as a (preface to a) response to a question and new information in both the NNSs' and NSs' dialogic genres, the instances in this category are further examined to see if the NNSs use *oh* as a (preface to a) response in the same way as the NSs do.

Table 5.16 presents the proportion of the seven types of co-occurrence across the three sub-corpora of the dialogic genres. *Oh* is most often used as a (preface to a) response. The proportion is highest, at 81.3%, in the NNSs' dialogues, as opposed to 76.9% and 68.3% in the two NS sub-corpora. The highlighted percentages show the very different distribution of

the use of *oh* as a (preface to a) response. Two distinctions between the use by the NNSs and NSs may be drawn. The NSs tend to use *oh* in fixed phrases, such as *oh yeah*, *oh yes*, *oh no*, *oh really*, *oh God*, *oh dear* and *oh thank you* (see Section 5.3.2 above for the collocates of *oh*).

Another distinction is that *oh* as a (preface to a) response to questions is used more than three times as often in the NNSs' dialogues (15.3%) as in the NSs' highly interactive discourse mode (5.0%) and the private direct conversations (4.3%). Nevertheless, this comparison is not based on the number of questions in the corpora. It is difficult to conclude that the NNSs use *oh* as a (preface to a) response to a question more often than the NSs do. One possible explanation for this is that there are far more instances of questions in the NNSs' dialogues than elsewhere, since the speakers are constantly being asked to exchange ideas and turn-changing is therefore constant.

Table 5.16: Percentages of types of co-occurrence of *oh* in the three sub-corpora of the dialogic genres

Co-occurrence \ Percentage \ Corpus	SECCL (NNSs' dialogues)	MICASE (NSs' highly interactive discourse mode)	ICE-GB (NSs' private direct conversations)
1. Reported speech	0.3	2.5	3.3
2. Hesitation markers, pauses; repetitive words	3.7	0.0	0.0
3. Repairs; rephrasing	2.0	1.4	1.7
4. Opening and changing a topic	6.7	4.3	5.7
5. As a (preface to a) response	15.3	5.0	4.3
to a question	2.7	4.3	3.3
to new information	44.0	37.8	27.7
to new information (fixed phrase)	19.3	29.1	33.0
6. Showing emotions	4.3	6.1	11.7
7. Implying a cognitive process	1.3	8.3	8.3
Unclassified	0.3	1.1	1.0
Occurrences: 300 (random samples)	100.0	100.0	100.0

If the corpora had been tagged with questions, it would have been an ideal way to compare how the NNSs and NSs use *oh* to respond to questions. When the present study was undertaken, however, it was not possible to count the number of questions in the corpora under investigation. To investigate if the NNSs tend to use *oh* to respond to questions, an alternative method was used. Two interrogatives, *what* and *which*, were searched across the three sub-corpora of the dialogic genres. On the basis of the raw frequencies of *what* and *which*, a set of 100-line concordance samples was extracted. Each of these 100 instances was manually examined and the instances of *what*- and *which*-questions with *oh*-prefacing responses were counted.

For example, the raw frequency of *what* in SECCL was 2,330. The *Concord* tool in *WordSmith 4* (Scott 2004) was set to randomly select one in every 23 instances of *what*. This made it more likely that the tool would return 100 concordance lines equally distributed across texts. Of the 100 instances of *what*, 62 were found in questions. Among these questions, 13 instances (21%) were found to be followed by *oh* as a (preface to a) response.

Table 5.17 below shows that the NNSs use *oh* as a (preface to a) response to *what*- and *which*-questions more often than the NSs do. In the NNSs' dialogues, 21% of the responses to the *what*-questions are prefaced with *oh*, as opposed to 6% in the NSs' highly interactive discourse mode in MICASE and no instance in the private direct conversations in ICE-GB. Similarly, *oh*-prefaced responses to *which*-questions represent 16% in the NNS data, as opposed to 9% and 0% in the NS dialogic genres in MICASE and ICE-GB.

Table 5.17: Proportion of *oh* prefacing *what*- and *which*-questions in the three sub-corpora of the dialogic dialogic genres

Item \ Frequency \ Corpus	SECCL (NNSs' dialogues)	MICASE (NSs' highly interactive discourse mode)	ICE-GB (NSs' private direct conversations)
Raw frequency of <i>what</i>	2,330	5,030	1,317
Instances of <i>what</i> -questions (out of a 100-line concordance sample)	62	36	49
Instances of <i>oh</i> -prefaced responses	13 (21%)	2 (6%)	0 (0%)
Raw frequency of <i>which</i>	503	917	378
Instances of <i>which</i> -questions (out of a 100-line concordance sample)	32	23	7
Instances of <i>oh</i> -prefaced responses	5 (16%)	2 (9%)	0 (0%)

This further investigation supports the initial finding that the NNSs have a tendency to use *oh* as a preface to a response to a question, shown in Table 5.16 above. One possible explanation for this is L1 transfer. In Chinese, the equivalent of *oh* (哦 or 噢) is commonly used as an acknowledgment token, which is probably similar to the use of *yes* (*yeah*) for expressing listenership. To test this hypothesis, the speech of Chinese speakers should be compared with that of speakers of other L1s, but this is beyond the scope of this thesis. I want to leave it as an open question here and look at some of the available examples.

The three highlighted instances of *oh* in Excerpts (5.5.4), (5.5.5) and (5.5.6) are peculiar in that they do not seem to fit into the main functions of *oh* discussed in the literature (see Section 5.2). No evidence indicates that they are being used as a marker of change-of-state, nor are they a token of the receipt and acknowledgment of new information. Their prosodic

information is referred to in order to confirm that the use of *oh* in these three excerpts lacks the rising tone which would show emotion. *Oh* in Excerpts (5.5.4) and (5.5.5) is made in a neutral tone and *oh* in Excerpt (5.5.6) in a falling tone. If para-language, such as facial expressions and gestures, had been available for reference, it would have been easier to specify the use of *oh*. (The availability of multi-modal corpora, recently announced, will make it possible to analyse a given item with its audio-visually aligned data (Knight, Evans, Carter and Adolphs 2009).)

(5.5.4)

P: Turn-initial

E: *Oh* prefaces Speaker A's answer, Item (2), to Speaker B's question, Item (1)

F: As a preface to a response to a question

B: I don't think so. I think your intonation is very good. So I think you don't worry about it. You just try your best to speak it out, and don't worry so much about, don't worry so so much and I don't know... eh... what the time the speak contest begins<begin>. ⁽¹⁾ **What's the time it begins<begin>?**

A: **Oh** | ⁽²⁾ **tomorrow night.**

B: Tomorrow night?

A: I have little time to prepare for it.

(SECCL: C97-11-07)

(5.5.5)

P: Turn-initial

E: *Oh* prefaces Speaker B's answer, Item (2), to Speaker A's question, Item (1)

F: As a preface to a response to a question

B: Oh... oh, I... I, oh, I think you should take it easy, you know. The university life is very wonderful. I think if you want to be the very successful in the university, I think you should, at least, have two points to follow.

A: ⁽¹⁾ **Which two point?**

B: **Oh** |, ⁽²⁾ **first is you should be very good at your academic studies.**

(SECCL: C00-29-15)

(5.5.6)

P: Turn-initial

E: *Oh* prefaces Speaker B's answer, Item (2), to Speaker A's question, Item (1)

F: As a preface to a response to a question

B: What's it? Go ahead!

A: OK! A friend of mine is graduating this year, and would like to ask me for some advice... whether it is a good idea for a high... educate... graduate to go abroad to study. ⁽¹⁾ **What do you think about?**

B: **Oh** |, ⁽²⁾ **in my opinion, I think he'd better stayed at out home!**

(SECCL: C01-67-14)

In NSs' speech, *oh* has been found to preface responses to questions in the London-Lund Corpus, reported by Stenström (1984) and Aijmer (2002: 127-130) and in conversations, identified by Heritage (1984, 1998). Heritage (1998: 294-295) investigates *oh*-prefaced responses to inquiries and reported that *oh*, as a marker of change-of-state, indicates "a problem about a question's relevance, appropriateness, or presuppositions". For instance, in Excerpt (5.5.7) below, Speaker 3's *oh*-prefaced response to Speaker 1's question implies that

Speaker 1's presupposition is problematic.

(5.5.7)

P: Turn-initial

E: *Oh* prefaces Speaker 3's response, Item (2), to Speaker 1's question, Item (1)

F: As a preface to a response to a question

S3: this weekend she said she was gonna be gone. so she said get it to her today. i think

S1: ⁽¹⁾ **did she say she was gonna be gone next week too?**

S3: **oh** | ⁽²⁾ **but, y-**

S2: she said she's leaving Wednesday, until Sunday. [S1: oh okay] so, i'm assuming she'd get it back to us, Monday or Wednesday. [S1: yeah] or, some email (something)

S1: yeah

S3: yep

S2: and then if you get a email from her can you, send it, to us, to see what she says? i i or me i'm just curious. [S1: sure] like what she had to say about it... does anybody have to like, (change) it? or

(MICASE: SGR565SU144)

The investigation of types of co-occurrence of *oh* (see Table 5.16) in the three sub-corpora of the dialogic genres and the further examination of *oh*-prefaced responses to *what*- and *which*-questions (see Table 5.17) reveal that the NNSs tend to use *oh* as a preface to a response to a question and the implications conveyed by *oh* are probably different from NSs' understanding of *oh*. If this finding is generalised to the speech of Chinese NNSs, the distinctive use of *oh* may cause problems in communication when Chinese speakers of English talk with NSs.

5.6 Chapter summary and conclusions

Fraser (1990: 392) notes that *oh* is seldom spoken in a neutral tone. If *oh* is spoken with a rising tone, it may show the speaker's surprise, with a falling tone, disappointment and displeasure and with a rising then falling tone, a warning. Heritage (1984) finds *oh* in his data with falling tone functions as a receipt token. For the present study, prosodic information would have revealed more reliable results. However, the sound files of the NS speech under investigation were not all available and if they had been, the investigation of prosody of each instance of *oh* was beyond the feasible manual work. In addition to prosodic information, facial expressions and gestures, which can be analysed in multi-model corpora, will no doubt contribute to the analysis of the use of *oh* in future studies.

From the observed linguistic evidence, it seems that the Chinese NNSs manage well the use of *oh* co-occurring with the seven types of co-occurrence. Although the use of *oh* with more straightforward co-occurrence, such as reported speech, hesitation markers and responses, occurs more often in the speech of the NNSs than in that of the NSs, the use of *oh* to convey subtle meanings, such as rethinking and recasting, is also found in the NNSs' speech. In Excerpt (5.6.1), from an NS conversation, *oh* introduces the recast utterance by Speaker A. In Excerpt (5.6.2), from an NNS dialogue, Speaker A seems to re-evaluate Item (1) and to replace Item (1) with Item (2).

(5.6.1)

P: Turn-medial

E: Item (1) is recast by Item (2)

F: To indicate a cognitive process has just been completed

A: ⁽¹⁾ **I went to the King's mingle on**

Oh | ⁽²⁾ **I told you all that yeah**

B: Mm <,>

A: That was quite good though <,>

(ICE-GB: S1A-039)

(5.6.2)

P: Turn-medial

E: Item (1) is replaced by Item (2)

F: To indicate a cognitive process has just been completed

B: Ok, in the English department. I will take part in the English department, but my English level is very low, can you give some advice?

A: Ok, no problem. If you want to study English very well, you should know the English foundation is read<red> read<red>, you must read ⁽¹⁾ **some original English books**. |**Oh**], because you are freshman, at first you can read ⁽²⁾ **some easy ones**.

(SECCL: C00-82-13)

In addition to the similar usages identified across these two groups of speakers, one particular use made by the NNSs is to be found. Chinese NNSs seem to have a tendency to use *oh* as an acknowledgment token, which does not carry the implications of *oh* as used by NSs. It is arguable here that the distinctive use of *oh* can be acknowledged as a feature of a non-native variety of English. The global community of English speakers should be educated to accept a local variety, rather than teaching the Chinese NNSs to use *oh* differently.

Using DMs says nothing about right and wrong in relation to syntax and semantics, as shown in Excerpts (5.5.4), (5.5.5) and (5.5.6) above. Neither sentence grammar nor semantics has much to say about DMs. As demonstrated in the analysis of DM *oh*, DMs are probably culture-specific and context-dependent.

As the issue raised in Section 1.4 of Chapter 1, whether NNSs should aim to become native-like in speaking has been a matter of ongoing debate. It is argued that learners of English should not be punished for not speaking like NSs and they can, if they wish, keep

their cultural identity, as long as the use of NNS language leads to a satisfactory development of interpersonal relations.

CHAPTER 6: ANALYSIS OF *WELL*

6.1 Introduction

This chapter discusses the common word *well*. It begins with my hypotheses on the use of *well*, followed by a survey of the literature. As shown in the preceding two chapters on *like* and *oh*, a bottom-up approach is employed. The analysis presents first the frequency data and patterns of all instances of *well* and those of its discourse use (Type B). The major analysis is the discourse aspects of Type B *well*, looking at its positions in utterances/turns and the collocation phenomena surrounding *well*. On the basis of the identified types of co-occurrence, I suggest some functions of Type B *well*.

In the analysis of *like* (Chapter 4), it is found that the Chinese NNSs are more likely than the NSs are to use Type A *like*. It seems that the NNSs are less likely to employ the discourse-use of *like* as the NSs do. In this chapter, I investigate the word *well*, which is similar to the word *like* in Chapter 4 in that there is a clear-cut distinction between Type A *well* and Type B *well* and in that the NNSs are probably more familiar with the use of Type A than Type B. Therefore, I hypothesise that in the NNSs' speech, Type A *well* is primarily used, whereas in the NSs' speech, Type B *well* is predominantly used, as pointed out in the literature (e.g. Stenström (1994) and Biber *et al.* (1999)). If this is so, it is hypothesised that the uses of Type B *well* by the NNSs are not as varied as those by the NSs. The aim is to find out how similar or different the NNSs' and NSs' speech are in the use of *well*. The hypotheses are tested within the framework of the core research questions addressed in Section 1.1.2 of Chapter 1.

The distinction between Type A *well* and Type B *well* is clear-cut and can in most cases be drawn without difficulty. As mentioned in Section 3.3.2, to distinguish Type A *well* from Type B *well* is mostly a straightforward reference to the word classes and co-occurring syntactical structure. Additionally, Sinclair and Mauranen's *Linear Unit Grammar (LUG)* analysis (2006) is used to look at Type B *well*. All these will be reviewed in the next section.

As with *like* in Chapter 4, the results from the manual classification of Types A and B are compared with the frequencies retrieved from ICE-GB. The resemblance shown in Appendix 5 adds credibility to my manual examination process.

6.2 Previous studies of *well*

This section begins with a discussion of the semantic and syntactical aspects of *well*, which are relevant to the use of Type A *well*. Type B *well* cannot be fitted into any traditional grammatical classification, whereas the *LUG* analysis is able to accommodate both Type A and Type B *well*. Examples are given below to demonstrate the roles of both types of *well* in speech.

6.2.1 Semantic aspect of *well*

In most grammar books, five uses of Type A *well* are listed: adverb, adjective, noun, verb and convention (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 1650-1651). *Well* as an adverb, adjective, noun or verb conveys different meanings. As an adverb, *well* modifies a verb placed before it, which means something is done completely, thoroughly, to a high standard or to a great extent, as in Excerpt (6.2.1) below. *Well* as an adverb can be used before an adjective or a prepositional phrase to emphasise them, as in Excerpt (6.2.2). As an adjective, *well* is used to describe someone as healthy, as in Excerpt (6.2.3). Surprisingly, this word class has occasionally been mis-used as *good* by the Chinese NNSs, for example, *her English is very well* (SECCL: B01-30-34) and *you must prepare a well dress* (SECCL: B02-61-18).

(6.2.1)

..... But when I was a child, I was very envy of them that they could **swim** very **well**, but nobody can teach me, could teach me.....

(SECCL: B99-08-33)

(6.2.2)

..... The unusual method and the unusual behavior of my teacher was quite **well-known** soon in the whole middle school.....

(SECCL: B01-100-30)

(6.2.3)

..... Eh..., he said to me that eh... he he had some disease in the before, but now he didn't he was **well**.

(SECCL: B01-100-34)

As a noun, a *well* means a hole in the ground from which water or another liquid can be extracted. As a verb, if a liquid *wells*, it comes to the surface and flows out and if an emotion *wells* in a person, it becomes strong (*Collins COBUILD Advanced Learner's English*

Dictionary 2006: 1650-1651).

Type B *well* is different from the above-mentioned four word classes. DM is not a word class. *Well* as a DM does not carry any literal meanings but serves several pragmatic functions. These roles are discussed on the basis of its collocation phenomena in Section 6.4 below.

6.2.2 Syntactical aspect of *well*

In terms of syntax, *well* as an adverb, adjective, noun or verb behaves as follows:

VERB + *well* as adverb

well as adverb + ADJECTIVE

BE + *well* as adjective

ARTICLE + *well* as noun

SUBJECT + *well* as verb

Well as a DM follows no syntactical rules since DM is not a word class. Although DMs, in most cases, freely mark anywhere in discourse, there are some syntactic restrictions. *Well* as a DM cannot split a compound expression, for example, such as **I'm going to the city well centre*. Nor is it usually placed before a coordinated pronoun, such as **Maria and well I have a meeting this afternoon* (Crystal 1988: 48). In the NNSs' and NSs' speech under investigation, *well* as a DM usually appears in extra-clausal positions, which are further categorised into utterance/turn-initial, -medial and -final positions (see the discussion in Section 6.4.1 below). *Well* occurring in intra-clausal positions is described in the *LUG* analysis (Sinclair and Mauranen 2006).

6.2.3 Linear Unit Grammar analysis of *well*

It is demonstrated below that both Types A and B *well* fit into the *LUG* analysis. *Well* as an adverb in Example (6.2.4) below and *well* as an adjective in Example (6.2.5) are M elements in the *LUG* analysis (see Appendix 4 for a list of the labels in *LUG*). In both cases, the use of *well* makes a propositional contribution.

(6.2.4)

..... If I want to get along well with each other, we must... communicated communicate
M M- +M MA +M-
each other often and... know well.
+M OT M

(SECCL: B02-61-27)

(6.2.5)

..... Eh..., he said to me that eh... he he had some disease in the before, but now he didn't
OI M- OT OI MF +M MS OT M M
he was well. But why he was always wearing the heavy clothes.
M OI M- +M- +M

(SECCL: B01-100-34)

In Example (6.2.6), *well* as a DM is an O element, which does not augment knowledge but makes the discourse flow. It is further categorised as an OI element rather than an OT element because *well* does not create cohesion at the textual level.

(6.2.6)

..... Eh... eh... teachers laughed <left>... laughed and students also laughed, and we have...
OI OI MA M OT M OT MA
and we had a good time. Yeah... good time. Well, I think... I think um... I think we...
OI M OI M OI OI OI OI MF
we pleasant... we were pleasant <present>
MF M

(SECCL: B00-29-23)

6.2.4 Previous studies of *well* as a discourse marker

Well is one of the most widely-discussed DMs. It seems to be beyond dispute that *well* is a central DM, but there have been various approaches to investigating *well*, for example, Svartvik's paraphrasing approach (1980), Schiffrin's coherence-based approach (1987), Jucker's relevance-theoretical approach (1993) and Cuenca's contrastive analysis (2008).

A comprehensive study by Svartvik (1980) looks at the occurrences of *well* in nine conversations extracted from the London-Lund Corpus of Spoken English. The functions of *well* are suggested on the basis of paraphrases in English and Swedish equivalents, not a theory or an existing framework. *Well* is discussed as a "qualifier" and a "frame" (Svartvik 1980: 173). *Well*, as a "qualifier", serves as a connector between the preceding and following contexts. The sub-categories include: 1) agreement (English paraphrases *yes* and *indeed* and Swedish counterparts *ja(visst)* and *jo(visst)*), 2) reinforcement (English paraphrases *actually* and *certainly* and Swedish counterparts *(ju) faktiskt* and *nog*), 3) non-straight and incomplete answer to the *wh*-question and 4) non-direct or qualified answer. This use of *well* is predominantly in turn-initial, linking two turns. *Well* as a "frame" is primarily placed

non-initially for 1) marking the closing of the discourse and focusing the new discourse, 2) introducing explanations and clarifications, 3) introducing quotations and 4) functioning as “editing marker” for self-repairs (Svartvik 1980: 175). All the functions discussed in his study are based on the basic pragmatic use of *well* as a sharing device for maintaining social relationships. As the use of *well* is rather context-dependent, Svartvik acknowledges that *well* performs many more functions than are listed above.

Schiffrin’s coherence-based approach (1987) is derived from a corpus of sociolinguistic interviews. In the participation framework that she proposes, *well* acts as a signal to interlocutors when the coming utterance lacks coherence (e.g. disagreement and insufficient or unexpected answers). *Well* is mainly a marker of response, prefacing a response to a *wh*-question more often than a *yes/no* question. Schiffrin claims that *yes-no* questions often require either affirmative or negative responses, whereas *wh*-questions offer more options for interlocutors, thereby requiring more cognitive work. A similar finding is reported in Svartvik’s study (1980: 169). It indicates that prefatory *well* occurs more frequently with responses to *wh*-questions than to *yes/no* and tag questions. However, in some *yes/no* questions, the responses are not merely a *yes* or *no*. (Examples are given in Section 6.5.2.10 below.)

Jucker (1993) re-investigates, in terms of relevance theory (see Section 2.3.2 of Chapter 2 for more details), examples of *well* analysed in previous studies. In his study, four uses of *well* are identified: 1) “as marker of insufficiency”, 2) “as a face-threat mitigator”, 3) “as a frame marking device” and 4) “as a delay device”. These uses of *well* indicate “a shift in the relevant context” and orient the interlocutors to process the following utterance in a new or adjusted cognitive environment which may be against the interlocutor’s expected background or the background developed by the previous utterance (Jucker 1993: 451).

Unlike Schiffrin’s (1987) and Jucker’s (1993) studies and similar to Svartvik’s approach (1980), Cuenca’s work (2008) adopts an indirect approach to the investigation of *well*, conducting a contrastive analysis on the instances of *well* in a film transcription and their functional counterparts in the Catalan and Spanish languages. Although this study is limited to the instances of *well* in one film transcription, *well* is analysed in a broader context and compared with two translated versions. This study indicates that the meanings of *well* can be organised in two broad “spaces”: modal and structural functions. The former includes the feature of “downtoning” and contributes to interactional meanings (e.g. agreement) and the

latter includes the feature of “continuity” and makes a contribution to the textual meanings (e.g. change of topic) (Cuenca 2008: 1388).

In addition to the four above approaches, *well* is also documented in the work of Biber *et al.* (1999), Schourup (2001), Müller (2005) and Carter and McCarthy (2006). The limited space in this thesis does not allow more detailed discussion of the studies of *well*. The major functions of *well* which have been identified in previous studies are summarised below.

Well as a DM is found as a device for structuring the discourse, for example, in managing topics and reported speech and indicating digression and continuation in speech.

Well is one of DMs used for speakers to initiate the discourse and open a topic (Svartvik 1980: 175, Biber *et al.* 1999, Carter and McCarthy 2006: 214-215) as well as closing a topic (Svartvik 1980: 175, Carter and McCarthy 2006: 214-215) or ending a conversation (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 1651).

Well marking the beginning of reported speech has been widely discussed in previous studies, such as Svartvik (1980: 175), Schiffrin (1987: 124), Fraser (1990: 389), Redeker (1990: 374), Jucker (1993: 438) and Müller (2005: 113-115). It is generally acknowledged that it is unknown whether or not the instance of *well* has been spoken in the original utterance.

Well is used to indicate a shift of topic or signal a change in the direction of the discourse that might have been expected by the interlocutor (Carter and McCarthy 2006: 212, 219). In narratives, *well* may indicate a shift of episodes (Watts 1989: 853).

Well can be interpreted as an indication of continuation, suggesting that something is about to be said (Svartvik 1980: 175, Biber *et al.* 1999: 1087, Schourup 2001: 1043, *Collins COBUILD Advanced Learner's English Dictionary* 2006: 1650). It can be pragmatically seen as a floor holder.

The above four uses of *well* are used mainly for structuring the discourse. The following three uses contribute to the construction of a proposition. First, *well* signals the following contrasting points, disagreement, unexpected and surprising answers (Biber *et al.* 1999: 1087, Carter and McCarthy 2006: 153, 727). *Well* in this use can be taken as a mitigator for softening the impact of the speech. Second, *well*, usually in intra-clausal position, to indicate self-repair, has been reported in a number of studies (e.g. Svartvik (1980: 175), Schiffrin (1987) and Biber *et al.* (1999)). It can also indicate revision or rephrasing (Carter and McCarthy 2006: 153). Third, *well* in extra-clausal position is used to emphasise a key point in

the narrative. In intra-clausal position, *well* may be placed before a key point, word or phrase. Schourup (2001: 1038-1039) compares *well* with *after all* and *moreover* and finds that only *well* can be used “to ‘focus down’ on the choice of a single word or phrase”.

It is reasonable to assume that the frequencies of the uses of *well* vary across types of genre. For example, it is likely that, in conversations, there are more opportunities for speakers to use *well* for opening and closing a topic and, in narratives, there may be more instances of *well* indicating the shift of episodes and prefacing the coda of the story. The NNSs’ monologues under investigation are mainly narratives; however, they are peculiar in that the speaker is asked to retell an experience of theirs or a particular event, which are not voluntary and spontaneous narratives. It is possible in them the NNSs do not feel driven to attract the listener’s attention. The NNSs’ dialogues are produced according to the rubrics set up in the oral examinations. The rubrics assign the speakers to take a position either for or against a given proposition. Thus, it will not be surprising if *well* prefacing disagreement occurs frequently in the NNSs’ dialogues. Due to the variations in the types of activity in the six sub-corpora, it is anticipated that the frequency of *well* will vary to some extent across sub-corpora.

On the whole, some functions of *well* as a DM are widely regarded as typical (e.g. signalling unexpected and insufficient answers and introducing negative opinions and contrary positions) and some are differently reported, mainly because of the differing approaches adopted in the studies. Some of the previous studies of *well* are based on specific theories (e.g. Jucker (1993)) and some use self-established or existing frameworks (e.g. Carlson (1984) and Schiffrin (1987)). Instances of *well* for analysis in the literature come from various sources, such as contrived language, corpus data, film transcripts, etc. The functions identified in the above-mentioned studies provide an overall sense of the use of *well* as a DM and the background knowledge for the present study, although the categories for discussion in the present study are not based on any existing frameworks or schemas. Due to the different contexts in the six sub-corpora under investigation, the uses of Type B *well* identified in the speech of the NNSs and NSs are not limited to those described in the literature.

6.3 Frequency information in the speech of the non-native speakers and native speakers

6.3.1 Overall frequency of *well*

The overall frequency of *well* is shown in Table 6.1 below. The NNSs' monologues and dialogues reveal respectively 512 and 1,384 occurrences of *well*. The raw frequencies are normed on a basis of 10,000 words. The normalised frequencies show that *well* is more often used in the dialogues than in the monologues (23.2 vs. 15.2 times per 10,000 words). The same trend is also found in the NSs' speech. There are more instances of *well* in the dialogic genres than in the monologic genres in MICASE and ICE-GB.

As with the word *like* in Chapter 4, the distinction between Type A *well* and Type B *well* can, in most cases, be drawn without difficulty. (A dividing line, however, between Type A and Type B in the cases of *you know*, *I mean*, *you see* and *I think* is more difficult to establish. These markers will be discussed in turn in the next two chapters.) The instances of *well* are manually grouped into Types A and B. This classification reveals that *well* is not primarily used as a DM in the NNSs' monologues and dialogues and the NSs' monologic genres. The percentages of Type B in these four sub-corpora, shown in Table 6.1, range from 16% to 52%. In contrast, 90% and 85.7% of the occurrences of *well* in the two sub-corpora of the NSs' dialogic genres are used as a DM. This finding corresponds to work done by Biber *et al.* (1999: 1096), which indicates that *well* is primarily a DM in American and British conversations.

Table 6.1: Frequency information of *well* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percent-age (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	512	15.2	48 out of 300 ^a	16.0	2.4
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	1,384	23.2	112 out of 300 ^b	37.3	8.7
MICASE: 13 transcripts of highly monologic (American NSs)	134,096	200	14.9	104	52.0	7.8
MICASE: 48 transcripts of highly interactive (American NSs)	577,996	2,116	36.6	270 out of 300 ^c	90.0	32.9
ICE-GB: 70 unscripted monologues (British NSs)	153,646	353	23.0	158	44.8	10.3
ICE-GB: 90 private direct conversations (British NSs)	185,000	1,521	82.2	257 out of 300 ^d	85.7	70.5

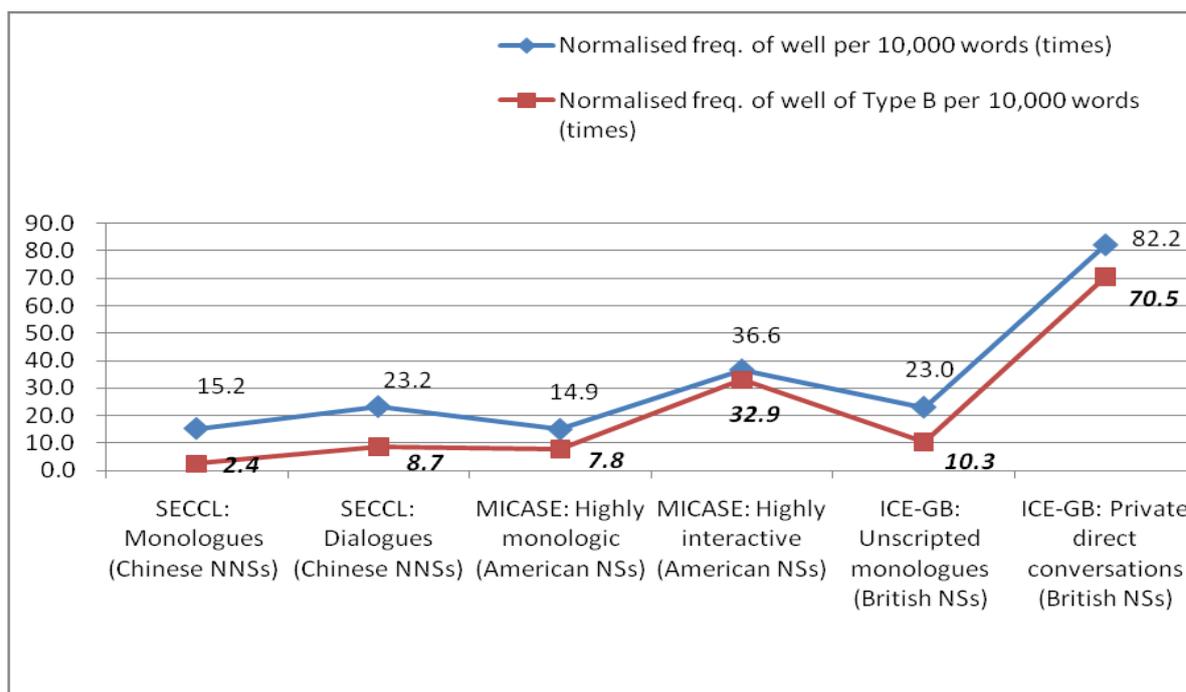
* Three sets of random samples of 100 are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B per 10,000 words are based on an extrapolation of the percentages of the Type B word.

a, b, c and d in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

The raw frequencies of *well* are normed on a basis of 10,000 words and the normalised incidence, ranging from 14.9 to 82.2 times across the six sub-corpora, is shown in Table 6.1 above. The same normalisation is used on the frequencies of Type B *well*, which range from 2.4 to 70.5 times. Figure 6.1 below shows a comparison of the normalised frequencies of *well*. It can be clearly seen from the two W-shaped curves that there are more instances of *well* in the dialogic genres than in the monologic genres. This supports the hypothesis that the more interactive the genres or types of activity are, the more DMs occur.

Figure 6.1: Comparison of normalised frequencies of *well* across sub-corpora



The results of the statistical test shown in Appendix 6 indicate that the differences between two types of genres and between two groups of speaker are sometimes highly significant. The LL scores between the monologic genres and the dialogic genres are very high with -966.54 between Corpora A1 and A2, -969.87 between Corpora B1 and B2 and -1047.74 between Corpora C1 and C2. These negative scores indicate the under-representation of DM *well* in the sub-corpora of the monologic genres. Interestingly, *well* is over-represented in the Chinese NNSs' monologues, as opposed to the two NS sub-corpora (LL: +142.58 between Corpora A1 and B1; LL: +112.27 between Corpora A1 and C1), while it is under-represented in the Chinese NNSs' dialogues (LL: -891.44 between Corpora A2 and B2; LL: -1864.41 between Corpora A2 and C2).

6.3.2 Collocates of *well*

Tables 6.2 to 6.7 present the patterns of *well* in the six subsets extracted from SECCL, MICASE and ICE-GB. The two patterns of the NNSs' monologues and dialogues (see Tables 6.2 and 6.3) reveal similar collocates. The collocates immediately to the left, *very*, *as*, *do*, *study*, *sing* and *quite*, shown in boldface, are indications of the use of Type A *well*. Such fixed expression as *as well (as)* and such phrasal verb as *get along* also suggest that *well* is Type A. These correspond to my manual classification of the three sets of random samples of *well*

extracted from the two NNS sub-corpora. Most of the instances of *well* in the NNSs' speech are Type A.

Well as an adverb also co-occurs with *very*, representing about 33% (96 out of 293 instances in Table 6.2) in the monologues and 19% (57 out of 297 instances in Table 6.3) in the dialogues. As an adverb, *well* modifies the preceding verbs, such as *do*, *study*, *learn*, *sing* and *get*. These verbs reflect that the topics in the monologues and dialogues are student-related.

Further investigation reveals that, in Table 6.3, the collocates to the left, *a* and *b*, are mostly the identification of the two speakers in the dialogues. They indicate that *well* is used in turn-initial position. It is very likely that *well* in this position is Type B.

Table 6.2: Pattern of *well* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (31)	i (20)	do (17)	very (96)	well (293)	and (43)	i (32)	i (13)	i (22)
2	the (13)	to (18)	learn (15)	as (17)		i (27)	the (14)	the (13)	and (15)
3	we (9)	can (14)	i (11)	do (13)		as (17)	my (13)	in (9)	was (9)
4	to (9)	you (11)	to (10)	it (12)		in (16)	he (9)	other (8)	in (8)
5	she (7)	the (10)	it (9)	along (11)		with (15)	was (8)	very (8)	the (8)
6	want (6)	he (8)	get (8)	quite (11)		but (12)	each (8)	but (7)	that (6)
7	my (6)	could (7)	is (7)	english (8)		eh (10)	and (7)	to (6)	um (6)
8	play (6)	a (7)	very (7)	study (6)		at (9)	then (6)	so (5)	to (6)
9	and (6)	eh (6)	me (6)	sing (5)		because (9)	we (6)	is (5)	have (5)
10	well (6)	learn (6)	will (6)	a (4)		you (8)	you (5)	have (5)	when (5)

Table 6.3: Pattern of *well* in the non-native speakers' dialogues in SECCL

	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	you (20)	can (27)	get (21)	very (57)	well (297)	i (41)	i (28)	think (21)	a (25)
2	i (15)	to (15)	do (19)	b (42)		with (29)	think (22)	i (17)	i (15)
3	the (11)	the (14)	it (15)	a (32)		in (20)	the (16)	you (15)	the (14)
4	can (11)	b (10)	to (14)	on (15)		a (16)	you (14)	a (13)	you (9)
5	to (10)	you (10)	a (11)	study (13)		but (15)	know (10)	b (10)	think (9)
6	they (9)	is (10)	can (10)	do (12)		and (14)	my (10)	the (8)	and (9)
7	a (8)	do (9)	job (9)	it (9)		b (14)	them (7)	um (7)	b (8)
8	how (8)	we (7)	not (8)	as (9)		you (14)	your (7)	it (6)	so (8)
9	do (7)	i (7)	english (7)	quite (7)		eh (9)	so (6)	but (6)	eh (6)
10	think (7)	of (6)	b (5)	along (7)		so (7)	a (6)	eh (5)	in (6)

In the NSs' monologic genres in MICASE and ICE-GB, most collocates immediately to the left of *well* (see Tables 6.4 and 6.5), such as *very*, *as*, *perfectly*, *pretty*, *really* and *extremely*, reveal that *well* is Type A.

Table 6.4: Pattern of *well* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (12)	the (6)	to (9)	as (34)	well (196)	as (15)	you (13)	the (9)	that (9)
2	of (9)	we (5)	and (5)	very (14)		if (11)	the (11)	a (6)	a (8)
3	what (7)	and (5)	it (4)	it (4)		the (10)	a (7)	that (6)	the (7)
4	a (6)	this (5)	the (4)	is (4)		in (9)	we (6)	you (5)	in (6)
5	in (4)	about (4)	of (4)	was (4)		that (7)	in (5)	is (5)	of (5)
6	it (4)	of (3)	work (4)	perfectly (3)		you (7)	of (5)	know (4)	if (5)
7	you (4)	or (3)	in (3)	pretty (3)		uh (6)	is (4)	to (4)	just (5)
8	to (3)	would (3)	is (3)	see (3)		what (5)	this (4)	not (4)	this (4)
9	was (3)	you (3)	was (3)	say (3)		so (5)	not (4)	and (3)	to (4)
10	so (3)	to (3)	uh (3)	really (3)		i (4)	know (4)	one (3)	very (3)

Table 6.5: Pattern of *well* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (15)	the (21)	the (17)	as (62)	well (343)	as (17)	the (28)	the (21)	to (13)
2	a (12)	a (12)	it (10)	uh (15)		the (16)	s (15)	is (9)	that (12)
3	in (11)	that (11)	and (9)	very (11)		i (12)	is (12)	a (7)	the (12)
4	to (10)	uh (11)	he (6)	may (10)		and (12)	uh (9)	it (7)	of (10)
5	of (9)	of (10)	to (6)	is (8)		in (12)	that (9)	be (6)	and (10)
6	and (9)	to (8)	s (6)	and (6)		it (11)	a (8)	and (6)	a (9)
7	is (6)	it (7)	a (6)	extremely (5)		this (9)	to (8)	uh (6)	in (7)
8	s (5)	in (6)	that (5)	a (5)		that (9)	was (8)	you (6)	uh (7)
9	he (5)	for (5)	is (5)	it (5)		there (8)	of (6)	not (5)	is (6)
10	uh (5)	and (5)	this (5)	uhm (4)		he (8)	it (5)	that (5)	he (5)

The collocates immediately to the left of *well* in Tables 6.6 and 6.7 (the NSs' dialogic genres) are rather different from those in Tables 6.4 and 6.5 (the NSs' monologic genres). The collocates shown in both boldface and italics, *like*, *okay*, *so*, *say*, *um*, *uh*, *uhm*, *yeah*, *yes* and *no*, suggest that *well* is used as Type B. This supports the findings of the manual examination. As shown in Table 6.1, over 85% of the occurrences of *well* in the NSs' dialogic genres revealed as DMs.

Table 6.6: Pattern of *well* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	we (5)	and (7)	you (6)	as (12)	well (191)	i (29)	you (20)	that (12)	the (11)
2	that (4)	paragraph (3)	to (4)	like (7)		it (13)	i (11)	i (11)	you (8)
3	particular (3)	people (3)	might (3)	do (6)		we (11)	the (10)	a (9)	it (8)
4	so (3)	i (3)	like (3)	okay (4)		but (10)	was (9)	you (9)	a (7)
5	think (3)	the (3)	two (3)	very (4)		that's (9)	is (9)	it (6)	that (7)
6	you (3)	about (3)	or (3)	so (3)		if (9)	think (8)	actually (6)	have (5)
7	uh (2)	to (3)	it (3)	say (3)		it's (9)	we (7)	we (5)	is (5)
8	the (2)	might (2)	but (3)	um (3)		okay (8)	what (6)	is (5)	on (5)
9	to (2)	um (2)	be (2)	uh (3)		you (8)	a (5)	like (5)	with (5)
10	in (2)	that (2)	they (2)	that (3)		what (8)	are (5)	not (4)	do (5)

Table 6.7: Pattern of *well* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (14)	and (11)	i (16)	as (26)	well (299)	i (76)	s (37)	i (23)	i (17)
2	i (11)	the (10)	it (12)	yeah (16)		it (26)	i (13)	s (11)	it (10)
3	a (11)	i (9)	a (8)	know (10)		that (14)	it (12)	not (9)	a (9)
4	the (9)	of (9)	the (7)	uhm (9)		you (14)	was (10)	the (9)	to (9)
5	to (7)	s (9)	you (7)	uh (9)		yes (10)	think (10)	it (8)	s (9)
6	you (6)	to (8)	s (7)	yes (9)		no (9)	you (10)	is (7)	the (8)
7	was (6)	so (6)	or (7)	it (8)		we (8)	don't (9)	that (7)	that (8)
8	know (6)	it (6)	yeah (6)	said (8)		they (8)	m (8)	got (6)	you (8)
9	uhm (5)	that (5)	that (6)	no (7)		she (7)	can (6)	think (6)	don't (5)
10	s (5)	do (5)	of (4)	very (6)		he (7)	mean (6)	have (6)	in (5)

The patterns shown in Tables 6.2 to 6.7 reveal the tendency of *well* towards either Type A or Type B and this tendency supports the conclusion drawn from the manual classification of Type A *well* and Type B *well*.

To learn more about the use of Type B, the patterns of Type B *well* in the NNSs' and NSs' speech are produced. In the NNSs' monologues (see Table 6.8), there are no prominent collocates. The frequencies of other collocates to the right/left are probably too low to suggest the use of *oh*.

In Table 6.8, the collocate immediately to the left, 2, refers to the beginning of the transcription. 4 (8.3%) out of 48 instances of *well* are used to open the monologues. In Table 6.11, the first two collocates immediately to the left, *b* (42) and *a* (31), mostly refer to the identification of the speakers in the dialogues. Therefore, 73 (65%) out of 112 occurrences in the dialogues open a turn. In spite of these different percentages (8.3% vs. 65%), it could not be concluded that *well* is used more often to open an utterance in the dialogues than in the monologues, because, as explained in Chapter 3, each monologue is taken as a single utterance in the analysis, while there are many utterances/turns in each of the dialogues.

Table 6.8: Pattern of Type B *well* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (5)	of (3)	i (6)	2 (4)	well (48)	i (8)	i (6)	the (3)	was (4)
2	she (4)	a (3)	good (4)	said (3)		eh (4)	s (3)	to (3)	i (4)
3	the (3)	to (2)	said (3)	ok (2)		and (4)	the (2)	point (2)	you (3)
4	a (3)	the (2)	very (3)	um (2)		it (3)	that (2)	saw (2)	we (2)
5	well (2)	would (2)	that (2)	well (2)		one (2)	what (2)	very (2)	that (2)
6	said (2)	very (2)	say (2)	us (2)		we're (2)	we (2)	easy (2)	another (2)
7	i (2)	teacher (2)	birthday (2)	music (2)		what (2)	will (2)	company (2)	interested (2)
8	classes (2)	happy (2)	jazz (2)	jenny (2)		it's (2)	special (2)	have (2)	people (2)
9	night (2)	anybody (2)		friends (2)		next (2)	all (2)	job (2)	
10	me (2)	said (2)		knew (2)		now (2)	a (2)	hold (2)	

Table 6.9: Pattern of Type B *well* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (6)	b (8)	a (11)	b (42)	well (112)	i (32)	think (18)	think (12)	the (10)
2	in (6)	the (7)	it (6)	a (31)		you (9)	i (16)	i (5)	a (8)
3	the (5)	your (5)	that (3)	oh (5)		but (8)	know (6)	that (4)	have (5)
4	you (5)	a (4)	idea (3)	well (5)		it (4)	the (5)	a (4)	i (4)
5	a (5)	of (4)	so (3)	eh (4)		maybe (4)	have (5)	you (4)	to (3)
6	think (4)	original (3)	b (3)	professors (3)		eh (4)	you (3)	my (4)	this (3)
7	than (3)	to (3)	plan'so'er (2)	ah (2)		to (3)	maybe (3)	they (3)	their (3)
8	of (3)	about (3)	study (2)	but (2)		in (3)	some (2)	it (3)	that (3)
9	decision (3)	well (2)	to (2)			then (3)	seem (2)	eh (3)	eh (3)
10	i (3)	this (2)	time (2)			there (2)	we've (2)	but (3)	she (2)

In the patterns of the NSs' speech, shown in Tables 6.10 to 6.13, it is found that Type B *well* co-occurs with the quoting verb *SAY*. This suggests that *well* is used to mark the boundary between the speaker's own utterance and quoted speech. This use is further discussed below in Section 6.4.2.2. *Well* is also found to co-occur with the hesitation markers *uh* and *uhm*. However, the frequencies are too low to indicate any importance.

Table 6.10: Pattern of Type B *well* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (7)	the (5)	to (5)	it (4)	well (100)	if (9)	you (7)	is (4)	a (7)
2	what (5)	is (3)	the (4)	see (3)		the (8)	we (6)	that (4)	that (5)
3	a (4)	this (3)	and (4)	say (3)		you (6)	know (4)	you (4)	know (3)
4	to (2)	we (3)	it (3)	ask (3)		what (5)	the (4)	a (3)	just (3)
5	what's (2)	about (3)	asked (2)	was (3)		i (4)	of (3)	know (3)	this (3)
6	we (2)	would (3)	these (2)	this (2)		it's (4)	are (3)	to (3)	the (3)
7	that (2)	and (3)	come (2)	that (2)		it (3)	i (3)	so (2)	in (3)
8	do (2)	way (2)	you (2)	saying (2)		they (3)	a (3)	want (2)	if (3)
9	but (2)	do (2)	we (2)	here (2)		that (3)	was (2)	the (2)	have (2)
10	he (2)	a (2)	two (2)	cells (2)		for (2)	this (2)	i (2)	part (2)

Table 6.11: Pattern of Type B *well* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (10)	the (7)	the (15)	uh (13)	well (148)	the (12)	s (13)	the (12)	a (6)
2	to (6)	that (6)	and (8)	uhm (4)		it (10)	is (11)	is (7)	of (5)
3	in (4)	of (5)	to (4)	well (3)		i (9)	was (6)	a (5)	is (4)
4	a (4)	a (5)	a (4)	really (2)		this (7)	i (4)	it (4)	in (4)
5	s (3)	to (4)	this (3)	ok (2)		that (7)	of (4)	just (4)	we (4)
6	of (3)	at (3)	with (3)	say (2)		my (6)	the (4)	that (4)	the (4)
7	uhm (3)	on (3)	at (3)	the (2)		there (5)	uh (3)	not (4)	that (4)
8	uh (3)	uh (3)	it (3)	says (2)		he (5)	are (3)	ladies (3)	and (4)
9	there (3)	for (3)	which (2)	esterases (2)		a (5)	you (3)	be (3)	to (3)
10	do (3)	in (3)	use (2)	header (2)		in (5)	a (3)	to (2)	uh (2)

Table 6.12: Pattern of Type B *well* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	we (4)	and (7)	to (4)	like (7)	well (161)	<i>i</i> (29)	you (19)	that (12)	the (10)
2	particular (3)	the (3)	it (3)	okay (4)		it (12)	<i>i</i> (11)	<i>i</i> (10)	it (8)
3	you (3)	paragraph (3)	like (3)	that (3)		we (11)	is (9)	you (9)	you (8)
4	that (2)	um (2)	two (3)	uh (3)		it's (9)	the (9)	a (9)	a (7)
5	the (2)	to (2)	but (3)	so (3)		that's (9)	was (8)	actually (6)	that (7)
6	uh (2)	do (2)	and (2)	say (3)		if (8)	think (8)	is (5)	have (5)
7	to (2)	already (2)	be (2)	me (3)		the (8)	we (7)	like (5)	with (5)
8	so (2)	is (2)	you (2)	um (3)		you (8)	what (5)	we (5)	on (5)
9	in (2)	going (2)	we (2)	oh (3)		okay (7)	a (5)	it (4)	like (4)
10	<i>i</i> (2)		was (2)	xx (2)		what (7)	this (5)	not (4)	is (4)

Table 6.13: Pattern of Type B *well* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (14)	and (10)	<i>i</i> (15)	yeah (16)	well (256)	<i>i</i> (74)	s (36)	<i>i</i> (18)	<i>i</i> (15)
2	<i>i</i> (10)	s (8)	s (7)	know (10)		it (26)	<i>i</i> (11)	not (9)	it (10)
3	the (7)	<i>i</i> (8)	the (7)	uh (9)		you (14)	was (10)	it (8)	a (9)
4	was (6)	to (8)	it (7)	uhm (9)		that (14)	think (10)	the (8)	to (8)
5	to (6)	the (7)	or (7)	yes (9)		yes (9)	don't (9)	is (7)	s (8)
6	a (6)	so (6)	you (7)	it (8)		they (8)	it (9)	s (7)	the (7)
7	know (5)	it (6)	yeah (6)	said (8)		we (8)	you (8)	think (6)	that (7)
8	that (5)	of (5)	a (5)	no (7)		she (7)	m (8)	all (5)	don't (5)
9	you (5)	you (4)	that (5)	ah (5)		no (6)	ve (6)	that (5)	you (5)
10	uh (5)	do (4)	mmm (4)	well (4)		there (6)	mean (6)	they (5)	good (4)

In the patterns shown in Tables 6.8 to 6.13, the DM collocations are identified. There is clearly some preference for the orders of DM collocations, e.g. *well, anyway, let's...* rather than *anyway, well, let's...* (Fraser 1990: 395). Carter and McCarthy (2006: 153) report that *well* occurs in clusters, such as *well actually, well as a matter of fact, well well well, well really* and *oh well*. In the NNSs' dialogues (see Table 6.9), some DM collocations, *oh well, well I think, well you know* and *well but*, are found and in the NSs' speech (see Tables 6.10 to 6.13), the DM collocations used are *well well, ok well, oh well, yeah well* and *so well*.

6.4 Discourse aspects of *well*

In this section, the positions in an utterance/turn where Type B *well* occurs are first described and then the linguistic items which Type B *well* tends to co-occur with are discussed. All the occurrences of Type B *well* in the NSs' highly monologic discourse mode in MICASE (104) and the unscripted monologues (158) in ICE-GB were manually analysed, but the sheer number of occurrences of *well* in the NNSs' monologues (512) and dialogues (1,384), the NSs' highly interactive discourse mode in MICASE (2,116) and the private direct conversations in ICE-GB (1,521) made them unmanageable for manual analysis; therefore,

the random sampling procedure described in Section 3.3.6 of Chapter 3 was used for the manual examination.

6.4.1 Positions in an utterance/turn

In this section the positions of Type B *well* in an utterance/turn are discussed. Its distribution and percentages in the six sub-corpora under investigation are shown in Table 6.14 below. In general, it is typical for Type B *well* to occur in extra-clausal positions. Most of the instances appear in extra-clausal utterance-medial position in the three monologic genres and in extra-clausal turn-initial position in the three dialogic genres. There is no marked difference in the distribution of the positions in an utterance/turn of Type B *well* across the two types of genre and between the two groups of speakers, except a difference in proportion between the NNSs' monologues and dialogues. In the NNSs' monologues, about two-thirds of the occurrences of *well* occur in extra-clausal positions (66.7%), whereas a much higher proportion, 92.9%, occurs in the NNSs' dialogues.

Table 6.14: Distribution of the positions of Type B *well* in an utterance/turn

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode			ICE-GB (NSs): Unscripted monologues	
	Random samples (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	
Positions in an utterance of <i>well</i>	48	100	104	100	158	100	
Extra-clausal: utterance-initial	4	8.3	10	9.6	14	8.9	
Extra-clausal: utterance-medial	28	58.3	67	64.4	121	76.6	85.4
Extra-clausal: utterance-final	0	0.0	0	0.0	0	0.0	
Intra-clausal: after an M-	7	14.6	15	14.4	10	6.3	
Intra-clausal: after an MA	1	2.1	0	0.0	0	0.0	
Intra-clausal: after an MF	1	2.1	7	6.7	6	3.8	
Intra-clausal: others	7	14.6	5	4.8	7	4.4	
Unclassified	0	0.0	0	0.0	0	0.0	
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode			ICE-GB (NSs): Private direct conversations	
	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	
Positions in a turn of <i>well</i>	112	100	270	100	257	100	
Extra-clausal: turn-initial	90	80.4	176	65.2	168	65.4	
Extra-clausal: turn-medial	14	12.5	51	18.9	57	22.2	87.5
Extra-clausal: turn-final	0	0.0	4	1.5	0	0.0	
Intra-clausal: after an M-	2	1.8	21	7.8	16	6.2	
Intra-clausal: after an MA	0	0.0	1	0.4	0	0.0	
Intra-clausal: after an MF	4	3.6	15	5.6	9	3.5	
Intra-clausal: others	2	1.8	2	0.7	7	2.7	
Unclassified	0	0.0	0	0.0	0	0.0	

6.4.1.1 *Well* in extra-clausal position

Generally, in the NNSs' and NSs' monologic genres under investigation, most of the instances of Type B *well* are placed in the extra-clausal utterance-medial position, as shown in Excerpt (6.4.1).

(6.4.1)

P: Utterance-medial the last point, is simply to say, that, it really has to translate to, reproductive success, it's not just that it helps you, uh, uh do well, but it helps it increases the chances of your offspring surviving, and, and your offspring in turn reproducing. **[well]** the last assumption just says <SEG TYPE="READING"> species produce more offspring than survive to be adults </SEG> really, that some organisms die before procreating themselves.

(MICASE: LEL500JU034)

In the three dialogic genres, a large proportion of the instances is found in the extra-clausal turn-initial position, as Excerpt (6.4.2) exemplifies:

(6.4.2)

P: Utterance-initial

A: What do you think about it?

B: **Well**, for me I just... you know... I don't agree... to this kind of... trend, because I think... eh... .. high school graduates are very young and they are not psychologically mature. I don't think they should go abroad during this period of time. I just think they... should finish... the college in China then they can go abroad.

(SECCL: C01-50-31)

It can be seen in Table 6.14 above, across the six sub-corpora, that only in four instances of *well* is the word placed in turn-final position in the sub-corpus of the highly interactive discourse mode in MICASE. Of these four instances, three are preceded by *oh* and one instance co-occurs with hesitation markers.

6.4.1.2 *Well* in intra-clausal position

Across the six sub-corpora, a small proportion, 13% on average, of the occurrences of *well* is placed in an intra-clausal position, mostly occurring after an M- element, as Excerpt (6.4.3) exemplifies:

(6.4.3)

P: M- + *well* + +M

..... according to the ship, the star and the earth are rushing past it and that distance is contracting by that same factor of ten. and so the people aboard the ship say |well| the star isn't a hundred light years away. it's only ten light years away.

(MICASE: LEL485JU097)

6.4.2 Contexts where Type B *well* tends to occur

The positions of Type B *well* in an utterance/turn are described in the previous section and the results are referred to in the present discussion of the contexts where *well* tends to occur. The types of co-occurrence of *well* are discussed in order of the strength of evidence, from the stronger linguistic evidence to intuitive interpretation. As mentioned earlier, although linguistic evidence is used to decide categories for discussion, sometimes intuition is inevitably used to interpret the use of DMs.

Type B *well* is found to co-occur with the following ten categories: 1) hesitation markers, pauses, repetitive words and restarts, 2) reported speech, 3) repairs and rephrasing, 4) opening/closing of a topic and concluding remarks, 5) questions, 6) transitions and shifts of topic, 7) as a preface to a response, 8) disagreement and negative evaluation, 9) key information and 10) as a continuer.

each of the four NS sub-corpora.

It is impossible to specify why the speaker hesitates, but possible reasons are 1) the speaker hesitates or pauses on purpose to hold the floor while searching for content or lexis and to sound less direct and 2) the speaker is interrupted, thereby being forced to formulate his/her ideas again.

6.4.2.2 *Well* co-occurring with reported speech

Well co-occurring with reported speech marks the boundary between the speaker's own utterance and the speech being quoted, as shown in Excerpt (6.4.6). This use has been widely discussed in previous studies, such as Svartvik (1980: 175), Schiffrin (1987: 124), Fraser (1990: 389), Redeker (1990: 374), Jucker (1993: 438) and Müller (2005: 113-115).

(6.4.6)
P: M- + *well* + +M And the other person said, "Well, Who's that boy? There is no boy at all.
E: Quoting verb *said* That's... That's Jay's friend. That's the Bona."
F: To signal reported speech (SECCL: B02-01-03)

In the NNSs' monologues under investigation, *well* co-occurring with reported speech represents 33.3%, but there is no such instance in the NNSs' dialogues. The figures in the NSs' speech are not as high as that in the NNSs' monologues. In the NSs' highly monologic and highly interactive discourse mode in MICASE, this use accounts for 14.4% and 4.1% respectively. In the unscripted monologues and the private direct conversations in ICE-GB, it represents 3.8% and 4.3% respectively. In SECCL and MICASE, *well* marking reported speech is more frequently used in the monologic genres.

6.4.2.3 *Well* co-occurring with repairs and rephrasing

Well is found to be followed by a correction or rephrasing of the term or statement before *well*. In Excerpt (6.4.7) below, *well* prefaces the correction of Item (1). *Well* seems to suggest that a repair has been done and it serves to mark a repair.

(6.4.7)
P: MF + *well* + OI A: Now she doesn't want <,> to do this sort of thing a lot if you see what I
E: Item (2) repairs mean <,> and uh <,> ⁽¹⁾ she does like |well| I think ⁽²⁾ she actually likes it
Item (1) but <,> has a sense of proportion.
F: To signal a repair (ICE-GB: S1A-061)

A number of previous studies have indicated that *well*, particularly in intra-clausal position, signals a self-repair (Svartvik 1980: 175, Schiffrin 1987: 123, Biber *et al.* 1999: 1987) and revision or rephrasing (Carter and McCarthy 2006: 153). This is almost never used by the NNSs, with only one instance in the data and it is one of the least frequent types of co-occurrence in the NNSs' speech, with 7.7% and 3.3% in the highly monologic and highly interactive discourse mode in MICASE respectively. In ICE-GB, it accounts for 4.4% and 3.1% in the unscripted monologues and the private direct conversations, respectively.

6.4.2.4 *Well* co-occurring with the opening and closing of a topic and concluding remarks

Some DMs are used for speakers to organise the discourse (Svartvik 1980: 175, Carter and McCarthy 2006: 214). *Well* as a DM may be used to open an utterance/turn and provide orientation for the listener (Biber *et al.* 1999: 1074, 1086-1087). In Excerpt (6.4.8) below, Speaker A uses *well* to begin a new topic and in Excerpt (6.4.9), Speaker A begins his/her unscripted speech with *well*.

(6.4.8)

<p>P: Turn-initial E: Item (1) is a new topic F: To mark the opening of a topic</p>	<p>Task 3 A: Hello, Tom. B: Hello, James. A: Well, ⁽¹⁾ as a new freshman, eh... I have been iner..... this campus for almost a week. And I think everyone everything here iser... new and attractive to me, but I also have some problems. B: Yeah.</p>
--	--

(SECCL: C00-11-23)

(6.4.9)

<p>P: Utterance-initial E: <i>Well</i> at the beginning of the utterance F: To mark the opening of a topic</p>	<p>A: Well Chairman ladies and gentlemen uhm I'm always very suspicious when people uh introduce me uh as being a national figure since I've never forgotten uh the first time I uh appeared on Question Time <></p>
---	---

(ICE-GB: S2A-023)

In addition to *well* occurring at the opening of a topic, it occurs at the closing of a topic. To organise the discourse, speakers can use *well* to close a topic (Svartvik 1980: 175, Carter and McCarthy 2006: 214) or to end a conversation (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 1651). In Excerpt (6.4.10) below, *well* is followed by *anyway*, which also signals an ending of the conversation.

(6.4.10)

P: Turn-initial

E: Item (1) closes a topic

F: To mark the closing of a topic

A: Maybe it's time for me to consider it my myself thoroughly and carefully, right?

B: Yes, I think you have done a lot of... of that.

A: **Well**,⁽¹⁾ **anyway thank you for your advice. I will consider it later.**

B: You are welcome.

A: Thank you.

(SECCL: C99-66-02)

It is found that *well* precedes a concluding remark. In Excerpt (6.4.11) below, the speaker was talking about one of her embarrassing events. After describing what happened, she used *well* to preface her conclusion. Such use as prefacing the coda of a story is identified in the monologic genres only.

(6.4.11)

P: Utterance-medial

E: Item (1) closes a topic

F: To mark the closing of a topic

Secretly she told me that: "Hey, you worn your T-shirt inside out." Oh, my god. I looked down on myself and found I really... put my... put my T-shirt inside out. Well it's it's really terrible and I looked around to find... only to find everybody was liking at me. **Well**...⁽¹⁾ **it's really an... embarrassing <embarra>... embarrassing thing that I'll never forget.** It's too terrible so from that time on everyday when... when I have already dressed myself I'll...

(SECCL: B02-01-31)

6.4.2.5 *Well* prefacing questions

Well is found to preface a question. In Excerpt (6.4.12) below, Speaker A uses *well* before raising a question. This seems to make her question sound more indirect. *Well* prefacing a question is one of the least frequent types of co-occurrence.

(6.4.12)

P: Turn-initial

E: *Well* prefaces a question

F: To sound less direct

B: And I think this kind of activities may also help you to improve your academic knowledge and besides that you may also try to know more people and make more friends And I I think that this can broaden your horizon

A: **Well**... **will I have the chance to make many foreign friends to practice my oral English?**... or... do I have much opportunity to have a part time job?

B: Well I think maybe now on this campus maybe you can not. But in the future... you may be you have such kind of chances

(SECCL: C00-65-32)

The instances of *well* prefacing a question can also fit into the category of *the opening of a topic*, but since the interrogative is obvious evidence and this type of co-occurrence is identified in the investigations of other DMs, it forms a category for discussion here in order

to be consistent with the analyses of other words and phrases.

6.4.2.6 *Well* co-occurring with transitions and shifts of topic

The previous four sub-sections discuss *well* co-occurring with reported speech, repairs and rephrasing, the opening and closing of a topic and concluding remarks and questions. *Well* in these contexts to some degree suggests a transition in discourse. In some cases, transitions and topic shifts are obvious. These instances are therefore grouped together.

Well signals a transition. In Excerpt (6.4.13), the speaker talked about her first part-time job, which was a good experience because she made a good deal of money. She used *well* to make a transition from the first job to the second job. In Excerpt (6.4.14), Item (1) *let's go back* explicitly indicates the shift of a topic.

(6.4.13)

P: Utterance-medial

E: Item (1) separates the preceding discourse from the following discourse

F: To mark a transition

Task 2

I thought I begin to do part-time job several years ago in a factory. Because I'm a secretary... learner<learnerer> in a secondary school, I'm the one who has almost the most top<topest> education that let me do counting, typing, copying, answering the message manager's calls<exceptary> I did it for a whole summer vocation, and earned a thousand Yuan. It's a large sum of money to me at that time. It was excited<exciting>. It was exciting to me <al> too. I told everyone I knew <knowed> I had earned such lar such large sum<sums> of money to anybody I knew<know>. **Well,** ⁽¹⁾ **next the job**<jobs> **I had** <have> I had<have> done it was <it's> not so happy<happiness> not um happy one <oned>as before, because I found the owners of the factory<factories>er... so males... um...

(SECCL: B98-21-03)

(6.4.14)

P: Utterance-medial

E: Item (1) separates the preceding discourse from the following discourse

F: To mark a transition

Olympia confronts you in the present... the servant figure would have been another suggestion, um that this is modern Paris and that if you're going to find a nude in modern Paris it's going to be in circumstances, that aren't socially acceptable... |well| ⁽¹⁾ **let's go back.** um the bouquet itself, would have said a lot to people at the time um because it comes wrapped in paper you see she's holding up a bouquet wrapped in paper,

(MICASE: LEL320JU143)

Well prefaces a topic shift or signals a change in the direction of the discourse which might have been expected by the interlocutor (Carter and McCarthy 2006: 212, 219). In Excerpt (6.4.15), Speaker A prefaces his or her response with *well*, which changes the direction of their discussion and the response is against speaker B's expectation.

(6.4.15)

- P:** Turn-initial
E: Item (1) separates the preceding discourse from the following discourse
F: To mark a transition
- B: Yeah, just as I, just, I mentioned above, I, I... can make friend with the... foreign students. I can talking English with them and I can improve my English very quickly and, very very quickly and, and very, very <...> very efficiently, yeah, ah, so you think so?
A: **Well**, and ⁽¹⁾ **there is another problem**. I think the tuition tuition fee in the foreign, in the, in overseas, in the abroad is very expensive. How can you afford it?

(SECCL: C01-08-01)

Well occurring at a transition and the shift of a topic accounts for 6.3% and 4.5% in the NNSs' monologues and dialogues respectively. In contrast, it is highly represented in the NSs' highly monologic discourse mode (44.2%) in MICASE and the unscripted speech (44.9%) in ICE-GB. In these two sub-corpora, a particular type of transition is very frequent. 29.8% (31) out of 104 instances of *well* in the highly monologic discourse mode in MICASE and 10% (16) out of 158 instances in the unscripted speech in ICE-GB are found to occur between the speaker's self-raised question and response, as shown in Excerpt (6.4.16). This use is particularly frequent in the NSs' monologic genres.

(6.4.16)

- P:** Utterance-medial
E: Item (1) is a question and Item (2) is the response from the same speaker
F: To mark a transition
- if you take mouse melanoma cells like the ones we've been talking about today, inject them into the tail vein, of a mouse, ⁽¹⁾ **what's likely to happen?** **|well|** ⁽²⁾ **from the diagram on page twenty-four** you know from the veins you get eventually pumped into the, uh right side of the heart,

(MICASE: LEL175SU106)

6.4.2.7 *Well* co-occurring with disagreement and negative evaluation

It is found that *well* co-occurs with disagreement and negative evaluation. In Excerpt (6.4.17), *well* prefaces disagreement and it seems to serve as a cushion to soften the force of the speech. In Excerpt (6.4.18), *well* is found to preface a negative evaluation and it can function as a mitigator in this context.

(6.4.17)

P: Turn-initial
E: Item (1) is disagreement
F: To mitigate disagreement

B: No, um... I think... um... I reports in modern society... is equal especially the males and females. Uh... and if... and to... according to them to entrance <inrance> examination, females have <hawe> scored higher than most males so the collage should admit student by scores... um... so I... I think... eh... .. I don't think the department should stick to the original <adri> <adrinal> plan.

A: **Well,** ⁽¹⁾ **I don't quite agree with you.** You know... um... for a... as for a <gelo> geology department, ability is very important.

(SECCL: C02-50-19)

(6.4.18)

P: Turn-initial
E: Item (1) is negative evaluation
F: To mitigate negative evaluation

B: Yes, I know that there are lots of... famous professors and there are also lots of famous professors in China. And I think and I think you are going your friend... is planning to study college abroad, right? <A: Yeah.> And I think he is too young

A: **Well,** ⁽¹⁾ the teachers in China are good, ⁽¹⁾ **but still not so good compared to the teachers overseas.**

(SECCL: C01-08-04)

This type of co-occurrence is seldom found in the monologic genres. In the dialogic genres, it is highly represented in the NNSs' dialogues, accounting for 30.4%. This phenomenon can be attributed to the test-taking context where the two speakers are asked to exchange ideas and therefore disagreement is expected.

6.4.2.8 *Well* as a preface to a response

Well as a preface to a response to a question, unsurprisingly, occurs more often in the dialogic genres than in the monologic genres. Due to the generic constraints, there are almost no instances of it in the three monologic genres. This type of co-occurrence is more frequently used in the NNSs' dialogues, with 23.2%. In a slightly lower proportion, it accounts for 19.3% in the NSs' highly interactive discourse mode and 18.3% in the NSs' private direct conversations.

As a preface to a response to a question, *well* is followed by responses which can be insufficient as answers, in disagreement or with reservations and negative evaluation. In Excerpt (6.4.19), *well* as a preface to the response to Speaker A's question is followed by some indications of uncertainty (Items (1), (2) and (3)), which show that Speaker B's answer is insufficient. In this case, *well* as a preface to a response seems intended to mitigate the insufficient answers.

(6.4.19)

P: Turn-initial

E: *Well* as a preface to a response to a question; Items (1), (2) and (3) indicate uncertainty

F: To mitigate insufficient answers

A: What sort of activities physical activities were available?

B: Well ⁽¹⁾ **I suppose** uhm the <,> the standard kind of physiotherapy <,> when you asked for it <,> uhm <,> and well sports ⁽²⁾ **I guess** <,> But <,> I mean ⁽³⁾ **I'm not necessarily interested in doing sports**

(ICE-GB: S1A-003)

In Excerpt (6.4.20), *well* as a preface to the response to the speaker's question is followed by disagreement (Item (1)). *Well* in turn-initial position seems to signal to the listener the disagreement which follows.

(6.4.20)

P: Turn-initial

E: *Well* as a preface to a response to a question; Items (1) is disagreement

F: To mitigate disagreement

B: Eh... .. I'm fine... thank you. And what about you?

A: Well. Me too... well one of my friends<friend> is graduating this year. She would like to go to over... go... overseas for... for... for college education.

What do you think of that?

B: Well, I think it was a good idea. ⁽¹⁾ **But actually I think we should finish the college in China first.**

A: Well, I don't think so... eh... go oversea for college study is better.

(SECCL: C01-99-28)

In Excerpt (6.4.21), *well* as a preface to the response to Speaker A's question co-occurs with the hesitation marker *eh* and pauses, which could suggest that Speaker B is taking time to search for contents or lexis. This use has been previously discussed.

(6.4.21)

P: Turn-initial

E: *Well* as a preface to a response to a question; hesitation marker *eh* and pauses

F: To search for contents or lexis

A: I am interested in the university social... socialists. I want to take part in.

Eh... .. **would you give me some advice please?**

B: Well,... eh... our university have... have a Germen... have Germen society. And also a society about environment preservation and a painting school. There are all quite good professions. You can take part in any of them to your content.

(SECCL: C00-58-33)

The literature (e.g. Svartvik (1980: 169) and Schiffrin (1987: 104-105)), discussed in Section 6.2, reports that prefatory *well* occurs more frequently with responses to *wh*-questions than to *yes/no* and tag questions, probably because *wh*-questions offer more options for interlocutors and, at the same time, require more cognitive work. However, sometimes the response to a *yes/no* question also requires effort and time, as shown in Excerpt (6.4.21) above. In Excerpt (6.4.22) below, the responses to the question from the other participants are

not simply a *yes* or *no*. It is possible that *well* is used as a device for gaining more thinking time or as a mitigator to soften the impact of an insufficient answer.

(6.4.22)

<p>P: Turn-initial E: <i>Well</i> as a preface to a response to a question; Items (1) and (2) are non-answers F: To mitigate indirect/insufficient answers</p>	<p>SU-f: <OVERLAP1> do we have to, say exactly what </OVERLAP1> they are? we can't say they're, something SU-f: well⁽¹⁾ we can't ignore 'em, <OVERLAP1> we're doing an inventory. </OVERLAP1> SU-f: <OVERLAP1> look ⁽²⁾ let's talk about it afterwards. </OVERLAP1> (MICASE: LAB175SU026)</p>
---	---

Well occurring as a preface to a response, whether made or not, is particularly exclusive to the two NS dialogic genres. As shown in Excerpt (6.4.23), the turn-changing is constantly negotiated. S7's use of *well* probably indicates that s/he tries but fails to gain the floor. It is possible that the response following *well* is interrupted and is not provided.

(6.4.23)

<p>P: Turn-initial E: <i>Well</i> as a preface to a response F: To gain the floor</p>	<p>S12: <OVERLAP1> and you can see it in the m- in the model right here. </OVERLAP1> S13: <OVERLAP1> yeah it works better there. </OVERLAP1> S6: yeah. S7: well S12: <OVERLAP1> actually right here as well. </OVERLAP1> S7: <OVERLAP1> actually, no over there is where </OVERLAP1> you want to point to S13: this SS: yeah right if it's contiguous with the wall right? (MICASE: STP125JG050)</p>
--	---

6.4.2.9 *Well* co-occurring with key information

Well is flexible as regards positions in utterances. In extra-clausal position, *well* seems to be used to emphasise key information in the narrative. In intra-clausal position, *well* may be placed before a key point, word or phrase. Schourup (2001: 1038-1039) compares *well* with *after all* and *moreover* and finds that only *well* can be used “to ‘focus down’ on the choice of a single word or phrase”, as in Excerpt (6.4.24) below. This type of co-occurrence is rare in the speech under investigation of both NNSs and NSs.

(6.4.24)

P: +M- + *well* + +M

E: Item (1) is a key word

F: To draw attention to the following key information

A: Yeah, that's true. People work in the joint ventures are only workaholic, right?

B: Yeah.

A: Well, but compared with... with working in the government, is a little bit, |well|, ⁽¹⁾advantages, right?

B: I think so.

(SECCL: C99-66-02)

6.4.2.10 *Well* as a continuer

The category of *well* as a continuer is not set up on the basis of a specific type of linguistic evidence. It is mainly based on my interpretations of contexts. *Well* is interpreted as an indication of continuation, suggesting something is about to be said (Svartvik 1980: 175, Schourup 2001, *Collins COBUILD Advanced Learner's English Dictionary* 2006: 1650). In Excerpt (6.4.25) below, Speaker B's use of *well* can be seen as a device for holding the floor and signalling continuation.

(6.4.25)

P: Turn-initial

E: Continuation of the earlier topic

F: To hold the floor and signal continuation

Task 3

A: Hello, Richard. I... I have been offered two jobs, one is working in a government, and the other is in a joint-venture. Now I feel difficult to decide which one I should take.

B: **Well**. First let's look at the government. You know if you work in the government, you have a better chance to serve the people. I think that is good for your future.

(SECCL: C99-21-16)

In addition to the prospective quality, *well* as a continuer suggests retrospective consideration (Svartvik 1980: 177, Schourup 2001: 1043). Prefatory *well* indicates that the old information, i.e. what has already been said, is taken into consideration. In Excerpt (6.4.26) below, speaker B's utterance is closely related to the previous turn.

(6.4.26)

P: Turn-initial

E: Continuation of the earlier topic

F: To hold the floor and signal continuation

A: And you know, I'm not good at writing.

B: **Well**, this is easy. You can... ah... you can look for someone who can<come>conversation very good. You ask him or her to write... write for<to> you.

(SECCL: C97-11-24)

This retrospective linking function of *well* as a DM can also be used to explain *well* as a preface to a response, as discussed in the preceding section.

This type of co-occurrence is frequent, with larger proportions in the dialogic genres, but

similarly represented across the NNSs and NSs' speech with 16.7% in the NNSs' monologues and 7.1% in the dialogues. It accounts for 18.3% in the NSs' highly monologic discourse mode, 8.9% in the highly interactive discourse mode in MICASE, 24.7% in the unscripted monologues and 17.5% in the private direct conversations in ICE-GB.

6.4.2.11 Unclassified instances of *well*

In the six sub-corpora under investigation, four instances of *well* in the NNSs' speech and twelve in the NSs' speech are found impossible to classify, because there is generally not enough linguistic information to interpret the use of *well* with any certainty. For example, in Excerpt (6.4.27) below, as the markup in ICE-GB indicates, what follows *well* is unclear, making it unclassifiable.

(6.4.27)

B: That's what she's going for <,> Societies Secretary <,>

A: **Well** | I <unclear-words> <,>

B: Who's Ed Dickinson

(ICE-GB: S1A-070)

6.4.2.12 Summary of the contexts where Type B *well* tends to occur

The types of co-occurrence which tend to co-occur with Type B *well* are summarised in this section. Tables 6.15 to 6.20 below illustrate the distribution of the positions in an utterance/turn of Type B *well*.

It is found that *well* co-occurs with *hesitation markers, pauses, repetitive words and restarts*. *Well* is also used *as a preface to a response*. In these two contexts, *well* seems to signal that the speaker is searching for content or lexis and *well* seems to be used to gain and hold the floor.

Well co-occurring with *reported speech, repairs and rephrasing, the opening and closing of a topic and concluding remarks and questions* suggests, to some extent, suggests a transition in discourse. In some cases, the transitions and topic shifts are obvious. These instances are therefore grouped together in the category of *transitions and shifts of topic*.

The instances of *well* in the categories of *disagreement and negative evaluation* and *as a preface to a response* can be interpreted as a mitigator. *Well* co-occurring with key information and *well* as a continuer seem to be used to draw listeners' attention to what follows or to hold the floor.

Due to the variations in the types of activity in the six sub-corpora, it is anticipated that

the distribution of the identified co-occurrence of *well* varies to some extent across sub-corpora. In the NSs' monologues, the most frequent type of co-occurrence is in *reported speech* and in the dialogues, the most frequent one is in *disagreement and negative evaluation*. These two types of co-occurrence are not as frequently represented in the NSs' speech, in which the category of *transitions and shifts of topic* is highly represented.

Table 6.15: Distribution of co-occurrence of *well* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	18.8			6	12.5					1	2.1	1	2.1	1	2.1
2. Reported speech	33.3			3	6.3			7	14.6					6	12.5
3. Repairs; rephrasing	0														
4. Opening/closing of a topic; concluding remarks	8.3	4	8.3												
5. Questions	0														
6. Transitions; shifts of topic	6.3			3	6.3										
7. Disagreement; negative evaluation	4.2			2	4.2										
8. As a preface to a response	0														
9. Key information	10.4			5	10.4										
10. As a continuer	16.7			8	16.7										
Unclassified	2.1			1	2.1										
Occurrences: 48 out of 300 (random samples)	100.0	4	8.3	28	58.3			7	14.6	1	2.1	1	2.1	7	14.6

Table 6.16: Distribution of co-occurrence of *well* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	19.6	12	10.7	3	2.7			1	0.9			4	3.6	2	1.8
2. Reported speech	0														
3. Repairs; rephrasing	0.9			1	0.9										
4. Opening/closing of a topic; concluding remarks	7.1	7	6.3	1	0.9										
5. Questions	3.6	4	3.6												
6. Transitions; shifts of topic	4.5	4	3.6	1	0.9										
7. Disagreement; negative evaluation	30.4	28	25.0	5	4.5			1	0.9						
8. As a preface to a response	23.2	26	23.2												
9. Key information	0.9			1	0.9										
10. As a continuer	7.1	7	6.3	1	0.9										
Unclassified	2.7	2	1.8	1	0.9										
Occurrences: 112 out of 300 (random samples)	100.0	90	80.4	14	12.5			2	1.8			4	3.6	2	1.8

Table 6.17: Distribution of co-occurrence of *well* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	6.7	1	1.0	1	1.0						3	2.9	2	1.9	
2. Reported speech	14.4							12	11.5				3	2.9	
3. Repairs; rephrasing	7.7			3	2.9			1	1.0		4	3.8			
4. Opening/closing of a topic; concluding remarks	1.9	2	1.9												
5. Questions	4.8	2	1.9	1	1.0			2	1.9						
6. Transitions; shifts of topic	44.2	2	1.9	44	42.3										
7. Disagreement; negative evaluation	1.0			1	1.0										
8. As a preface to a response	1.0	1	1.0												
9. Key information	0														
10. As a continuer	18.3	2	1.9	17	16.3										
Unclassified	0.0														
Occurrences: 104	100.0	10	9.6	67	64.4			15	14.4			7	6.7	5	4.8

Table 6.18: Distribution of co-occurrence of *well* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	10.7	6	2.2	2	0.7	1	0.4	6	2.2	1	0.4	12	4.4	1	0.4
2. Reported speech	4.1							10	3.7					1	0.4
3. Repairs; rephrasing	3.3			3	1.1			3	1.1			3	1.1		
4. Opening/closing of a topic; concluding remarks	3.0	5	1.9	3	1.1										
5. Questions	7.4	19	7.0	1	0.4										
6. Transitions; shifts of topic	24.4	36	13.3	27	10.0	3	1.1								
7. Disagreement; negative evaluation	15.9	40	14.8	2	0.7			1	0.4						
8. As a preface to a response	19.3	51	18.9	1	0.4										
9. Key information	1.9	2	0.7	2	0.7			1	0.4						
10. As a continuer	8.9	14	5.2	10	3.7										
Unclassified	1.1	3	1.1												
Occurrences: 270 out of 300 (random samples)	100.0	176	65.2	51	18.9	4	1.5	21	7.8	1	0.4	15	5.6	2	0.7

Table 6.19: Distribution of co-occurrence of *well* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	10.1			4	2.5			2	1.3			6	3.8	4	2.5
2. Reported speech	3.8							4	2.5					2	1.3
3. Repairs; rephrasing	4.4			2	1.3			4	2.5					1	0.6
4. Opening/closing of a topic; concluding remarks	7.6	12	7.6												
5. Questions	1.9			3	1.9										
6. Transitions; shifts of topic	44.9	1	0.6	70	44.3										
7. Disagreement; negative evaluation	2.5			4	2.5										
8. As a preface to a response	0														
9. Key information	0														
10. As a continuer	24.7	1	0.6	38	24.1										
Unclassified	0														
Occurrences: 158	100.0	14	8.9	121	76.6			10	6.3			6	3.8	7	4.4

Table 6.20: Distribution of co-occurrence of *well* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; repetitive words; restarts	14.8	11	4.3	12	4.7			3	1.2			7	2.7	5	1.9
2. Reported speech	4.3							10	3.9					1	0.4
3. Repairs; rephrasing	3.1			4	1.6			1	0.4			2	0.8	1	0.4
4. Opening/closing of a topic; concluding remarks	0.8	2	0.8												
5. Questions	5.1	9	3.5	4	1.6										
6. Transitions; shifts of topic	14.8	21	8.2	17	6.6										
7. Disagreement; negative evaluation	17.5	39	15.2	6	2.3										
8. As a preface to a response	18.3	45	17.5	2	0.8										
9. Key information	0.4							1	0.4						
10. As a continuer	17.5	32	12.5	12	4.7			1	0.4						
Unclassified	3.5	9	3.5												
Occurrences: 257 out of 300 (random samples)	100.0	168	65.4	57	22.2			16	6.2			9	3.5	7	2.7

6.5 Chapter summary and conclusions

The frequency information and collocates of *well* are used as a point of entry into the data. The frequencies of *well* in the six sub-corpora reveal that there are more instances of *well* in the dialogic genres than in the monologic ones. This finding is expected; *well* is often used in more interactive genres.

As I hypothesised, the manual classification of Type A *well* and Type B *well* reveals that *well* is not primarily used as a DM by the NNSs, but is frequent among the NSs in their highly interactive discourse mode and private direct conversations. In addition, the prominent collocates of *well* in the patterns across the six sub-corpora indicate some uses of Type A *well* and Type B *well*, as well as identifying some DM collocations, such as *oh well*, *well I think*, *well you know* and *well but* in the NNSs' speech and *well well*, *ok well*, *oh well*, *yeah well* and *so well* in the NSs' speech.

The uses of *well* as a DM are discussed on the basis of its collocation phenomena in regard to placing it in an utterance/turn. In general, Type B *well* occurring in extra-clausal positions is common. Most of the instances appear in extra-clausal utterance-medial position in the three monologic genres and in extra-clausal turn-initial position in the three dialogic genres. There is no marked difference in the distribution of the positions in an utterance/turn of Type B *well* across the two types of genre and between the two groups of speakers. However, there are marked differences in the distribution of the types of co-occurrence of *well* across corpora. This can be attributed to the variations in the types of activity in the six sub-corpora. For example, the NNSs' monologues are mainly narratives, in which there are more opportunities for using *well* to mark reported speech. It is concluded that the use of *well* is sensitive to genre and dependent on context.

The difficulty of classification based on collocation phenomena has to be acknowledged, in particular when *well* occurs in the extra-clausal turn-initial position. More than one type of co-occurrence, such the structural evidence (e.g. as a topic opener and as a preface to a response), lexical co-occurrence (e.g. emphatic lexis) and surrounding proposition (e.g. a disagreement), can be found in a single instance of *well*, for instance, Excerpt (6.4.21), repeated here as Excerpt (6.5.1) for ease of reading. One example of *well* does not necessarily fit without controversy into one category or another. I coded the instance of *well* in Excerpt (6.5.1) in the category of *as a preface to a response*, according to my subjective judgment. Arguably, however, it also fits into the category of *hesitation markers, pauses, repetitive*

words and restarts.

(6.5.1)

P: Turn-initial

E: *Well* as a preface to a response to a question; hesitation marker *eh* and pauses

F: To search for contents or lexis

A: I am interested in the university social... socialists. I want to take part in.

Eh... .. **would you give me some advice please?**

B: **Well**,... **eh**... our university have... have a German... have German society. And also a society about environment preservation and a painting school. There are all quite good professions. You can take part in any of them to your content.

(SECCL: C00-58-33)

It is evident that DMs perform some functions in discourse. The classification can be made; however, it is difficult because DMs are multifunctional. This causes difficulties in investigating DMs and in reaching agreement on the use being made of them.

This study uses three publicly available corpora to investigate the use of DMs. The selection of the SECCL, MICASE and ICE-GB is justified in Section 3.1 of Chapter 3. The use of corpora with different mark-ups nevertheless has its limitations. The NNS corpus, SECCL, clearly identifies speaker change, but the NS corpora, MICASE and ICE-GB, annotate similar information in the mark-up language, which cannot easily be searched and identified. This causes difficulties for an investigation conducted across sub-corpora, giving rise to some questions. For example, Section 6.4.2.1 above discusses the co-occurring hesitation markers, pauses, repetitive words and restarts and Type B *well* in turn-initial position is about twice as frequent in the NNSs' speech as in the NSs' speech. Initial *well* marking hesitation and diffidence can be a notably NNS usage. However, it is difficult to test this hypothesis. The number of turns in SECCL can be easily counted by searching the speaker identities, *a* and *b*, and then finding the proportion of turns beginning with *well*, but this cannot be done in MICASE and ICE-GB. If I hypothesise that the Chinese NNSs tend to use *well* to mark hesitation and diffidence and NSs are more likely to use *ok*, which sounds more confident than *well* does, it is difficult pro tem to find the answer in the case of these corpora, which must be left as a topic for future research.

CHAPTER 7: ANALYSES OF *YOU KNOW, I MEAN* AND *YOU SEE*

7.1 Introduction

This chapter begins an examination of two-word phrases, which are quite distinct from the three words, *like*, *oh* and *well*, investigated in the preceding chapters. The ambiguous grammatical roles of *you know*, *I mean* and *you see* make it difficult to draw a distinction between non-discourse use (Type A) and discourse use (Type B). This is further discussed below.

There are three reasons why these three phrases are discussed together in this chapter. First, *you know* and *I mean* have usually been investigated together in previous studies. For greater ease in reviewing, they are also discussed together in the present study. Second, the grammar aspects of these three phrases are similar. Third, some of the collocation phenomena surrounding them are shared. To save space in this thesis and avoid overlapping discussions, they are discussed together in this chapter.

This study focuses on the use of Type B *you know*, *I mean* and *you see* by the Chinese NNSs and the NSs. It aims to give a detailed description of these three phrases from the aspects of grammar and discourse. Attention is paid to the discourse use of *you know*, *I mean* and *you see* (rather than to the non-discourse use), for three reasons: first, they are found to be primarily used as DMs in NSs' conversations (Biber *et al.* 1999: 1096). Indeed, *you know* and *I mean* are two of the most frequently-used DMs in the NS speech (Carter and McCarthy 2006: 214). Previous research on their discourse uses has shown fruitful results, for example in Östman (1981), Schourup (1985), Holmes (1986), Schiffrin (1987), Erman (1987), Biber *et al.* (1999), Fox Tree and Schrock (2002), Macaulay (2002), Müller (2005) and Carter and McCarthy (2006). However, little research in the use of DMs by NNSs has been carried out, though Müller's (2005) and Fung and Carter's studies (2007) are two examples of the few exceptions (see the detailed discussion in Chapter 2).

Second, each of these phrases has two very distinct uses. One of these is the non-discourse use. The phrases *you know*, *I mean* and *you see* are consequences of the two open-choice decisions, made from the paradigm of pronouns (e.g. *you*, *I*, *he*, *she*, *we*, etc.) + *know(s)*, *mean(s)* or *see(s)*. Each of these is independently selected and grammatically constructed. This is quite different from *you know*, *I mean* and *you see* for discourse use (Type B), which are chosen by the "idiom principle" (Sinclair 1991: 110). Type B *you know*, *I mean*

and *you see* have no variation of the phrase, no independent choice or paradigmatic selection. There is no negative form or past tense for Type B *you know*, *I mean* and *you see*. They are used as fixed phrases.

Third, Type B *you know*, *I mean* and *you see* seem to carry some negative connotations, such as lacking of confidence, unclear thinking, incompetent social skills, undesirable speaking idiosyncrasies and the like. The use of these DMs is associated with the style of unskilful speakers (O'Donnell and Todd 1991: 69). However, they are still in constant use by NSs. As Crystal (1988: 48) argues, *you know* is used “as the oil which helps us perform the complex task of spontaneous speech production and interaction smoothly and efficiently”.

The following sections set out my hypotheses about the use of *you know*, *I mean* and *you see* and my research questions, followed by a discussion of the grammar aspects of the phrases, with emphasis on the grammatical ambiguity, which causes difficulty in the distinction between Type A *you know*, *I mean* and *you see* and Type B. As mentioned in the preceding chapters, a bottom-up approach is employed. The analysis first presents the frequency information and patterns of *you know*, *I mean* and *you see*, to give an overall picture of the use of these three phrases in the six sub-corpora under investigation. The major part of the analysis is the discourse aspects of Type B *you know*, *I mean* and *you see*, looking at the positions in an utterance/turn and the collocation phenomena surrounding them. The identification of co-occurrence leads to the interpretations of the functions of Type B *you know*, *I mean* and *you see*.

7.1.1 Hypotheses and research questions

In my analyses in the preceding chapters, some DMs (e.g. *like* and *well*) are under-represented in the Chinese NNSs' speech and some DMs are over-represented (e.g. *oh*) as compared with the NS data. In this chapter, I hypothesise that *you know*, *I mean* and *you see* are probably under-represented in the NNSs' speech, due to generic constraints. I also hypothesise that there are more instances of Type B phrases in the dialogic genres, as these phrases include the first and second pronouns, which may occur more often in the presence of other speakers.

To test my hypotheses and to produce a thorough description of the use of *you know*, *I mean* and *you see*, the following questions are addressed:

1. What are the distributions of the three phrases *you know*, *I mean* and *you see* in the

speech of the NNSs and NSs?

2. What do the collocates of *you know*, *I mean* and *you see* reveal about their uses?
3. How do the NNSs and NSs use Type A *you know*, *I mean* and *you see* and Type B?
4. What other DMs do Type B *you know*, *I mean* and *you see* co-occur with?
5. Where do Type B *you know*, *I mean* and *you see* appear in an utterance/turn?
6. With what types of co-occurrence or in what contexts do Type B *you know*, *I mean* and *you see* tend to occur?

Question 1 above asks the proportion of Type A phrases and Type B phrases in the NNSs' and NSs' speech and ascertains whether Type B phrases are under-represented in the NSs' speech, as I have hypothesised. Questions 2 to 4 concern the uses of these three phrases and the answers to them support the manual classification of *you know*, *I mean* and *you see* between Type A and Type B (answering Question 1), while validating some claims about the use of Type B phrases, which are based on types of co-occurrence and contextual information (answering Questions 5 and 6). The answers to the above questions help to reveal the use made by the NNSs and NSs of DMs. The importance of the research questions is discussed in more detail in Section 1.1.2 of Chapter 1.

7.1.2 Ways of distinguishing *you know*, *I mean* and *you see* between non-discourse use (Type A) and discourse use (Type B)

It has been explained in the process of distinguishing between Types A and B set out in Section 3.3.2 that one difficulty in investigating the phrases *you know*, *I mean* and *you see* in utterance/turn-initial is their syntactical ambiguity. *You know*, *I mean* and *you see* in clause-initial position can be the subject and verb of a main clause followed by an embedded clause as its object, i.e. the use of Type A. For instance, ***You know*** *there are many foreigners also foreign teachers in our... on our campus.* (SECCL: C00-65-34) This example is ambiguous in that it either means *You know (that) there are many foreigners also foreign teachers in our campus* or *You know, there are many foreigners also foreign teachers in our campus*. In the latter, *you know* is used as a DM, a separate unit outside the clause structure and belonging to Type B in this study.

Biber *et al.*'s three criteria (1999: 1076-1078) (see Section 3.3.2) for determining "utterance launchers" as DMs is used and exemplified below. To distinguish the highlighted

you know in Excerpt (7.1.1) below, *that* is added after *you know*. This is grammatically correct and what follows, *you might wanna do that* is part of the *that*-clause. However, in this the context, Speaker 1 is offering a possibility, since *you know* can be syntactically and semantically optional and can be placed in other positions in the turn.

(7.1.1)

S2: <OVERLAP1> yeah i saw </OVERLAP1> a course like Introduction to World Politics or something

S1: right that, <OVERLAP1> right </OVERLAP1>

S1: <OVERLAP1> which is </OVERLAP1> Poli Sci one-sixty

S2: okay

S1: **you know**, you might wanna do that, and then you might wanna take, this biological anthro course or a beginning geology course or,

S2: <OVERLAP2> mhm </OVERLAP2>

S1: you know something like that, and then if you decide you wanna do geology or, bio anthro, then you could use that, poli sci for social science distribution.

(MICASE: ADV700JU023)

In addition to Biber *et al.*'s three criteria (1999: 1076-1078), *LUG* (Sinclair and Mauranen 2006) is used to facilitate the classification of the instances of *you know*, *I mean* and *you see* between Types A and B. Although in the *LUG* analysis subjective judgement is unavoidable, investigating these phrases in a new model is of some help. More detailed discussion is provided in Section 7.2.2 below.

As with *like* and *well*, I compare my tagging with that in the ICE-GB corpus. I was surprised that my manual classification of Types A and B and the tagging in ICE-GB show similar results, shown in Appendix 5. The frequencies of Type B *you know*, *I mean* and *you see* in the manual classification mostly resemble the numbers retrieved by the accompanying software *ICECUP*, except for the case of *you see* in the sub-corpus of the unscripted monologues in ICE-GB. Whereas the two processes, my manual classification and the tagging in ICE-GB, do not give identical results, they are similar enough to add confidence and reliability to my manual classification of Types A and B, as well as the tagging in ICE-GB.

7.2 Previous studies of *you know*, *I mean* and *you see*

7.2.1 Grammatical aspect: Syntactical structure

The syntactical structure of Type A *you know*, *I mean* and *you see* is mentioned in Section 7.1.2 above. Type B *you know*, *I mean* and *you see*, in general, have flexible positions in a clause, but are governed by some syntactical rules. Crystal and Davy (1975: 92) claimed that

you know can be placed utterance-initially, -medially and -finally, but in utterance-medially, it occurs “always at a point of major grammatical junction”. Crystal (1988: 48) pointed out that, first, *you know* in sentence-initial position is usually followed by a statement and second, *you know* as a unit of DM cannot be inserted into a compound expression, for instance, such as “**I went to New you know York*”. Third, it is not usually placed before a coordinated pronoun, such as “**John and you know I left early*”. These syntactical regulations applied to DM *you know* also apply to *I mean* and *you see*.

The above syntactic tendencies, based on Crystal’s observation or intuition, are likely to be true in the NS speech because in the above two examples, *New York* and *John and I* are generally processed as a single unit by NSs. Nevertheless, in the NSs’ speech under investigation, a few instances of *you know*, e.g. the two instances of *you know* in Excerpt (7.2.1) below, do not conform to this syntactic tendency.

(7.2.1)

and of course, this doesn't_ you know, when the light goes through glass and slows down, uh that's another matter because it's moving relative to the glass and so forth but, uh when we're looking at light or any electromagnetic |**you know**| uh, wave in a vacuum uh

.....

so, let's take a concrete example. you'll see what i mean. suppose i'm in oh |**you know**| a, a something coasting along like this at constant velocity, and i throw a ball. okay?.....

(MICASE: LEL485JU097)

NNSs, however, might process language in a different way. *New York* might not be a single unit for NNSs. In NNSs’ speech, *you know* as an insert in a compound expression might be found, in part, because some collocations are not strong enough for NNSs to use as a unit and in part because NNSs are not familiar with them. NNSs might produce *I’m going to the city you know centre*. In this case, the speaker might think of *city* first and formulate *centre* later; therefore, *city centre* is not necessarily a single unit. The way of processing language can be taken into consideration in the syntactical roles of Type B *you know*, *I mean* and *you see*.

Based on such example as (7.2.1) and possible NNS use, it is argued that DMs do not follow traditional syntactical rules and it would be problematic to describe the positions of DMs with grammatical labels. The following section presents how *LUG* (Sinclair and Mauranen 2006) is used to describe where DMs occur.

7.2.2 Linear Unit Grammar analysis of *you know*, *I mean* and *you see*

It has been discussed that the grammatical status of *you know*, *I mean* and *you see* is ambivalent because they can express their literal meanings (Type A) and also can take on the role of DMs (Type B). In the previous section, the three criteria of Biber *et al.* (1999: 1076-1078) can be applied to determine “utterance launchers” as DMs, which means that *you know*, *I mean* and *you see* function alone, not as clause components.

Because of the ambiguous role of these three phrases, it is often necessary to look at the occurrences in a larger discourse. Sinclair and Mauranen’s *LUG* analysis (2006) is adopted to distinguish whether *you know*, *I mean* and *you see* are message- or interaction-based. As mentioned earlier (see Chapter 2), *LUG* is designed as a descriptive bottom-up approach to grammar and is intended to be compatible with most conventional grammars. In the *LUG* analysis, the three phrases are classified into two functional elements, an M- element and an OI element. The latter functions as a marker in discourse and is referred to as Type B in the present study. The former contributes its semantic meaning and is referred to as Type A.

In real-time communication, the speaker and hearer process meaning incrementally (Brazil 1995). *You know*, *I mean* and *you see*, as M elements, increment the shared knowledge of the interlocutors. As OI elements, they mainly contribute to the aspects of the interaction, such as initiating, maintaining and structuring the interaction and controlling the timings.

The following excerpts are taken from the NNSs’ speech and analysed with *LUG*. When the phrase *you know* is followed by a *that*-clause, it is very likely that the instance of *you know* is an M element, conveying the listener(s)’ awareness of something, as in the two instances of *you know* in Excerpt (7.2.2). (See Appendix 4 for a list of the labels in the *LUG* analysis.)

(7.2.2)

B: Why? I don't agree with you. They shouldn't stick to the original plan. **You know** that that
M M M- +M M- MF OT
females get higher marks than males. So they should enroll the higher marks person.
+M OT M- +M

A: But but **you know** that score is not everything. Higher score does doesn't mean higher ability.
OI OI M- OT +M M- MR
You should know that.
M

(SECCL: C02-61-10)

The instance of *I mean* in Excerpt (7.2.3) is assigned as an OI element, for what follows is an imperative verb *try* and it is no longer grammatically correct once *that* is inserted. This

(7.2.4)

A: **You know** I never did get to spin. But I / was like

B: That was cool. (AmE) (Biber et al. 1999: 1075)

You know can provide orientation in discourse, followed by either a new topic or a topic shift.

Second, *you know* is used for turn management. *You know* in utterance/turn-initial position serves a turn-taking function. Sometimes, it serves more than one function at a time, such as opening the first topic and taking a turn (Holmes 1986: 5-6, Erman 1987: 52). *You know* in utterance/turn-final position has a turn-yielding function, but Erman (1987: 53) claims that this is not one of its main functions because in her data *you know* in utterance/turn-final position occurs less often. Additionally, it is suggested that this function refers to intonation information. *You know* spoken with rising intonation tends to ask for feedback from the hearer and *you know* with falling intonation seems to relinquish the floor to the hearer. In both of these cases, *you know* serves a turn-yielding function (Östman (1981) cited in Holmes (1986: 6) and Müller (2005: 148)).

Third, *you know* as a signal for reported speech has been extensively discussed in the studies of the NS speech (Erman 1987: 115, Schiffrin 1987: 282, Redeker 1991: 1163, He and Lindsey 1998: 143, Erman 2001: 1342, Müller 2005: 167-171). NSs start with a DM to report speech, which may be a way of saying that they are now starting to report speech. However, whether the speech is direct or indirect is unknown. Redeker (1990) argues that the DM prefacing reported speech may be interpreted as either part of the direct quotation, as in Example (7.2.5) below, or being added by the current speaker to introduce quotations, as in Example (7.2.6).

(7.2.5)

A-and he says well I don't want to make a profit on it. (1990: 374)

(7.2.6)

He says, you know, pack and go! That's it! You're out o'here! (1990: 374)

Fourth, *you know* is found to co-occur with certain adjectives and adverbs, such as *important*, *main* and *right now* and be assigned the function of prefacing information of importance and salience (He and Lindsey 1998: 140-141). In the case of this kind, *you know* also has the interactional function of drawing the hearer's attention (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 797).

Fifth, the speaker uses *you know* to signal a search for lexical words or content information and it frequently co-occurs with pauses or other DMs (Erman 1987: 121, Müller 2005: 158-160). *You know* also marks linguistic imprecision when the speaker is not certain which lexical words to choose (Holmes 1986: 10).

Sixth, *you know* is used in clause-medial position primarily to clarify what has just been said (Erman 1987: 114, Crystal 1988: 47). For instance, *He's just got a new BMX – you know, one of those tough little bikes* (Crystal 1988: 47).

Seventh, *you know* prefacing a repair has been discussed in detail by Erman (1987: 141-181), who classifies repairs into four types: repetition, restart, insertion and correction. Holmes (1986: 11-12) identifies *you know* as signalling a false start, followed by a change of syntactic structure. In Müller's study (2005: 162), she found that almost one-third of her American NSs use this function, whereas less than 4% of her German participants employ it.

The last two functions are less frequently used. *You know* is used to introduce exemplifications (Erman 1987: 114). It can also mark approximations, as in *She said you're, you're nice, you're pretty, you know whatever* (Erman 2001: 1348). In Müller's study (2005), *you know* marking approximations is one of the less frequent functions and is used more by NSs than German NNSs.

Compared with *you know*, *I mean* serves fewer textual functions in discourse. It primarily serves to 1) act as an utterance opener, 2) manage turns, 3) mark a self-repair and 4) elaborate, clarify, modify or expand what has been said.

First, the same as *you know*, *I mean* is labelled as an "utterance launcher" (Biber *et al.* 1999: 1075) to open a turn, as shown in Example (7.2.7).

(7.2.7)

A: **I mean** are these the same, these are the same?

B: Uh huh. Those are kind of further back. (AmE) (Biber *et al.* 1999: 1075)

Second, *I mean* in utterance/turn-initial position is used for the speaker to take a turn. Unlike *you know* in utterance/turn-final position which has a turn-yielding function, Erman (1987: 52-53) found that *I mean* seldom occurs in final position, but it does, when the speaker is forced to give up the turn.

Third, as Type A *I mean* implies, Type B *I mean* is used to mark a self-repair. The repair can be a self-correction, as in *I know he's Portuguese, I mean Brazilian, but he's probably read quite widely in Latin American literature* (Carter and McCarthy 2006: 107) or a word

substitution such as what Schiffrin (1987: 301) terms *replacement repair*, as exemplified in Example (7.2.8).

(7.2.8)

Sally: *Were your parents pretty strict or ...*

Irene: *Not at all. And not t'my disadvantage. **I mean** not t'my advantage as I- I see it now because I got everything I wanted then. (1987: 301)*

The last main function of *I mean* is to elaborate, clarify or modify what has just been said (Erman 1987: 118-119, Carter and McCarthy 2006: 107). Schiffrin (1987: 296) suggests that “the literal meaning of the expression ‘I mean’ influences its function”; *I mean* is used to signal the speaker’s modification, expansion and explanation of the speaker’s prior talk.

The above short accounts of the textual functions of *you know* and *I mean* are supported by the collocation phenomena adopted in the present study. The categories for discussion in Section 7.4.2 are types of co-occurrence of *you know* and *I mean*, which are the basis for suggesting their functions.

In terms of interactional functions, the five main functions accorded to *you know* are 1) marking shared or general knowledge, 2) appealing for acceptance, 3) appealing for patience, understanding and sympathy, 4) softening the force of utterance and 5) acting as a question.

First, as Type A *you know* implies, the proposition followed by Type B *you know* may be either knowledge shared between the speaker and the hearer (Holmes 1986: 8, Schiffrin 1987: 309-310, Biber *et al.* 1999: 197, Müller 2005: 177-181, Carter and McCarthy 2006: 221) or general knowledge and common experience (Schiffrin 1987: 274-275). It seems that *you know* is used to get the hearer involved and to directly draw the hearer’s attention.

Second, *you know* is found to mark new information, but the speaker requests the hearer to resort to his or her knowledge or experience and further to raise the possibility of the hearer’s acceptance of the new information (Biber *et al.* 1999: 1077).

Third, *you know* serves the function of appealing. As noted above, at the textual level, *you know* is found to signal a search for lexical words or content information and it frequently co-occurs with pauses or other DMs (Erman 1987: 121, Müller 2005: 158-160). It is difficult to specify the interactional functions associated with this use, but it seems reasonable to assume that the speaker aims to stall for time and *you know* is used to appeal for the hearer’s patience (Erman 1987: 137). It can also be interpreted as an appeal for understanding when the speaker is unable to find an appropriate expression and to provide sufficient argument

(Müller 2005: 181-182). Another possibility is that when *you know* co-occurs with an embarrassing experience or personal information, the speaker seems to use *you know* to appeal for sympathy (Holmes 1986: 10).

Fourth, *you know* is found to co-occur with a critical or negative comment and the use of *you know* seems to make the statement less direct and to soften the force of the utterance (Holmes 1986: 10, Carter and McCarthy 2006: 108). Crystal and Davy (1975: 91-92) term *you know*, *I mean*, *sort of* and *you see* as “softening connectives” or “softeners”, which primarily serve to change the speaking style to informal.

Last, *you know* referring to prior statements can be used to check the hearer’s understanding (Carter and McCarthy 2006: 108), while referring to the subsequent statements can be taken as a reduced form of the question beginning *do you know* to prepare the hearer for the coming information (Schiffrin 1987: 287). *You know* in these two cases acts as a question, but in the present study the second of these is treated as a use of Type A *you know*.

Three main interactional functions served by *I mean* are: 1) orientating the speaker’s talk, 2) acting as a hesitation marker and 3) acting as a mitigator. Like *you know*, Type A *I mean*, it is suggested, affects its functions in discourse (Schiffrin 1987: 309-310). *I mean* orientates the speakers’ own talk and gains the hearer’s attention. *I mean* is found to act as a hesitation marker, which usually co-occurs with pauses (Erman 1987: 119, Carter and McCarthy 2006: 108). *I mean*, like *you know*, can be used as a delaying device. In addition, as a mitigator, *I mean* is followed by an adjustment for what has been said when the speaker lacks confidence and would like to reduce his/her commitment (Erman 1987: 119, Fox Tree and Schrock 2002: 733).

The interactional functions of *you know* and *I mean* require more interpretation based on intuition and contexts. This causes problems for the present study in classifying the uses of DMs, as it aims to describe the uses of DMs on the basis of linguistic evidence. It is sometimes impossible to categorise the uses of DMs without resorting to intuitive interpretation, thereby describing the uses in relation to a continuum of strong linguistic co-occurrence and the intuitive interpretation of contexts.

7.2.4 Previous studies of *you see* as a discourse marker

There are far fewer studies of *you see* than the great number of them on the phrases *you know* and *I mean* as DMs, discussed in the preceding section. This is probably because *you see* is

found to be less common than *you know* and *I mean* (Biber *et al.* 1999: 1096-1097).

It is reported in the work of Biber *et al.* (1999: 1097) that *you see* is predominantly used as a DM in NSs' conversations and that it is about eight times as frequently used in British English as in American English. In addition, Prodromou's study (2008) on second language (L2) users' conversations (42 proficient L2 users of English with different L1 backgrounds) and informal spoken English in the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and the British National Corpus (BNC) shows that *you see* occurs twice as frequently in L1 speakers' conversations as in L2s'. These findings are rather different from those in the present study. *You see* occurs slightly more often in the NNSs' dialogues than in the British NSs' conversations (see Section 7.3.2 below for more detailed discussion).

The general use of *you see*, like that of *you know* and *I mean*, in utterance/turn-initial position is labelled an "utterance launcher" and *you see* in medial or final positions serves as a signal of a coming explanation (Biber *et al.* 1999: 1075). Erman (1987: 117-118) argues that *you see* tends to occur in explanatory and argumentative discourse, allowing speakers to introduce a personal viewpoint. She interprets *you see* as a device of persuasion.

7.3 Frequency information about *you know*, *I mean* and *you see* in the speech of the non-native speakers and native speakers

The overall frequencies of the phrases *you know*, *I mean* and *you see* in the NNSs' and NSs' speech are shown in Tables 7.1, 7.2 and 7.4 respectively. These three tables provide the word counts of the six sub-corpora, raw and normalised frequencies of the phrases under investigation and those of the Type B phrases and their percentages of use as DMs.

In the cases where raw frequencies were below 400, manual classification of Types A and B was manageable. When those numbers exceeded 400, they were investigated by random sampling (see Section 3.3.7 for the sampling procedure). Three sets of 100-line concordance samples were classified manually and the instances of Type B were used for the investigation of the positions in utterances/turns and collocation phenomena.

7.3.1 Overall frequencies of *you know* and *I mean*

The instances of *you know* and *I mean* are manually grouped into Types A and B. This classification reveals that *you know* and *I mean* are primarily used as DMs. Across the six sub-corpora, the percentages of Type B range from 71% to 83.5% in the case of *you know* (see

Table 7.1) and from 66.7% to 93.7% in the case of *I mean* (see Table 7.2). Previous studies indicate that *you know* and *I mean* are predominantly used as DMs in the NS conversations (Biber *et al.* 1999: 1096). Based on the high percentages of Type B *you know* and *I mean*, it can be concluded that *you know* and *I mean* are primarily for discourse use in the speech of the NNSs and NSs under investigation.

Table 7.1: Frequency information of *you know* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percentage (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	303	9.0	228	75.2	6.8
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	3,263	54.7	246 out of 300 ^a	82.0	44.8
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	163	12.2	119	73.0	8.9
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	2,671	46.2	213 out of 300 ^b	71.0	32.8
ICE-GB: 70 unscripted monologues (British NSs)	153,646	79	5.1	66	83.5	4.3
ICE-GB: 90 private direct conversations (British NSs)	185,000	819	44.3	244 out of 300 ^c	81.3	36.0

* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B per 10,000 words are based on an extrapolation of the percentages of the Type B phrase.

a, b and c in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

Table 7.2: Frequency information of *I mean* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percentage (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	56	1.7	51	91.1	1.5
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	317	5.3	243	76.7	4.1
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	66	4.9	57	86.4	4.3
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	1,702	29.4	263 out of 300 ^a	87.7	25.8
ICE-GB: 70 unscripted monologues (British NSs)	153,646	21	1.4	14	66.7	0.9
ICE-GB: 90 private direct conversations (British NSs)	185,000	865	46.8	281 out of 300 ^b	93.7	43.8

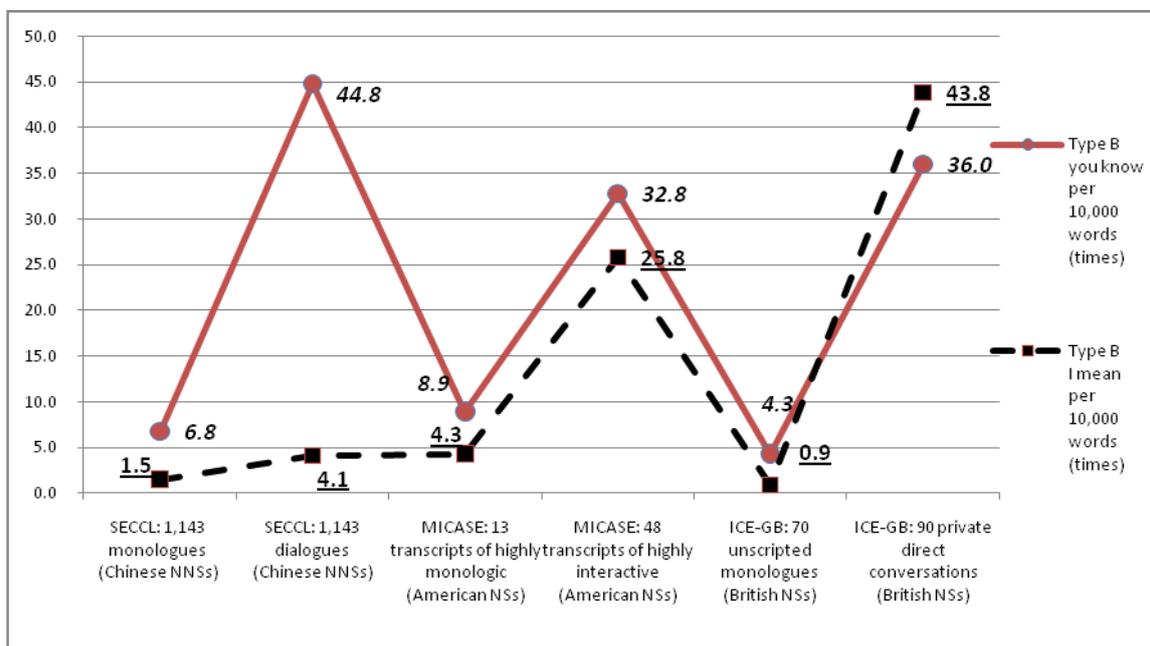
* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B per 10,000 words are based on an extrapolation of the percentages of the Type B phrase.

a and b in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

The raw frequencies of Type B are normed on a basis of 10,000 words and the normalised frequencies, ranging from 4.3 to 44.8 times for *you know* and from 0.9 to 43.8 times for *I mean*, are shown in Tables 7.1, 7.2 above and Figure 7.1 below. It is clear that there are slightly more instances of *you know* in the NNSs' dialogues than the NSs' speech and vice versa in the case of *I mean*. Interestingly, in terms of genre, there are apparently more instances of *you know* and *I mean* in the dialogic genres than in the monologic genres. This supports my hypothesis that the more interactive the genres or types of activity are, the more DMs occur.

Figure 7.1: Comparison of normalised frequencies of Type B *you know* and *I mean* across sub-corpora



As explained in Section 3.3.3 of Chapter 3, the tables in Appendix 6 present the results of statistical significance tests between the two types of genre and between the speech of the NNSs and NSs. In the case of *you know*, it is found that there is a statistically significant relationship between the two types of genre in SECCL (LL: -1259.88, p-value: < 0.0001), MICASE (LL: -284.35, p-value: < 0.0001) and ICE-GB (LL: -466.18, p-value: < 0.0001). The negative LL scores indicate that Type B *you know* is under-represented in the monologic genres. This supports the conclusion discussed previously that the more interactive the genre is, the more instances of Type B *you know* occur. Between the two groups of speakers, the differences in the monologic genres are not statistically significant (LL: -5.5 between Corpora A1 and B1 and LL: +11.53 between Corpora A1 and C1). However, in the dialogic genres, Type B *you know* is over-represented in the Chinese NNSs' dialogues, as opposed to the NSs' dialogic genres (LL: +110.11, p-value: < 0.0001 between Corpora A2 and B2 and LL: +26.91, p-value: < 0.0001 between Corpora A2 and C2).

In the case of Type B *I mean*, the differences between the two types of genre is statistically significant in the speech of the Chinese NNSs (LL: -50.06, p-value: < 0.0001 between Corpora A1 and A2) as well as in that of the NSs (LL: -324.92, p-value: < 0.0001 between Corpora B1 and B2; LL: -860.89, p-value: < 0.0001 between Corpora C1 and C2).

As with *you know*, *I mean* is also under-represented in the monologic genres. When the sub-corpus of the Chinese NNSs' monologues is compared with the two sub-corpora of the NSs' monologic genres, the difference between SECCL and MICASE is statistically significant, with the LL value of -27.91 (p-value: < 0.0001), but between SECCL and ICE-GB, the difference is not significant. In contrast, the differences between the Chinese NNSs' dialogues and the NSs' dialogic genres are highly significant (LL: -1040.84, p-value: < 0.0001 between Corpora A2 and B2; LL: -1330.43, p-value: < 0.0001 between Corpora A2 and C2). This shows that *I mean* is under-represented in the dialogues of the Chinese NNSs.

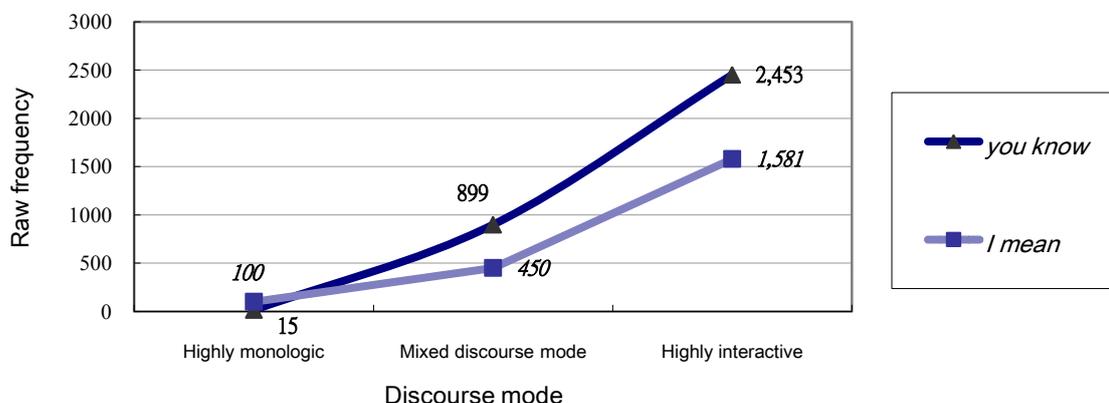
One contributing factor in the under-representation of *I mean* in the Chinese NNSs' speech may be the nature of test language. The NNSs' speech under investigation is likely to have been practised before recording, which leads to less use of *I mean* as a DM co-occurring with clarifications, explanations and elaborations (the major types of co-occurrence of *I mean*). Another possibility is that the NNSs' monologues are peculiar in respect of the absence of listeners. This is very different from the NSs' monologues, for example, lectures and colloquia, which are usually spoken to someone or a group of people. The NNSs' dialogues, in contrast, are more interactive, in that one listener is involved and the two speakers are required to take turns. This task is closer to the NSs' conversations for analysis. The presence of hearers may lead to the use of more DMs in order to manage turn taking, perform face-saving acts, mitigate criticism and the like.

The presence of interlocutors and the nature of the interactivity in the dialogues are assumed to be the reason why more instances of *you know* and *I mean* occur there than in the monologues. A quick on-line search in MICASE was done to check this assumption.

Since *you know* and *I mean* are primarily used as DMs in NSs' conversations (Biber *et al.* 1999: 1096), the raw frequencies of *you know* and *I mean* (including both Type A and Type B) in the three different discourse modes – highly monologic, mixed and highly interactive – in MICASE are compared, as in Table 7.3 below. (Mixed discourse mode means discourse in which neither monologic discourse nor interactive discourse predominates.) It is evident that the more interactive the discourse is, the more *you know* and *I mean* occur. Therefore, it can reasonably be argued that the degree of interactivity, as defined by MICASE, in a text affects how often *you know* and *I mean* are used.

Table 7.3: Raw frequencies of *you know* and *I mean* in three different discourse modes in MICASE

Discourse mode\Phrase	<i>you know</i>	<i>I mean</i>
Highly monologic discourse mode	15	100
Mixed discourse mode	899	450
Highly interactive discourse mode	2,453	1,581



The degree of interactivity in a text is identified as one of the factors in the use of *you know* and *I mean*. This means that comparing frequency requires direct comparable corpora, which are difficult to find. Therefore, frequency information comparison tends to be unreliable and the discussion of the phenomenon of overuse and underuse of DMs in the NNS speech is often misleading. In the present study, the frequency information is used as an entry point into the data and is treated with caution, thereby facilitating further analyses of the use of DMs. (The issue of comparability and the use of neutral terms *over-* and *under-representation* are discussed in Section 2.2.1 of Chapter 2.)

Another attribution to the use of *you know* and *I mean* is the idiosyncrasies of individual speakers. The initial investigation indicates that the occurrences of *you know* and *I mean* are unevenly distributed in different texts. High users and low users of Type B *you know* and *I mean* can be identified. More discussion is provided in Chapter 10, based on a more qualitative text-based analysis.

An example of a high user was found in the NSs' speech of highly monologic discourse mode. A closer look at the distribution of *you know* and *I mean* in the 14 texts of the sub-corpus of the NSs' highly monologic discourse mode revealed an unusual text, which was not used for analysis. More than half the occurrences of *you know* (225 out of 388) and *I mean* (64 out of 130) in this sub-corpus were from this text, LES495JU063, which was a lecture, produced by a senior graduate with near-native proficiency. Excerpt (7.3.1) below is divided

into seven stretches with *you know* and *I mean* in boldface. The speaker seems to be making a pretence of coherence and, to a listener, the utterance seems to make sense, but it is not completely coherent. In Items (2) and (6), there are no main clauses and *you know* co-occurs with illegitimate ends. This case shows that *you know* and *I mean* occurring in contexts which lack coherence may hinder the progress of listeners' understanding.

(7.3.1)

- | | | |
|---|------------------------------|--|
| <p>(1) but see what we ha- what we have here, is you know ultimately you know long term animosity,</p> <p>(2) and, you know a lot of people you know if you if you ask them in Eastern Europe they'll,</p> <p>(3) these days i mean they're not very keen, on remembering either of course,</p> <p>(4) i mean those who were older of course they're not very keen on on remembering German domination, but they're ultimately the also not very keen on when it comes well,</p> <p>(5) you know um did you like the Soviet troops there?</p> <p>(6) mkay. so, even these days, and of course you know , if you think about it this long term in- animosity, and anxiety and fear installed,</p> <p>(7) i mean , if that wasn't there, if if the Soviets had established some you know common interest, more on legitimacy rather than based on force, f- you know , the the the the Poles, the Hungarians and the Czechs these days, or what is it a year ago or two years ago? it was last year when they were admitted to NATO? i mean that's, why, that's part of the reason why Eastern European countries these days are now very, uh you know , are very vociferous in asking for, membership in NATO and the European Union.</p> | <p>(MICASE: LES495JU063)</p> | <p>No main clause;
with illegitimate
end</p> <p>No main clause</p> |
|---|------------------------------|--|

Another example of a high user of *you know* is found in Text S2A-050¹⁴ in the sub-corpus of the unscripted monologues in ICE-GB. The high frequency of *you know* in this text skews the result in favour of Type B *you know*. In this text, of the 44 instances of *you know*, 43 instances are of Type B. These instances account for 65% of the 66 instances in the whole sub-corpus (see Table 7.2 above). If they are taken out, only 23 (29.1%) instances out of 79 are used as a DM would be left and this leads to the conclusion that *you know* is not primarily used as a DM in this sub-corpus. In addition, the 23 instances occur in 13 (out of the 70) unscripted monologues in ICE-GB. Therefore, it is reasonable to argue that Type B *you know* is under-represented in this sub-corpus, which would change one of the previous conclusions.

¹⁴ See Section 10.3.1 of Chapter 10 for the text-based analysis of Text S2A-050.

7.3.2 Overall frequency of *you see*

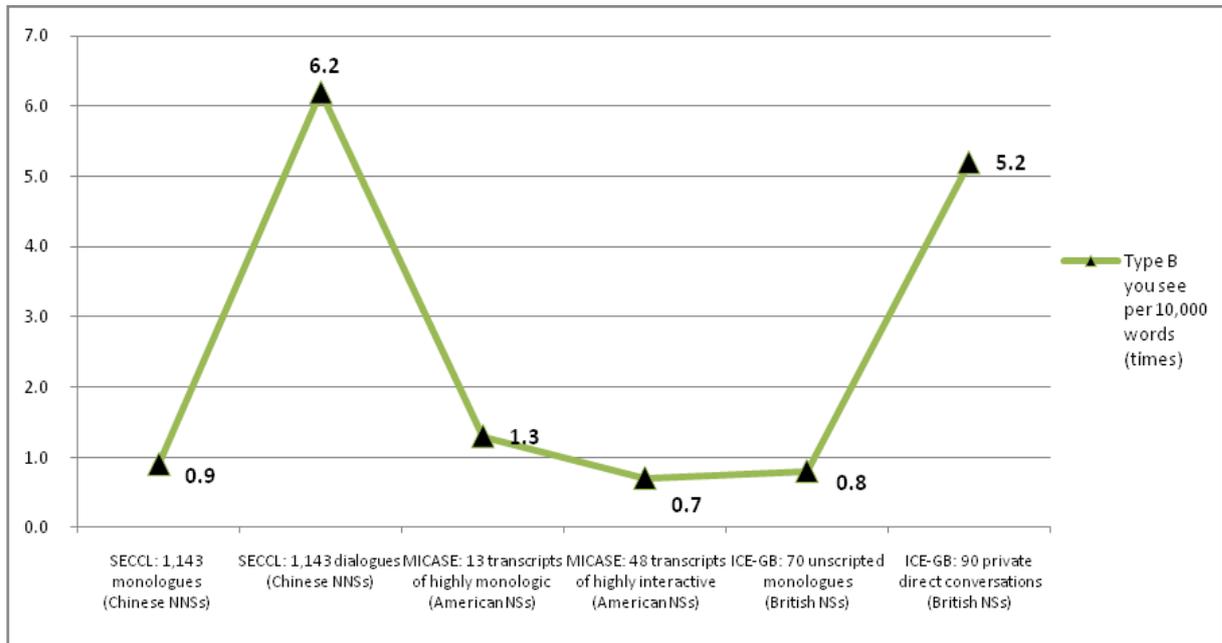
All the instances of *you see* are manually grouped into Types A and B. This classification reveals that *you see* is primarily used as a DM in the NNSs' speech, but not in the NSs' speech. This is different from Biber *et al.*'s finding (1999: 1097) that *you see* is typically a DM in NSs' conversations. In the NNSs' speech, over 74.4% of the instances of *you see* are used as a DM, while in the NSs' speech, the proportion of Type B *you see* ranges from 17.1% to 67.4%.

Table 7.4: Frequency information of *you see* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percentage (%)	Normalised freq. of Type B per 10,000 words (times)
SECCL: 1,143 monologues (Chinese NNSs)	336,303	39	1.2	29	74.4	0.9
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	452	7.6	403	89.2	6.2
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	58	4.3	17	29.3	1.3
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	228	3.9	39	17.1	0.7
ICE-GB: 70 unscripted monologues (British NSs)	153,646	35	2.3	13	37.1	0.8
ICE-GB: 90 private direct conversations (British NSs)	185,000	144	7.8	97	67.4	5.2

The raw frequencies of Type B *you see* are normed on a basis of 10,000 words and the normalised frequencies, ranging from 0.7 to 6.2 times. In terms of genre, there are apparently more instances of Type B *you see* in the dialogic genres than in the monologic genres, except in the two sub-corpora of MICASE.

Figure 7.2: Comparison of normalised frequencies of Type B *you see* across sub-corpora



The results of the test of statistical significance (see Appendix 6) indicate that there is a statistically significant relationship between the two types of genre in SECCL (LL: -206.81, p-value: < 0.0001) and in ICE-GB (LL: -57.92, p-value: < 0.0001), but in MICASE, the difference between the two types of genre is not significant (LL: +4.29). In the monologic genres, no statistical significance lies in the comparison between the speech of the Chinese NNSs and that of the NSs (LL: -1.53 between Corpora A1 and B1; LL: 0 between Corpora A1 and C1). The same is also found in the difference in dialogic genre between SECCL and ICE-GB (LL: +5.28). The difference between SECCL and MICASE is, however, statistically significant (LL: +337.48, p-value: < 0.0001). In the case of *you see*, the phenomenon of under-representation in the monologic genres is less obvious than *you know* and *I mean*, and the comparisons between the speech of the Chinese NNSs and the NSs in ICE-GB is not statistically significant.

7.3.3 Collocates of *you know*

Tables 7.5 to 7.10 present the patterns of all the instances of *you know* in the six subsets of SECCL, MICASE and ICE-GB under investigation. The patterns of the three monologic genres, shown in Tables 7.5 to 7.7, reveal some differences in the use of the phrase *you know* between the NNSs and NSs. In Table 7.5, it can be seen that the use of *do you know* is

relatively frequent. Further investigation reveals that most of them are reported speech. The hesitation marker *eh* is also one of the frequent collocates immediately to the left. In Tables 7.6 and 7.7, the highlighted collocates, *of*, *what* and *that*, seem to be indications of the use of Type A *you know*.

Table 7.5: Pattern of *you know* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (20)	very (15)	the (12)	do (22)	you know (301)	i (43)	i (20)	the (15)	i (16)
2	to (18)	the (13)	very (10)	eh (20)		the (24)	was (18)	i (15)	very (13)
3	and (13)	was (11)	my (8)	and (13)		that (23)	the (16)	a (15)	a (9)
4	very (11)	to (11)	you (8)	but (12)		eh (14)	is (13)	time (14)	the (9)
5	the (10)	a (11)	eh (8)	um (9)		my (12)	that (12)	very (11)	was (7)
6	a (9)	you (10)	said (7)	because (8)		it (12)	are (9)	is (9)	you (7)
7	was (8)	in (9)	but (7)	as (7)		in (12)	you (9)	was (8)	to (7)
8	it (6)	and (9)	in (6)	me (6)		at (12)	my (7)	in (8)	in (7)
9	my (6)	i (8)	i (6)	that (5)		and (10)	we (5)	you (6)	my (6)
10	he (5)	my (7)	me (6)	so (4)		you (9)	have (4)	have (6)	that (6)

Table 7.6: Pattern of *you know* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (10)	to (7)	a (6)	and (10)	you know (162)	that (12)	the (9)	that (8)	the (8)
2	of (5)	is (6)	that (5)	of (10)		the (9)	this (6)	the (6)	you (7)
3	or (4)	a (6)	but (5)	uh (6)		what (8)	i (6)	a (6)	to (6)
4	a (4)	in (5)	he (4)	that (5)		if (7)	a (6)	of (3)	a (5)
5	so (3)	the (5)	this (4)	again (4)		i (6)	is (4)	or (3)	of (4)
6	they (3)	you (4)	is (4)	um (4)		it's (6)	of (4)	see (2)	and (4)
7	um (3)	that (3)	most (3)	the (4)		and (6)	it (4)	sort (2)	uh (3)
8	which (3)	and (3)	so (3)	a (4)		a (4)	that (3)	they (2)	that (3)
9	this (3)	of (3)	some (3)	saying (3)		you (4)	you (3)	was (2)	right (2)
10	to (3)	as (3)	to (3)	do (3)		there's (3)	as (3)	years (2)	you're (2)

Table 7.7: Pattern of *you know* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (8)	and (6)	of (7)	of (5)	you know (79)	and (10)	the (9)	of (5)	it (3)
2	in (4)	the (5)	you (4)	as (4)		they (5)	i (7)	the (5)	to (2)
3	were (3)	of (4)	that (2)	and (3)		the (5)	uhm (4)	s (3)	this (2)
4	this (3)	a (4)	the (2)	uhm (3)		i (4)	and (3)	two (3)	was (2)
5	of (3)	i (3)	uhm (2)	uh (2)		that (4)	s (3)	you (3)	uh (2)
6	to (2)	which (2)	s (2)	you (2)		this (3)	would (2)	thought (2)	there (2)
7	and (2)	uh (2)	bit (2)	thought (2)		a (3)	this (2)	very (2)	m (2)
8	as (2)	to (2)	all (2)	in (2)		in (3)	fact (2)	was (2)	i (2)
9	s (2)	it (2)	a (2)	so (2)		where (2)	d (2)	all (2)	that (2)
10	day (2)	kind (2)	i (2)	that (2)		you (2)	lot (2)	a (2)	s (2)

In Table 7.8, the collocates to the left and to the right, *a* and *b*, are mostly the identification of the two speakers in the NNSs' dialogues. They indicate that *you know* is used in turn-initial or -final positions. In Tables 7.9 and 7.10, the highlighted collocates indicate the use of Type A *you know*, e.g. *do you know*, *you know what I mean*, *you know where/what* and *(that) you know (that)*, in the NSs' speech.

Table 7.8: Pattern of *you know* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (15)	a (16)	a (25)	but (41)	you know (246)	i (39)	i (15)	a (27)	a (11)
2	i (13)	b (13)	b (16)	a (20)		the (18)	you (13)	i (18)	are (8)
3	to (11)	the (12)	the (8)	eh (19)		b (14)	is (11)	you (9)	i (8)
4	b (11)	is (7)	but (8)	b (10)		a (13)	the (10)	is (8)	freshman (8)
5	the (8)	of (6)	very (6)	and (9)		eh (12)	m (9)	the (7)	the (7)
6	in (7)	very (5)	you (6)	because (8)		it (11)	have (7)	are (6)	think (7)
7	is (7)	good (5)	um (6)	um (5)		in (11)	are (7)	many (6)	it (5)
8	think (5)	in (5)	my (6)	major (5)		we (10)	think (6)	in (6)	for (4)
9	it (4)	i (5)	yeah (5)	future (3)		um (8)	we (6)	of (5)	is (4)
10	but (4)	you (5)	it (4)	oh (3)		you (8)	s (5)	to (5)	you (4)

Table 7.9: Pattern of *you know* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (9)	you (11)	you (11)	like (16)	you know (289)	what (31)	i (19)	you (11)	you (9)
2	the (9)	the (10)	and (8)	do (12)		the (17)	you (13)	mean (10)	a (8)
3	to (9)	know (9)	that (8)	so (9)		i (11)	the (12)	of (9)	know (7)
4	you (8)	like (7)	a (7)	um (7)		it's (11)	a (8)	and (9)	that (6)
5	i (6)	that (7)	of (6)	that (6)		that (11)	just (6)	the (8)	like (5)
6	like (6)	is (6)	um (6)	and (6)		and (10)	it (5)	to (7)	the (5)
7	do (5)	do (6)	the (5)	think (5)		like (9)	i'm (5)	like (6)	of (4)
8	of (5)	just (5)	it (5)	uh (5)		if (8)	it's (5)	a (6)	it's (4)
9	so (5)	to (5)	because (5)	but (5)		where (8)	that (5)	know (5)	i (4)
10	a (4)	of (5)	this (5)	is (4)		you (8)	this (5)	have (4)	to (4)

Table 7.10: Pattern of *you know* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (12)	to (10)	<i>i</i> (16)	and (17)	you know (300)	<i>i</i> (26)	<i>i</i> (25)	s (12)	of (11)
2	you (12)	the (9)	it (12)	it (15)		you (15)	you (18)	that (10)	it (10)
3	<i>i</i> (12)	<i>i</i> (9)	to (11)	uhm (15)		that (14)	s (14)	the (10)	the (10)
4	it (10)	and (9)	the (11)	mean (14)		what (13)	the (12)	mean (7)	s (9)
5	to (7)	that (8)	and (11)	do (13)		the (11)	that (8)	a (7)	to (7)
6	they (7)	s (7)	that (8)	d (9)		they (10)	it (8)	and (7)	is (7)
7	s (7)	it (6)	a (7)	but (7)		and (9)	a (7)	it (6)	that (6)
8	that (5)	in (5)	how (6)	just (4)		when (9)	are (6)	of (6)	<i>i</i> (6)
9	know (5)	but (5)	in (6)	then (4)		it (9)	is (6)	all (5)	you (6)
10	a (5)	have (5)	all (6)	that (4)		so (8)	to (6)	<i>i</i> (5)	a (6)

To look more closely at the use of Type B, the patterns of Type B *you know* in the NNSs' and NSs' speech are presented. In the NNSs' speech (see Tables 7.11 and 7.12), *you know* often co-occurs with the hesitation markers *eh* and *um*. Other frequent collocates are *and* and *but*. These two could be DMs co-occurring with *you know*.

Table 7.11: Pattern of Type B *you know* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (16)	very (14)	the (10)	eh (19)	you know (227)	<i>i</i> (38)	was (16)	time (13)	very (12)
2	<i>i</i> (16)	was (11)	very (10)	and (11)		the (16)	the (15)	the (13)	<i>i</i> (12)
3	very (10)	to (11)	my (8)	but (10)		eh (13)	<i>i</i> (12)	very (11)	a (8)
4	and (9)	a (9)	eh (7)	um (7)		at (12)	that (11)	a (11)	was (7)
5	was (8)	the (7)	<i>i</i> (6)	because (7)		my (11)	is (11)	<i>i</i> (10)	that (6)
6	a (8)	<i>i</i> (7)	you (5)	it (4)		in (11)	are (8)	was (8)	the (6)
7	my (5)	and (7)	good (5)	me (4)		it (11)	you (7)	in (8)	in (6)
8	it (5)	in (6)	in (5)	that (4)		and (8)	my (6)	is (5)	to (5)
9	the (5)	you (6)	and (4)	school (4)		he (7)	we (4)	have (5)	my (4)
10	he (4)	me (5)	one (4)	said (3)		you (7)	m (4)	not (5)	good (4)

Table 7.12: Pattern of Type B *you know* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (15)	a (16)	a (25)	but (41)	you know (246)	<i>i</i> (39)	<i>i</i> (15)	a (27)	a (11)
2	<i>i</i> (13)	b (13)	b (16)	a (20)		the (18)	you (13)	<i>i</i> (18)	are (8)
3	to (11)	the (12)	the (8)	eh (19)		b (14)	is (11)	you (9)	<i>i</i> (8)
4	b (11)	is (7)	but (8)	b (10)		a (13)	the (10)	is (8)	freshman (8)
5	the (8)	of (6)	very (6)	and (9)		eh (12)	m (9)	the (7)	the (7)
6	in (7)	very (5)	you (6)	because (8)		it (11)	have (7)	are (6)	think (7)
7	is (7)	good (5)	um (6)	um (5)		in (11)	are (7)	many (6)	it (5)
8	think (5)	in (5)	my (6)	major (5)		we (10)	think (6)	in (6)	for (4)
9	it (4)	<i>i</i> (5)	yeah (5)	future (3)		um (8)	we (6)	of (5)	is (4)
10	but (4)	you (5)	it (4)	oh (3)		you (8)	s (5)	to (5)	you (4)

In the NSs' speech, it is also found that Type B *you know* co-occurs with the hesitation markers *um* and *uh*. Additionally, more varied DM collocations are found, such as *well you know*, *but you know*, *you know and*, *like you know* and *I mean you know*.

Table 7.13: Pattern of Type B *you know* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (9)	to (5)	but (5)	uh (6)	you know (119)	if (6)	the (7)	that (6)	the (5)
2	a (4)	a (5)	he (4)	the (4)		and (6)	a (4)	a (6)	you (5)
3	or (3)	the (4)	this (4)	um (4)		the (6)	in (3)	the (3)	a (5)
4	of (3)	in (4)	that (4)	a (4)		i (5)	is (3)	or (3)	and (4)
5	um (3)	you (4)	is (4)	again (4)		it's (4)	this (3)	see (2)	of (3)
6	so (3)	is (3)	a (4)	and (4)		a (4)	i (3)	was (2)	there (2)
7	this (2)	well (2)	they (3)	well (3)		uh (3)	as (2)	years (2)	you're (2)
8	are (2)	uh (2)	the (3)	saying (3)		there's (3)	basically (2)	this (2)	uh (2)
9	to (2)	augustus (2)	to (3)	is (3)		he (3)	you (2)	uh (2)	to (2)
10	with (2)	and (2)	time (3)	was (2)		what (3)	to (2)	in (2)	that (2)

Table 7.14: Pattern of Type B *you know* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (9)	the (8)	and (7)	like (12)	you know (206)	it's (11)	the (8)	and (9)	a (8)
2	you (8)	know (7)	that (6)	so (8)		the (11)	a (7)	you (9)	know (6)
3	and (8)	you (6)	a (6)	um (5)		i (11)	just (6)	like (6)	you (5)
4	to (6)	like (5)	the (5)	and (5)		and (10)	i (5)	to (6)	the (5)
5	like (5)	to (5)	of (5)	a (4)		like (9)	that (5)	the (5)	it's (4)
6	i (5)	that (5)	this (5)	is (4)		you (7)	like (5)	have (4)	of (4)
7	know (4)	is (4)	you (5)	but (4)		i'm (5)	and (5)	a (4)	that (4)
8	so (4)	just (4)	it (4)	that (4)		if (5)	don't (4)	of (4)	i (3)
9	of (4)	of (4)	um (4)	uh (4)		it (5)	it (3)	know (3)	uh (3)
10	it's (3)	see (3)	because (4)	it (3)		a (5)	it's (3)	then (3)	said (2)

Table 7.15: Pattern of Type B *you know* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (12)	i (9)	i (14)	uhm (13)	you know (244)	i (25)	i (15)	the (10)	s (9)
2	and (10)	to (8)	the (11)	and (13)		you (15)	s (13)	and (7)	the (9)
3	it (8)	and (8)	and (9)	it (13)		the (11)	you (13)	that (7)	it (8)
4	you (8)	it (6)	it (8)	mean (12)		they (10)	the (8)	s (6)	of (8)
5	they (6)	the (5)	a (7)	this (4)		it (9)	a (7)	it (6)	to (7)
6	to (5)	have (5)	to (7)	but (4)		when (9)	that (6)	re (5)	you (6)
7	a (5)	but (5)	that (7)	just (4)		so (7)	mean (6)	you (5)	that (6)
8	that (5)	s (5)	all (6)	said (3)		and (7)	are (6)	a (5)	i (5)
9	he (4)	in (5)	s (5)	of (3)		he (6)	think (5)	my (5)	a (5)
10	but (4)	of (4)	in (4)	right (3)		that (6)	is (5)	how (4)	he (4)

It is difficult to identify prominent collocates in the pattern of Type B *you know* in the sub-corpus of the unscripted monologues in ICE-GB (Table 7.16), because this genre is peculiar, in that, as discussed in the previous section, of the 66 instances of *you know* in Table 7.16, 43 instances are produced by one single speaker and also in that the low frequencies of collocates seem not to be prominent.

Table 7.16: Pattern of Type B *you know* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (8)	and (6)	of (6)	of (3)	you know (66)	and (10)	i (6)	of (5)	this (2)
2	in (4)	the (4)	the (2)	uhm (3)		the (5)	the (5)	the (5)	there (2)
3	were (2)	a (4)	that (2)	thought (2)		they (5)	s (3)	you (3)	uh (2)
4	to (2)	to (2)	you (2)	uh (2)		i (4)	uhm (3)	s (3)	to (2)
5	day (2)	which (2)	uhm (2)	and (2)		in (3)	and (2)	thought (2)	it (2)
6	of (2)	uh (2)	s (2)	in (2)		this (2)	would (2)	two (2)	i (2)
7	s (2)	it (2)	bit (2)	that (2)		you (2)	d (2)	a (2)	that (2)
8		i (2)	all (2)			a (2)	is (2)	all (2)	m (2)
9		of (2)	a (2)			it (2)	lot (2)	it (2)	
10		kind (2)	i (2)			cos (2)		and (2)	

The frequency information and collocates of *you know* are used as starting points. The frequencies of *you know* in the six sub-corpora reveal that the phrase *you know* is primarily used as a DM by the NNSs and NSs under investigation.

7.3.4 Collocates of *I mean*

The numbers of the occurrences of *I mean* in the monologic genres in SECCL (56), MICASE (66) and ICE-GB (21), shown in Table 7.2 above, are too low to reveal frequent collocates.

Therefore, the patterns of *I mean* in the monologic genres are omitted in this section.

Tables 7.17 to 7.19 present the patterns of all the instances of *I mean* in the three sub-corpora of the dialogic genres. The patterns of the Chinese NNSs' dialogues, the American NSs' highly interactive discourse mode and the British NSs' direct conversations reveal some differences in the use of the phrase *I mean* between the NNSs and NSs. In Table 7.17, it can be seen that the use of *what I mean* is frequent. The hesitation marker *eh* is also one of the frequent collocates immediately to the left. Tables 7.18 and 7.19 show that the cluster *you know what I mean* is frequent in the American NSs' speech and *I mean I* and *I mean it* occur often in both American and British NSs' speech.

In Table 7.17, the collocates to the left and to the right, *a* and *b*, are mostly the identification of the two speakers in the dialogues. They indicate that *I mean* is used in turn-initial or -final positions. This is attributed to the frequent turn-taking in the NNS data.

Table 7.17: Pattern of *I mean* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	you (12)	you (20)	the (20)	b (24)	i mean (315)	the (39)	you (18)	i (14)	the (18)
2	to (11)	a (20)	i (19)	a (18)		that (30)	the (13)	you (12)	a (11)
3	think (10)	the (18)	b (14)	what (17)		i (21)	can (12)	do (9)	your (9)
4	a (10)	b (17)	mean (14)	eh (16)		you (20)	i (12)	the (8)	b (8)
5	in (10)	i (15)	a (13)	i (13)		eh (14)	a (10)	mean (8)	can (8)
6	and (9)	eh (8)	to (8)	but (10)		b (12)	we (9)	in (8)	you (8)
7	have (9)	in (8)	know (8)	job (8)		a (11)	should (8)	a (8)	and (7)
8	do (8)	some (7)	yes (7)	um (8)		in (9)	eh (8)	of (6)	is (7)
9	the (8)	to (6)	think (6)	mean (7)		we (7)	mean (7)	b (6)	to (7)
10	eh (7)	study (6)	um (5)	life (6)		is (6)	to (7)	some (6)	in (7)

Table 7.18: Pattern of *I mean* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	to (6)	you (21)	know (18)	what (21)	i mean (279)	i (33)	you (14)	i (15)	the (8)
2	is (6)	the (14)	the (9)	but (12)		it (15)	the (13)	you (11)	i (8)
3	i (5)	and (6)	of (9)	so (11)		you (14)	i (12)	the (10)	a (8)
4	do (5)	of (5)	and (8)	know (8)		if (14)	is (10)	it (8)	to (7)
5	in (5)	a (5)	i (7)	right (6)		it's (14)	a (7)	of (6)	of (7)
6	the (5)	to (5)	you (6)	well (5)		like (10)	just (7)	a (6)	but (7)
7	that (5)	i (4)	in (6)	um (5)		and (9)	like (7)	know (6)	like (6)
8	of (5)	is (3)	this (5)	it (4)		this (7)	not (6)	to (6)	that (6)
9	you (5)	know (3)	don't (4)	that (4)		for (7)	it's (6)	just (5)	know (4)
10	she (3)	that (3)	think (4)	yeah (4)		the (7)	know (6)	is (4)	was (4)

Table 7.19: Pattern of *I mean* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (24)	it (14)	you (12)	but (21)	i mean (300)	i (72)	s (36)	it (13)	to (12)
2	of (13)	i (10)	to (10)	yeah (15)		it (38)	i (17)	i (10)	s (11)
3	that (9)	to (9)	yeah (9)	know (12)		that (16)	you (13)	of (9)	that (10)
4	the (9)	you (8)	a (8)	well (12)		you (15)	think (12)	you (8)	the (10)
5	a (9)	the (8)	i (8)	no (10)		the (12)	know (8)	the (8)	you (8)
6	it (7)	s (7)	s (7)	yes (10)		there (9)	m (7)	is (8)	i (7)
7	you (6)	and (7)	as (7)	uhm (9)		he (7)	would (6)	a (8)	it (7)
8	and (5)	that (7)	but (7)	it (7)		if (7)	was (6)	not (7)	of (7)
9	know (4)	of (6)	that (6)	uh (7)		we (7)	it (6)	do (7)	got (6)
10	really (4)	uhm (5)	and (6)	and (7)		a (7)	that (5)	they (7)	a (6)

Tables 7.20 to 7.22 present the patterns of the instances of Type B *I mean* in the three sub-corpora of the dialogic genres. The co-occurring DMs are highlighted. The DM collocations found in the NNSs' speech are *I mean I mean* and *but I mean*, whereas those found in the NSs' speech are varied, such as *but I mean*, *so I mean*, *well I mean*, *yeah (yes) I mean*, *no I mean*, and *I mean*, *you know I mean*, *I mean like*, *I mean and* and *I mean you know*.

Table 7.20: Pattern of Type B *I mean* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	have (9)	the (17)	the (18)	b (21)	i mean (241)	the (30)	you (14)	i (12)	the (18)
2	think (8)	a (16)	i (17)	a (15)		you (19)	can (10)	you (9)	your (7)
3	and (8)	b (15)	a (8)	eh (15)		i (19)	the (10)	the (8)	can (6)
4	you (8)	you (10)	b (7)	i (9)		eh (12)	a (8)	a (8)	to (6)
5	a (7)	i (10)	mean (6)	mean (7)		b (10)	should (8)	in (7)	should (5)
6	in (7)	in (7)	um (5)	but (7)		in (9)	eh (8)	do (6)	in (5)
7	to (7)	to (6)	to (5)	job (5)		we (7)	mean (7)	mean (6)	they (5)
8	um (6)	eh (6)	think (5)	life (5)		just (6)	i (7)	eh (5)	is (5)
9	eh (6)	some (6)	your (4)	um (5)		when (6)	to (6)	t (5)	a (5)
10	the (5)	are (5)	yes (4)	students (4)		a (6)	are (5)	of (5)	you (5)

Table 7.21: Pattern of Type B *I mean* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	is (6)	the (13)	the (9)	but (12)	i mean (244)	i (30)	the (12)	i (13)	i (8)
2	you (5)	you (6)	of (9)	so (11)		it's (14)	i (11)	the (10)	a (8)
3	to (5)	of (5)	i (7)	right (6)		you (14)	you (10)	you (9)	the (7)
4	of (5)	and (5)	and (6)	well (5)		if (13)	is (9)	it (7)	but (7)
5	the (5)	a (5)	this (5)	know (4)		it (11)	just (7)	to (6)	of (6)
6	that (4)	to (5)	in (5)	yeah (4)		the (7)	a (7)	of (6)	to (6)
7	i (4)	is (3)	don't (4)	um (4)		that's (7)	it's (6)	a (5)	that (5)
8	in (4)	i (3)	heavy (4)	it (4)		like (7)	not (6)	just (5)	know (4)
9	and (3)	know (3)	think (4)	like (3)		for (7)	know (6)	that (4)	was (4)
10	are (3)	that (3)	no (3)	that (3)		and (6)	can (6)	is (4)	want (4)

Table 7.22: Pattern of Type B *I mean* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (23)	it (13)	to (10)	but (20)	i mean (281)	i (69)	s (35)	it (12)	to (12)
2	of (13)	i (9)	yeah (9)	yeah (15)		it (37)	i (15)	i (9)	s (11)
3	that (9)	to (9)	you (9)	well (12)		you (14)	you (13)	the (8)	that (10)
4	the (8)	the (8)	a (8)	yes (10)		that (13)	think (12)	a (8)	the (8)
5	a (8)	and (7)	i (7)	no (10)		the (11)	know (8)	not (7)	of (7)
6	you (6)	s (7)	but (7)	uhm (9)		there (8)	m (7)	of (7)	i (7)
7	it (6)	of (6)	as (7)	know (9)		he (7)	would (6)	is (7)	it (6)
8	and (5)	uh (5)	that (6)	and (7)		we (7)	was (6)	they (6)	a (6)
9	to (4)	you (5)	it (6)	uh (7)		if (6)	it (5)	you (6)	you (6)
10	know (4)	uhm (4)	all (5)	it (7)		they (6)	re (4)	so (6)	re (6)

7.3.5 Collocates of *you see*

Similar to the case of *I mean*, the patterns of *you see* in the monologic genres in SECCL (39), MICASE (58) and ICE-GB (35) are omitted here because of the low frequencies and the very few frequent collocates.

Tables 7.23 to 7.25 present the patterns of all the instances of *you see* in the three sub-corpora of the dialogic genres. Like the cases of *you know* and *I mean*, the collocates of *you see* in the NNSs' dialogues are not as varied as those in the NSs' speech. In Table 7.23, it can be seen that *because you see* and *do you see*, the use of Type A *you see*, are frequent. Tables 7.24 and 7.25 show that the clusters *do (did) you see*, *can you see*, *what you see*, *when you see*, *that you see*, *if you see*, *you see what I mean* and *you see what I'm saying* are common in the speech of the NSs. Most of these collocates suggest that *you see* is Type A, serving as the subject and verb in a clause. This finding correlates with the manual classification of Types A and B. *You see* is not primarily used as a DM in the speech of the

NSs.

In Table 7.23, the collocates to the left, *a* and *b*, are mostly the identification of the two speakers in the dialogues. They indicate that *you see* is used in turn-initial position. This use is also found in the cases of *you know* and *I mean*.

Table 7.23: Pattern of *you see* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (37)	a (46)	b (42)	but (67)	you see (452)	the (48)	i (35)	is (20)	the (18)
2	i (23)	you (14)	a (42)	eh (28)		i (39)	you (23)	the (19)	a (15)
3	the (20)	i (13)	but (19)	b (23)		eh (33)	the (18)	have (16)	to (12)
4	b (18)	b (13)	you (18)	and (23)		that (26)	is (17)	i (16)	i (11)
5	you (16)	yes (13)	think (13)	a (18)		if (23)	are (15)	a (16)	are (10)
6	is (9)	to (10)	yeah (13)	because (17)		um (18)	in (15)	you (12)	is (10)
7	not (9)	of (9)	yes (9)	you (17)		a (18)	m (13)	are (11)	you (9)
8	to (9)	in (8)	i (8)	so (17)		you (15)	we (12)	in (10)	and (8)
9	eh (7)	yeah (8)	eh (8)	um (14)		there (11)	students (12)	we (9)	so (7)
10	do (7)	but (7)	the (8)	do (9)		it (11)	have (11)	of (9)	of (7)

Table 7.24: Pattern of *you see* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (7)	the (10)	see (8)	do (46)	you see (220)	what (36)	i'm (22)	saying (15)	the (15)
2	the (6)	you (9)	what (8)	what (10)		the (24)	i (15)	mean (11)	of (7)
3	a (6)	it (7)	the (6)	can (10)		this (16)	the (14)	see (6)	you (5)
4	you (6)	of (5)	that (5)	that (10)		that (15)	you (7)	you (6)	like (5)
5	of (5)	see (5)	um (5)	did (9)		it (13)	in (5)	that (6)	see (4)
6	just (4)	to (5)	it (5)	and (7)		a (7)	this (4)	the (5)	this (4)
7	mean (4)	so (5)	but (5)	when (6)		i (5)	that (4)	is (4)	that (4)
8	to (4)	and (5)	and (4)	so (5)		in (5)	and (4)	then (3)	one (3)
9	in (4)	a (4)	like (4)	if (5)		so (4)	little (3)	and (3)	right (3)
10	at (4)	is (4)	you (3)	then (3)		where (4)	what (3)	of (3)	um (3)

Table 7.25: Pattern of *you see* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (6)	the (13)	the (9)	do (6)	you see (144)	i (19)	i (20)	mean (10)	to (7)
2	i (4)	s (9)	you (4)	if (6)		the (10)	s (9)	you (5)	you (7)
3	that (3)	of (6)	that (4)	that (5)		what (10)	you (7)	s (4)	i (6)
4	he (3)	that (4)	uhm (3)	me (5)		that (10)	mean (4)	i (3)	the (6)
5	do (3)	uhm (4)	s (3)	can (5)		and (9)	don't (4)	is (3)	in (4)
6	and (3)	you (4)	and (3)	well (5)		it (7)	they (3)	they (3)	it (4)
7	to (3)	and (3)	so (2)	and (4)		if (6)	there (3)	want (3)	oh (3)
8	of (3)	to (3)	yours (2)	but (4)		he (5)	was (3)	as (3)	that (3)
9	be (3)	a (3)	yes (2)	you (4)		you (4)	ve (3)	got (3)	she (3)
10	the (3)	is (3)	what (2)	did (3)		well (3)	d (3)	would (2)	s (3)

To further examine the use of Type B, the patterns of Type B *you see* in the dialogic genres of SECCL, MICASE and ICE-GB are presented in Tables 7.26 to 7.28. In the NNSs' dialogues, *you see* co-occurs with *but*, *and*, *so*, *yes* and *yeah*, while in the NSs' speech, *you see* tends to follow *and*, *well*, *but* and *now*. In the NSs' speech, *you see* is not primarily used as a DM and the patterns shown in Tables 7.27 and 7.28 are based on a small number of instances; therefore, it is difficult to identify DM collocations. In Table 7.28, the frequencies of possible DM collocations, *well you see*, *but you see* and *now you see*, are very low, to indicate importance.

Table 7.26: Pattern of Type B *you see* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (34)	a (42)	a (38)	but (63)	you see (403)	the (40)	i (32)	is (18)	a (14)
2	i (21)	you (14)	b (37)	eh (27)		i (38)	you (19)	the (16)	the (13)
3	the (18)	yes (13)	you (17)	b (23)		eh (33)	is (17)	i (15)	to (11)
4	b (16)	i (12)	but (14)	and (20)		if (22)	the (16)	a (15)	is (9)
5	you (15)	b (9)	think (11)	because (17)		um (18)	are (15)	have (15)	you (8)
6	not (9)	of (9)	yeah (11)	so (16)		a (16)	m (13)	in (10)	i (8)
7	to (8)	to (9)	yes (9)	a (16)		you (15)	in (12)	are (10)	and (8)
8	is (8)	in (8)	i (8)	you (16)		there (11)	students (12)	you (9)	many (7)
9	yes (7)	yeah (8)	um (7)	um (14)		in (10)	we (11)	think (9)	so (7)
10	do (6)	think (7)	eh (7)	yes (8)		b (9)	have (11)	not (8)	um (7)

Table 7.27: Pattern of Type B *you see* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	you (2)	the (2)	um (3)	and (4)	you see (36)	this (5)	the (2)	that (2)	the (6)
2	to (2)		of (2)	like (2)		i (4)	this (2)	in (2)	and (2)
3	at (2)					so (3)	is (2)		
4						um (2)	that (2)		
5						the (2)			
6						like (2)			
7						that's (2)			

Table 7.28: Pattern of Type B *you see* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (5)	s (9)	the (7)	me (5)	you see (97)	i (16)	i (10)	want (3)	to (7)
2	i (3)	of (5)	s (3)	well (5)		and (9)	s (9)	got (3)	i (5)
3	that (3)	the (5)	you (3)	but (4)		if (6)	you (6)	you (3)	you (5)
4	and (3)	that (4)	and (3)	now (3)		he (5)	mean (4)	s (3)	it (3)
5	to (3)	is (3)	uhm (2)	you (3)		it (4)	don't (4)	they (2)	the (3)
6	he (3)	i (3)	yes (2)	yeah (2)		the (4)	they (3)	would (2)	she (2)
7	the (2)	and (2)	about (2)	yes (2)		you (4)	ve (3)	used (2)	that (2)
8	there (2)	with (2)	for (2)	and (2)		that (4)	d (3)	as (2)	out (2)
9	s (2)	but (2)	that (2)	so (2)		well (3)	what (2)	a (2)	get (2)
10	mean (2)	do (2)	not (2)	that (2)		because (3)	it (2)	don't (2)	be (2)

From the above tables, it can be concluded that the collocates of *you know*, *I mean* and *you see* in the NSs' speech are more varied than those in the NNSs' speech. This is very possibly due to the test-taking context in the NNS data, in which the speakers are restricted to discussing certain topics (see Appendix 1 for the topics discussed by the NNSs).

The availability of speaker identification in the transcripts of the NNSs' speech helps to identify that *you know* and *I mean* tend to occur in utterance/turn-initial or -final positions and *you see* tends to be placed utterance/turn-initially. This piece of information in MICASE and ICE-GB is included in the markup annotation and cannot be retrieved and shown in the patterns. More discussion about where Type B *you know*, *I mean* and *you see* occur in an utterance/turn is given in Section 7.4.1 below.

7.4 Discourse aspects of *you know*, *I mean* and *you see*

In this section, the positions where Type B *you know*, *I mean* and *you see* occur in an utterance/turn are first described and then the linguistic items with which these three DMs tend to co-occur are discussed.

As pointed out in the chapter on methodology (see Section 3.3.7), in the cases where the instances numbered more than 400, three sets of 100-line concordance samples were used. This random sampling procedure has been demonstrated to be a sufficient basis.

7.4.1 Positions in an utterance/turn

The positions in an utterance of *you know*, *I mean* and *you see* are described in this section. The term *utterance* is used in this thesis to refer to a stretch of speech produced by one speaker. The two main categories of position are extra-clausal and intra-clausal. The

extra-clausal positions are further divided into three sub-categories – utterance/turn-initial, -medial and -final. In the NNSs' monologues, each text is taken as an utterance and, in the dialogues, a turn indicated by Speaker A or B is seen as an utterance. Type B *you know*, *I mean* and *you see* occurring at the beginning of an utterance are categorised under utterance/turn-initial, utterance/turn-medial and utterance/turn-final position; the first, utterance/turn-initial position, includes those utterances with one or two organisational (O) elements (e.g. *well* and *yeah*) preceding *you know*, *I mean* and *you see*. Those occurring at the end of an utterance/turn are called *utterance/turn-final* and the remainder appearing in any extra-clausal positions in an utterance belong to the group called *utterance/turn-medial*.

The occurrences of *you know*, *I mean* and *you see* in the intra-clausal positions are examined by *LUG* (Sinclair and Mauranen 2006), because conventional grammars cannot satisfactorily assign a unit in spoken English. *You know*, *I mean* and *you see* are grouped according to message-oriented elements (M) or O elements immediately preceding the instances. The distribution and percentages of Type B *you know*, *I mean* and *you see* in the six sub-corpora under investigation are shown in Tables 7.29, 7.30 and 7.31 below.

It is clear in Table 7.29 that there is a marked difference in the distribution of the occurrences of Type B *you know* between the two groups of speakers. In the NNSs' speech, *you know* most often appears in an extra-clausal position, accounting for 86.4% and 91.9% in the monologues and dialogues respectively, while 50% more or less of the occurrences of *you know* occur in an extra-clausal position in the NSs' speech. Although the proportions seem to be very different, the most common position of *you know* is utterance/turn-medial in both the NNSs' and NSs' speech.

In terms of genre, there seems to be no great difference in the distribution of the positions across the two types of genre in MICASE and ICE-GB. However, there is a significant contrast between the NNSs' monologues and their dialogues in using *you know* in an extra-clausal position. 85.5% of the occurrences of *you know* in the monologues are classified into utterance-medial, as opposed to about half (50.8%) of those in the dialogues. This is because of the nature of these two types of activity and the speech unit used in the classification. In the NNSs' monologues, there is only one utterance in a text, but in the dialogues, the two speakers in a text produce many more utterances, which provide many more chances for *you know* to occur in turn-initial or turn-final.

Table 7.29: Distribution of the positions in an utterance/turn of Type B *you know*

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues	
	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)
Positions in an utterance of Type B <i>you know</i>	228	100	119	100	66	100
Extra-clausal: utterance-initial	1	0.4	0	0.0	0	0.0
Extra-clausal: utterance-medial	195	85.5	54	45.4	34	51.5
Extra-clausal: utterance-final	1	0.4	0	0.0	0	0.0
Intra-clausal: after an M-	11	4.8	31	26.1	9	13.6
Intra-clausal: after an MA	3	1.3	2	1.7	2	3.0
Intra-clausal: after an MF	2	0.9	11	9.2	7	10.6
Intra-clausal: others	15	6.6	21	17.6	14	21.2
Unclassified	0	0.0	0	0.0	0	0.0
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations	
	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)
Positions in a turn of Type B <i>you know</i>	246	100	213	100	244	100
Extra-clausal: turn-initial	76	30.9	11	5.2	12	4.9
Extra-clausal: turn-medial	125	50.8	98	46.0	119	48.8
Extra-clausal: turn-final	25	10.2	7	3.3	22	9.0
Intra-clausal: after an M-	7	2.8	50	23.5	49	20.1
Intra-clausal: after an MA	2	0.8	8	3.8	2	0.8
Intra-clausal: after an MF	3	1.2	16	7.5	11	4.5
Intra-clausal: others	8	3.3	22	10.3	29	11.9
Unclassified	0	0.0	1	0.5	0	0.0

Table 7.30 illustrates the distribution of the positions in an utterance/turn of Type B *I mean*. It can be seen that the distribution of the occurrences of Type B *I mean* between the two groups of speakers is similar. *I mean* appears more often in an extra-clausal position, accounting for 56.9% and 68.7% of the instances in the monologues and the dialogues respectively and over 75% on average in the NSs' speech.

In terms of genre, there seems to be no great differences in the distribution of the positions across the two types of genre in SECCL and MICASE. However, the contrast between the British NSs' unscripted monologues and their direct conversations is sharp. 57.1% of the occurrences of *I mean* in the monologues are classified into utterance-medial, as opposed to 87.9% of those in the private direct conversations.

Table 7.30: Distribution of the positions in an utterance/turn of Type B / *mean*

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues				
	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)			
Positions in an utterance of Type B / <i>mean</i>	51	100	57	100	14	100			
Extra-clausal: utterance-initial	0	0.0	0	0.0	0	0.0			
Extra-clausal: utterance-medial	29	56.9	56.9	47	82.5	82.5	8	57.1	57.1
Extra-clausal: utterance-final	0	0.0	0	0.0	0	0.0			
Intra-clausal: after an M-	1	2.0	2	3.5	1	7.1			
Intra-clausal: after an MA	8	15.7	43.1	2	3.5	17.5	0	0.0	42.9
Intra-clausal: after an MF	2	3.9	3	5.3	2	14.3			
Intra-clausal: others	11	21.6	3	5.3	3	21.4			
Unclassified	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations				
	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)			
Positions in an utterance of Type B / <i>mean</i>	243	100	263	100	281	100			
Extra-clausal: turn-initial	54	22.2	60	22.8	52	18.5			
Extra-clausal: turn-medial	100	41.2	68.7	143	54.4	79.8	183	65.1	87.9
Extra-clausal: turn-final	13	5.3	7	2.7	12	4.3			
Intra-clausal: after an M-	23	9.5	20	7.6	5	1.8			
Intra-clausal: after an MA	5	2.1	31.3	7	2.7	20.2	1	0.4	12.1
Intra-clausal: after an MF	9	3.7	16	6.1	19	6.8			
Intra-clausal: others	39	16.0	10	3.8	9	3.2			
Unclassified	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0

You see is not a central DM and there are few instances of *you see* as a DM in the NSs' speech. This makes the information about percentages less useful. For instance, the distribution of *you see* in the NSs' highly monologic discourse mode and the unscripted monologues are based on 17 and 13 instances respectively. Their percentage information is less convincing.

In Table 7.31, it can be seen that the distribution of the occurrences of Type B *you see* between the two groups of speakers and across the two types of genre is similar. *You see* is mainly used in an extra-clausal position in the speech of both the NNSs and NSs.

Table 7.31: Distribution of the positions in an utterance/turn of Type B *you see*

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues		
	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	
Positions in an utterance of Type B <i>you see</i>	29	100	17	100	13	100	
Extra-clausal: utterance-initial	0	0.0	0	0.0	0	0.0	
Extra-clausal: utterance-medial	28	96.6	17	100.0	10	76.9	76.9
Extra-clausal: utterance-final	0	0.0	0	0.0	0	0.0	
Intra-clausal: after an M-	0	0.0	0	0.0	2	15.4	
Intra-clausal: after an MA	0	0.0	0	0.0	0	0.0	
Intra-clausal: after an MF	0	0.0	0	0.0	0	0.0	
Intra-clausal: others	1	3.4	0	0.0	1	7.7	
Unclassified	0	0.0	0	0.0	0	0.0	
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations		
	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	
Positions in a turn of Type B <i>you see</i>	403	100	39	100	97	100	
Extra-clausal: turn-initial	138	34.2	12	30.8	13	13.4	
Extra-clausal: turn-medial	235	58.3	19	48.7	69	71.1	93.8
Extra-clausal: turn-final	20	5.0	5	12.8	9	9.3	
Intra-clausal: after an M-	5	1.2	2	5.1	1	1.0	
Intra-clausal: after an MA	0	0.0	0	0.0	0	0.0	
Intra-clausal: after an MF	3	0.7	0	0.0	1	1.0	
Intra-clausal: others	2	0.5	1	2.6	4	4.1	
Unclassified	0	0.0	0	0.0	0	0.0	

Tables 7.29, 7.30 and 7.31 above reveal some similarities and differences across the monologic and dialogic genres and also between the speech of NNSs and that of NSs. More details about the differences in the use of *you know*, *I mean* and *you see* and the factors which may lead to these differences are discussed in Section 7.4.2.

7.4.1.1 *You know* in extra-clausal position

Generally, in the NNSs' and NSs' speech under investigation, about half the instances of Type B *you know* are placed in extra-clausal utterance/turn medial position, as in Excerpt (7.4.1). This is particularly common in the NNSs' monologues, accounting for 85.5%.

(7.4.1)

P: Utterance-medial

Task 2

I think I will never forget my 18th birthday. **[You know]**, at that time, I was just a middle school student.

(SECCL: B00-65-12)

In the three sub-corpora of the monologic genres, *you know* occurring in utterance-initial and -final positions is not found in the NSs' speech. *You know* in utterance-initial and -final occurs only once in the NNSs' monologues.

In the three sub-corpora of the dialogic genres, as can be seen in Table 7.34 above, more instances of *you know* are placed in turn-initial and -final positions in the NNSs' dialogues, as shown in Excerpts (7.4.2) and (7.4.3). The proportion of *you know* in turn-initial position in SECCL (30.9%) is much larger than those in MICASE (5.2%) and in ICE-GB (4.9%) and that in turn-final in SECCL (10.2%) is slightly higher than those in MICASE (3.3%) and in ICE-GB (9%). These phenomena are possibly because turn-taking in the NNS data is very frequent, while in the NS data, the length of a turn is usually much longer.

(7.4.2)

P: Turn-initial

Task 3

A: Hello, Shirley. Hello, I'm really happy to meet you here.

B: Yeah, me too.

A: **You know** these days... eh... I had been really a bad mood. Do you know why?

B: I don't know.

(SECCL: C00-58-06)

(7.4.3)

P: Turn-final

A: I think it not very important because it is just temporary, **you know**...

B: You mean it is not necessary for me to change my character to cater for others.

(SECCL: C00-82-02)

7.4.1.2 *I mean* and *you see* in extra-clausal position¹⁵

The typical position of *I mean* is extra-clausal utterance/turn medial position. This is possibly due to the fact that the major types of co-occurrence of *I mean*, discussed in the next section, are clarifications, explanations and elaborations. When *I mean* co-occurs with clarifications, explanations and elaborations, there should be a previous utterance by the speaker or other interlocutors. Therefore, *I mean* tends to occur in extra-clausal utterance/turn medial position.

You see is predominantly used in extra-clausal positions, as shown in Table 7.38. About 92% on average of the occurrences are in extra-clausal positions. There are more instances of *you see* in turn-initial position in the NNSs' dialogues (34.2%) than in the NSs' highly interactive discourse mode (30.8%) and the private direct conversations (13.4%).

¹⁵ *I mean* and *you see* in extra-clausal positions are not illustrated in this section to save space in this chapter.

7.4.1.3 *You know* in intra-clausal position

In the NNSs' speech, a small proportion of *you know* occurs in an intra-clausal position, with 13.6% and 8.1% in the monologues and dialogues respectively. In contrast, about half the instances (54.6% and 45.1% in the two types of genre in MICASE and 48.5% and 37.3% in ICE-GB) are placed in intra-clausal positions in the NSs' speech. In spite of this difference in frequency, in both NNSs' and NSs' speech, *you know* in an intra-clausal position mostly occurs after an incomplete message (M-) element, as shown in Excerpt (7.4.4).

(7.4.4)

P: M- + *you know*
+ +M

B: And one day we had turkey cutlets
They brought it out and the girl I was sitting next to
And they were like |**you know**| pre-pack frozen bread orange breadcrumbs
she cut into it pushed the bits apart and there's this something stretched from
one bit to another and it was I'd swear to God it was chewing gum
(ICE-GB: S2A-038)

7.4.1.4 *I mean* and *you see* in intra-clausal position¹⁶

I mean is not primarily used in an intra-clausal position. On average, 23% of the occurrences of *I mean* are placed in intra-clausal positions in the NSs' speech and a slightly higher proportion, 37% on average, in the NNSs' speech.

You see in an intra-clausal position accounts for less than 8% in both the NNSs' and NSs' speech, except in the sub-corpus of the unscripted monologues, which reveals 23.1%.

7.4.2 Contexts where Type B *you know* tends to occur

By adopting a linguistic descriptive approach, the collocation phenomena surrounding *you know* are first identified. The types of co-occurrence of *you know* in the speech of the NNSs and NSs under investigation are discussed in order of the strength of evidence, from the stronger linguistic evidence to intuitive interpretation. As mentioned earlier, although the identified types of co-occurrence are used as discussion categories, it is sometimes unavoidable to resort to intuition for interpreting the functions of DMs. Occasionally some problems occur in assigning categories. How these instances are classified is accounted for, showing the kinds of problems faced and the decisions taken. The instances in ambiguous contexts, with no linguistic evidence and insufficient contextual information, remain unclassified in my analysis. Tables 7.32 to 7.37 at the end of this section illustrate the

¹⁶ *I mean* and *you see* in intra-clausal positions are not exemplified in this section to save space in this chapter.

proportion of the types of co-occurrence of *you know* in relation to positions in an utterance/turn in the six sub-corpora.

You know is found to co-occur with 1) hesitation markers, pauses and restarts, 2) reported speech, 3) repairs, 4) emphatic lexis and key information, 5) exemplifications, 6) clarifications and explanations, 7) contrasting and negative points, 8) the opening of a topic or narrative, 9) concluding remarks and 10) shared knowledge presumed by the speaker.

7.4.2.1 *You know* co-occurring with hesitation markers, pauses and restarts

You know co-occurring with hesitation markers, pauses and restarts accounts for similar proportions in the speech of the Chinese NNSs (10.1% and 13% in monologues and dialogues respectively), the American NSs (15.1% and 15%) and the British NSs (9.1% and 13.5%). In Excerpt (7.4.5), it seems probable that *you know* co-occurring with the hesitation marker *eh* and pauses is being used to stall for time. This type of co-occurrence is often found. The patterns of *you know* indicate that *eh* is one of the frequent collocates immediately to the left and to the right. As with *like* (see Section 4.4.2.1), it is concluded that *you know* may also be used to indicate a search for content information or lexical words.

(7.4.5)

P: Utterance-medial E: Hesitation marker <i>eh</i> ; pauses F: To search for contents and lexical words	Eh... ... so every, every... eh... every one of us spent... eh... enjoyable... enjoyed<enjoy> the, the night... eh... in... eh... in the celebration. So... eh... you know ... eh ... that's a very for... unforgettable night.	(SECCL: B00-74-22)
---	--	--------------------

7.4.2.2 *You know* co-occurring with reported speech

You know is found to co-occur with reported speech, as in Excerpt (7.4.6) below. In the NNSs' speech, this type of co-occurrence is not frequent and it is found only in the monologues, seen in 12 (5.3%) out of 228 occurrences. However, *you know* preceding reported speech has been widely discussed in the studies of the NS speech (Erman 1987: 115, Schiffrin 1987: 282, Redeker 1991: 1163, He and Lindsey 1998: 143, Erman 2001: 1342). In the four sub-corpora of the NSs' speech under investigation, *you know* prefacing reported speech is one of the less frequent types of co-occurrence.

(7.4.6)

P: M- + OI + *you know* + M+

E: Quoting verb *said*

F: To mark reported speech

..... I began to lose my confidence <confident> but also but this time was <is> also this classmate came <come> to me and she encouraged me to she encouraged me to be confident. She said, “Eh... |you know| now you have many good result. I think you will also get a remarkable you will get a remarkable result in the final examination. Please be confident to yourself.”.....

(SECCL: B96-13-15)

7.4.2.3 *You know* co-occurring with repairs

You know is found to be followed by the correction of the term or statement prior to *you know*. In Excerpt (7.4.7) below, *you know* prefaces *a girl* to repair the previous pronoun, *him*. In this case, *you know* seems to suggest that a correction is formulating and it serves to mark a self-repair. This is one of the less frequent types of co-occurrence in the NNSs’ speech as well as in the NSs’ speech, accounting for less than 6%. Most of these instances are placed in intra-clausal positions.

In Müller’s study (2005) on German NNSs of English and American NSs, she found that only 3 out of 77 NNSs use *you know* once to mark a self-repair, whereas almost one third of NSs use it between one and three times for this purpose. Likewise, in the present study, Chinese NNSs of English seldom use *you know* to signal a repair. The use of *you know* in Excerpt (7.4.7) is unusual.

(7.4.7)

P: MF + *you know* + MR

E: Item (2) rephrases Item (1)

F: To signal a repair

A: Find her English short coming easily and correct it. It’s very good for English learning.

B: But in my opinion, she’s too young to be... independent.

⁽¹⁾ It’s hard for him, for |you know|... ⁽²⁾ it’s... difficult for a girl to live abroad when he is only a teenager and live abroad is very hard. Um, and, I don’t know he, whether he can beared the discriminations in, in other countries.

(SECCL: C01-50-11)

7.4.2.4 *You know* co-occurring with emphatic lexis and key information

The category of emphatic lexical words and important contextual information is one of the most frequent types of co-occurrence of *you know*. *You know* co-occurs with lexical indications of importance, such as *best*, *most*, *very*, *always*, *certainly* and *perfectly*, which serve as *intensifiers* (Martin and White 2005). In Excerpts (7.4.8) and (7.4.9), *you know* co-occurs with emphatic lexis, *best* and *perfectly* and this type of co-occurrence seems to suggest that *you know* is used to draw listeners’ attention to the statement with the emphasis surrounding *you know*.

(7.4.8)

P: Utterance-medial

E: Emphatic lexis *best*

F: To emphasise a statement

..... I think the five-year birthday party is my best memory and I can't forget and I hope I can... and I hope I can get a very birthday again like that. |**You know**| the **best** memory in a person's heart is very good and very scarce.... ..

(SECCL: B00-29-15)

(7.4.9)

P: M- + *you know* + +M

E: Emphatic lexis *perfectly*

F: To emphasise a statement

..... if you were against me at the beginning, that's just too b- bad. uh, with Augustus, he says |**you know**| perfectly well that you're very lucky that i've spared your life, but now that i have, uh why don't you come along and join my, join my party?

(MICASE: LEL215SU150)

In Excerpt (7.4.10), *you know* seems to emphasise the fact that Speaker B is from another province. Followed by speaker A's minimal response, *um*, Speaker B continues to say the environment is new to her; therefore, the fact that she is from another province is an important item of information in their dialogue.

(7.4.10)

P: Turn-medial

E: Item (2) suggests the possibility that Item (1) is an important information

F: To emphasise a statement

Task 3

A: Hi, Carol.

B: Hi, Lily. Nice to meet you.

A: Nice to meet you. Eh... So you are just enter our university. You are a freshman?

B: Yeah, yeah. As a matter of fact, I'm coming for your advice.

A: Oh, really? I'm glad...

B: I know you are sophomore. And you know I'm just enrolled in the university ander..... |**you know**| ⁽¹⁾ **I come from I come from other province.**

A: Um.

B: ⁽²⁾ **It is a wholly new environment for me.**

(SECCL: C00-82-23)

This type of co-occurrence is highly-represented in the NNSs' speech, with 57.9% and 17.9% in the monologues and dialogues respectively. Most of the instances occur in extra-clausal utterance/turn-medial position. Less than 5% occur in intra-clausal positions.

In the NSs' speech, emphatic lexis and key information are also among the most frequent types of co-occurrence, but the proportion is not as high as that in the NNSs' speech, with an average percentage of 18%. In addition, the instances are more often placed in intra-clausal positions.

7.4.2.5 *You know* co-occurring with exemplifications

You know is found to be followed by some examples of the previous statement. In Excerpt (7.4.11), *you know* prefaces the examples (Items (2) to (7)) of Item (1), *they turned to talk about me*.

(7.4.11)

P: Utterance-medial
E: Items (2) to (7)
exemplifies Item (1)
F: To introduce
exemplifications

..... At first they about some trivial things like study and our dorm living condition. But I was so surprised to find that ⁽¹⁾**they turned to talk about me**. [**You know**] ⁽²⁾**I am very skinny in shape** but ⁽³⁾**they exaggerated my face** and said I was, ⁽⁴⁾**I was ugly**, and they also said<say> ⁽⁵⁾**I had many bad habits**. ⁽⁶⁾**I used to stay up until midnight to go on my study**. But they said that it as noisy very much. And ⁽⁷⁾**I was not popular at all**. I felt very angry at that time.

(SECCL: B02-61-34)

In Excerpt (7.4.12), Item (2) is an example of Item (1). This type of co-occurrence suggests that *you know* may be used to introduce exemplifications. This use is less frequent with only 1.8% and 2% in the NNSs' monologues and dialogues respectively and with an average of 8% in the NSs' speech.

(7.4.12)

P: M- + *you know* +
MF
E: Item (2)
exemplifies Item
(1)
F: To introduce
exemplifications

A: Depending on the vessel and depending upon who we are it might take a number of different types of storm It might be ⁽¹⁾**something** <,> that would interest some and not others such as [**you know**] the the ⁽²⁾**the** <,> **invention of injection moulding**

(ICE-GB: S2A-040)

7.4.2.6 *You know* co-occurring with clarifications and explanations

You know co-occurring with clarifications and explanations is one of the frequent types of co-occurrence in the NSs' speech, but not in the NNSs' speech. There is no instance of it in the NNSs' monologues. However, in the NSs' speech, a high proportion of the occurrences of *you know* (31.1% and 19.7% in the two sub-corpora of MICASE and 16.7% and 15.2% in those two of ICE-GB) is found in the context of clarifying what has previously been said. Therefore, in the three monologic genres, there is no occurrence of *you know* in utterance-initial position, shown in Tables 7.32, 7.34 and 7.36. The instances of *I mean* in this category all have to be connected to a previous utterance.

In Excerpts (7.4.13) and (7.4.14), *you know* is followed by a clarification of what has been said. In the first example, the speaker uses *you know* to preface the explanation of

farm-raised shrimp. In the second example, Speaker A clarifies why geology is a special major and *you know* is placed before the clarification.

(7.4.13)

P: Utterance-medial
E: Item (2) clarifies
 Item (1)
F: To signal the
 follow-up
 clarification

..... currently, on the open market. what one would want this for is to tint your flowers. it's a nice little reddish orange color. also, ⁽¹⁾ **farm-raised shrimp**, **[you know]** ⁽²⁾ **all those shrimps you buy in the grocery store**. farm-raised salmon. nobody wants to buy you know if they're pale right? so they add, astaxanthin to these guys so they make 'em nice and orange.

(MICASE: LES405JG078)

(7.4.14)

P: Turn-medial
E: Item (2) clarifies
 Item (1)
F: To signal the
 follow-up
 clarification

B: What's your opinion?

A: I think... I think... that batman should stick to their original plan. Because... because. .er... the ⁽¹⁾ **the ge geology is a special major [you know]**, it ⁽²⁾ **it acquire... it inquires the student to have a strong body to suit, to be suitable for... for their fue further work**. Um a... I I don't think the female can... can be very suitable. And besides... um... females have scored higher you know emn... but high score don't stand... doesn't stand for emn the higher ability. I s... I think <ink> so.

(SECCL: C02-31-24)

7.4.2.7 *You know* co-occurring with contrasting and negative points

A number of instances of *you know* are found to precede contrasting and negative points and *you know* may be used to claim consensus and further to cushion the impact of the comment, as in Excerpt (7.4.15).

(7.4.15)

P: Turn-medial
E: *but*
F: To mitigate a
 contrasting point

B: It's would it's also very difficult if if <laugh> communication is <,> virtually nil

A: It is but that surely would improve with time <,>

B: Well I think uh yes I mean I think it will **But [you know]** that's uh that's not something I can <,> start in on straight away

(ICE-GB: S1A-031)

In Excerpt (7.4.16), the speaker uses *you know* with a negative self-evaluation and *you know* probably elicits sympathy and acts as a mitigator. As Crystal (1988: 47) argues, *you know* in sentence-initial is often used to “soften the force of what we are saying – a verbal equivalent to a gentle hand on the shoulder”. It seems reasonable to assume that the more interactive the context is, the more *you know* with this function occurs. This is true in the speech of the NNSs and NSs under investigation.

- (7.4.16)
- | | | |
|--|--|--------------------|
| P: Utterance-medial
E: Negative self-evaluation follows
F: To mitigate a negative point | But... but one... but one day,... um... um... there're... um... at a middle time test... middle-term test. [You know], I'm... I'm did not good... my maths is not good... was not good , so I asked him to tell, to give, to ge... to give his answer to me. | (SECCL: B96-05-14) |
|--|--|--------------------|

7.4.2.8 *You know* occurring at the opening of a topic or narrative

You know occurring at the opening of a topic or narrative is a particular type of co-occurrence in the NNSs' speech, representing 2.6% and 8.9% in the monologues and dialogues respectively. It only occurs three times (1.2%) in the sub-corpus of the private direct conversations in ICE-GB and no instance is found in the other three NS sub-corpora. *You know* in utterance/turn-initial position or in utterance/turn-medial but close to the beginning of the utterance is found to indicate the opening of a topic or narrative, as in Excerpts (7.4.17) and (7.4.18) below.

- (7.4.17)
- | | | |
|---|---|--------------------|
| P: Utterance-medial
E: Formula, <i>one day</i> , for opening a narrative
F: To indicate the beginning of a topic/narrative | Task 2
I already hate John, you know whoer..... often played kicker..... wi... on me.
One day, [you know] I was I had a revenge over him. I wanted to take a revenge. | (SECCL: B97-01-04) |
|---|---|--------------------|

- (7.4.18)
- | | | |
|--|--|---------------------|
| P: Utterance-initial
E: Nothing relevant precedes
F: To indicate the beginning of a topic/narrative | Task 3
A: Hi, Betty!
B: Hello, Mary!
A: You know the geology department's examination's result has come out.
What you um... what you think about the department? The decision. An
Does it should um... stick to its original plan?
B: I don't think so. | (SECCL: C02-100-27) |
|--|--|---------------------|

7.4.2.9 *You know* prefacing concluding remarks

Similar in frequency to the previous type of co-occurrence, *you know* prefacing concluding remarks is least often used. In Excerpt (7.4.19) below, the speaker was talking about how her friend, Tingting, helped her study while she was in hospital. In the end, the speaker used *you know* to mark the concluding statement.

(7.4.19)

P: Utterance-medial
E: Item (1) is the ending of the narrative and Item (2) is a concluding remark
F: To indicate a conclusion

Then, um... everyday she told she told she worked together with me for several hours told... find my my.... my mistakes in the study. ⁽¹⁾**So I get up with the classmates very soon I'm very grateful to tingting.** |**You know**], ⁽²⁾**that is the very important term I get spirit.**

(SECCL: B96-13-02)

You know prefacing concluding remarks occurs 5 times (2.2%) in the 228 instances in the NNSs' monologues and 4 times (1.9%) out of 213 instances in the sub-corpus of highly interactive discourse mode in MICASE; no instance is found in ICE-GB.

7.4.2.10 *You know* co-occurring with shared knowledge presumed by the speaker

You know is found to precede common sense and mutual knowledge shared by speakers. It is difficult to be precise, but it seems that *you know* is used to build consensus and mark solidarity. The NNSs under investigation are expected to have much common experience, mainly because they are fellow students. In Excerpt (7.4.20) below, the statement following *you know* is a common sense explanation of the anti-smoking practices in school. In Excerpt (7.4.21), given the fact that the two speakers are fellow students, it is assumed that the statement after *you know* is mutual knowledge.

(7.4.20)

P: Utterance-medial
E: Shared (unrestricted) knowledge follows
F: To build consensus

He is a person who fond of smoking very much... who is fond of smoking very much. And... |**you know**], **smoking will do great harm to our human bodies.** So our middle school authority just abandoned it.

(SECCL: B01-30-28)

(7.4.21)

P: Utterance-medial
E: Shared (restricted) knowledge follows
F: To build consensus

B: I mean it is a very... very good way and also very important way, and eh... I will... I will tell you another good way, that you can speak English to native lan... eh... native speaker. |**You know** | **there are many foreigners also foreign teachers in our... on our campus.**

A: So... you can... eh... seek them and speak English to them, and to... eh... communicate with them. It is... very good way, you... you can find many good expressions and so... pure... pure pronunciations.

(SECCL: C00-65-34)

In Excerpt (7.4.22), *you know* follows an unfinished utterance (an M- element) and this seems to invoke shared knowledge between Speakers A and B, suggesting that we share

knowledge about what someone can do at half past five in the morning (e.g. have a cup of coffee). This use of *you know* serving as a substitute is more common in the NSs' speech.

(7.4.22)

P: M- + *you know* +
M-

E: Unfinished,
suggesting shared
knowledge

B: If it's something like half past five in the morning when you could easily sit down and have |**you know**| By the time the luggage has come through and planes are delayed anyway you might as just might just as well say well come and meet us at eight or something and make it a bit decent for everyone

A: Well I don't want to go anyway

(ICE-GB: S1A-006)

In the above three examples, *you know* could be used to build consensus. This use represents 13.6% and 12.2% in the Chinese NNSs' monologues and dialogues and it occurs more frequently in the NSs' speech, with 10.9% and 23% in the American NSs' highly monologic and highly interactive discourse mode and 13.6% and 19.3% in the British NSs' unscripted monologues and private direct conversations.

7.4.2.11 Problematic and unclassified instances of *you know*

Three problems in classifying the instances are illustrated in this section. First, more than one type of co-occurrence is observed. Second, instances cannot be objectively classified due to the lack of linguistic evidence. Last, the number of instances is too low to be put into a category. (These problems also arise in the classification of *I mean* and *you see*.)

More than one type of evidence may be observed. In these cases, *you know* is grouped into the category with stronger evidence. In Excerpt (7.4.23) below, *you know* co-occurs with emphatic lexis, such as *really* and *very* and negative self-evaluation, *I'm not good at English*. It is problematic to tell which statement, the preceding or the following one, *you know* is attached to. However, since the speaker's friend is the topic, this instance is assigned to emphasise the previous statement.

(7.4.23)

P: Utterance-medial

E: Emphatic lexis and
negative
self-evaluation

F: Primarily
emphasise a
statement

Task 2

My friend, I'd better tell you the experience when I was in the middle school. I had a very good... friend. **Really**, he is a **very** good person. |**You know**|... **I'm not good at English**. When I was <am> in the, when I was in the middle school... ah, my good friend named Xiaoming gave me very great help.

(SECCL: B96-13-32)

In Excerpt (7.4.24), *you know* co-occurs with the vocative hesitation marker *um* and pauses. It can be used to indicate a search for content information or lexis. It also co-occurs with emphatic lexis *must*, suggesting the importance of *learning how to solve the differences*. The phrase *how to say* shows that the speaker is not sure about the contents or lexis; therefore, it is more likely that this instance of *you know* is used as a delaying device to indicate a search for contents or lexis.

(7.4.24)

P: MF + *you know* + M-

E: Vocative hesitation markers, pauses and emphatic lexis

F: Primarily indicate a search for contents or lexis

A: Um... um. In the university... um... we usually... um... there <they> are... there <they> are... um... usually many people, many people... you know, we are... we are from... different places and um... we must have... some differences between us, just like... um... our ideas <idea>, our... habits. So... I think you must... **um... um... um** you **must um... how to say... |you know| you must** um... learn to... how to **um...** solve the differences... between you um.

B: And, do you think that we should um... keep... o... our... difference and... just... accept other people's habits and ideas?

(SECCL: C00-58-02)

In Excerpt (7.4.25) *you know* is placed in turn-initial position but is used to emphasise the previous turn. Only a few instances of this are found.

(7.4.25)

P: Turn-initial

E: Emphatic lexis *should*

F: To emphasise the speaker's previous turn

B: Do you think so?

A: Yes. It's better. But I think we **should** study in the college in China first and then go abroad study.

B: No.

A: **You know?**

B: No, no. I think there is no... necessity to study in our country.

A: But if you start study abroad, there are no friends and you'll feel lonely.

(SECCL: C01-08-09)

The second problem with the categorisation of co-occurrence is some instances cannot be objectively classified, due to the shortage of linguistic evidence. For example, in Excerpt (7.4.26) the language surrounding *you know* cannot be used to suggest any type of co-occurrence and therefore this instance remains unclassified.

(7.4. 26)

P: M- + *you know* +
+M

E: No relevant
evidence identified

F: Unclassified

..... mostly what you see is population size remaining stable. on the islands where Darwin was, was working he saw, **[you know]** certain sizes of populations. when he left England and then came back he saw certain populations of deer and so on and butterflies. and they were still about the same. even though there's this, potential for huge population growth generally populations remain stable.....

(MICASE: LEL175JU154)

The last problem is that the number of instances is too small for them to be put into a category. *You know* is found to co-occur with vague language, such as *a bit of*, *maybe* and *probably*. This type of co-occurrence is identified in the NSs' speech only. There is one instance in the highly interactive discourse mode in MICASE and one and seven instances in the unscripted monologues and the private direct conversations in ICE-GB respectively.

All the instances of Type B *you know* in the Chinese NNSs' speech are classified. In the American NSs' highly monologic discourse mode, 5 (4.2%) out of 119 instances are found impossible to classify and in the highly interactive discourse mode 9 (4.2%) out of 213. 8 (12.1%) out of 66 instances in the British NSs' unscripted monologues and 19 (7.8%) out of 244 instances in the private direct conversations also remain unclassified.

7.4.2.12 Summary of the contexts where Type B *you know* tends to occur

The types of co-occurrence which Type B *you know* tends to occur with are discussed and Tables 7.32 to 7.37 below illustrate the distribution of the positions in an utterance/turn of Type B *you know*. In terms of the positions in an utterance/turn of *you know*, there is no significant difference, except for a larger proportion of *you know* occurring in an intra-clausal position in the NSs' speech. In terms of frequency in relation to co-occurrence, there are a few differences between the speech of the NNSs and NSs. Most of the instances in the NNSs' speech illustrate two types of co-occurrence: 1) emphatic lexis and key information and 2) contrasting and negative points, whereas in the NSs' speech, the most frequent types of co-occurrence are 1) clarifications and explanations and 2) shared knowledge presumed by the speaker.

The above four frequent types of co-occurrence suggest three discourse functions of *you know*. *You know* co-occurring with emphatic lexis and key information and prefacing clarifications and explanations seems to draw listeners' attention. *You know* co-occurring with contrasting and negative points seems to act as a mitigator, appealing for acceptance or sympathy. *You know* co-occurring with shared knowledge presumed by the speaker seems to

build consensus.

Two types of co-occurrence, more often in the NNSs' speech, occur either at the opening of a topic/narrative or prefacing concluding remarks. This is probably because the texts in the NNS data are rather short, offering more chances for *you know* to occur at the beginning of a topic/narrative and of the concluding remarks.

Table 7.32: Distribution of co-occurrence of *you know* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	10.1			18	7.9			4	1.8					1	0.4
2. Reported speech	5.3			6	2.6			3	1.3					3	1.3
3. Repairs	1.3			1	0.4									2	0.9
4. Emphatic lexis; key information	57.9			121	53.1			1	0.4	3	1.3	1	0.4	6	2.6
5. Exemplifications	1.8			4	1.8										
6. Clarifications; explanations	0														
7. Contrasting and negative points	5.3			9	3.9			1	0.4			1	0.4	1	0.4
8. Opening a topic/narrative	2.6	1	0.4	5	2.2										
9. Concluding remarks	2.2			3	1.3	1	0.4	1	0.4						
10. Shared knowledge presumed by speaker	13.6			28	12.3			1	0.4					2	0.9
Unclassified	0														
Occurrences: 228	100.0	1	0.4	195	85.5	1	0.4	11	4.8	3	1.3	2	0.9	15	6.6

Table 7.33: Distribution of co-occurrence of *you know* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	13.0	5	2.0	21	8.5			4	1.6	1	0.4	1	0.4		
2. Reported speech	0														
3. Repairs	3.7			1	0.4			1	0.4	1	0.4	2	0.8	4	1.6
4. Emphatic lexis; key information	17.9	7	2.8	32	13.0	3	1.2							2	0.8
5. Exemplifications	2.0			5	2.0										
6. Clarifications; explanations	13.4	11	4.5	17	6.9	4	1.6							1	0.4
7. Contrasting and negative points	28.9	33	13.4	27	11.0	9	3.7	2	0.8						
8. Opening a topic/narrative	8.9	17	6.9	5	2.0										
9. Concluding remarks	0														
10. Shared knowledge presumed by speaker	12.2	3	1.2	17	6.9	9	3.7							1	0.4
Unclassified	0														
Occurrences: 246 out of 300 (random samples)	100.0	76	30.9	125	50.8	25	10.2	7	2.8	2	0.8	3	1.2	8	3.3

Table 7.34: Distribution of co-occurrence of *you know* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance-initial	%	Utterance-medial	%	Utterance-final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	15.1			2	1.7			3	2.5	1	0.8	7	5.9	5	4.2
2. Reported speech	4.2							4	3.4					1	0.8
3. Repairs	1.7											2	1.7		
4. Emphatic lexis; key information	16.0			7	5.9			7	5.9					5	4.2
5. Exemplifications	9.2			5	4.2			5	4.2					1	0.8
6. Clarifications; explanations	31.1			25	21.0			5	4.2	1	0.8			6	5.0
7. Contrasting and negative points	7.6			6	5.0			2	1.7					1	0.8
8. Opening a topic/narrative	0														
9. Concluding remarks	0														
10. Shared knowledge presumed by speaker	10.9			9	7.6			1	0.8			2	1.7	1	0.8
Unclassified	4.2							4	3.4					1	0.8
Occurrences: 119	100.0			54	45.4			31	26.1	2	1.7	11	9.2	21	17.6

Table 7.35: Distribution of co-occurrence of *you know* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn-initial	%	Turn-medial	%	Turn-final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	15.0	1	0.5	8	3.8			8	3.8			13	6.1	2	0.9
2. Reported speech	1.9							3	1.4					1	0.5
3. Repairs	6.1			1	0.5			1	0.5	6	2.8	2	0.9	3	1.4
4. Emphatic lexis; key information	7.5			5	2.3	1	0.5	8	3.8					2	0.9
5. Exemplifications	8.9	2	0.9	9	4.2			4	1.9					4	1.9
6. Clarifications; explanations	19.7	1	0.5	30	14.1			6	2.8			1	0.5	4	1.9
7. Contrasting and negative points	11.7	2	0.9	18	8.5			3	1.4				0.0	2	0.9
8. Opening a topic/narrative	0														
9. Concluding remarks	1.9			3	1.4	1	0.5								
10. Shared knowledge presumed by speaker	23.0	4	1.9	23	10.8	5	2.3	13	6.1	2	0.9			2	0.9
Unclassified	4.2	1	0.5	1	0.5			4	1.9					3	1.4
Occurrences: 213 out of 300 (random samples)	100.0	11	5.2	98	46.0	7	3.3	50	23.5	8	3.8	16	7.5	23	10.8

Table 7.36: Distribution of co-occurrence of *you know* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	9.1										5	7.6	1	1.5	
2. Reported speech	3.0			1	1.5								1	1.5	
3. Repairs	3.0									1	1.5	1	1.5		
4. Emphatic lexis; key information	27.3			11	16.7			4	6.1					3	4.5
5. Exemplifications	12.1			1	1.5			3	4.5					4	6.1
6. Clarifications; explanations	16.7			5	7.6			2	3.0			1	1.5	3	4.5
7. Contrasting or negative points	3.0			1	1.5									1	1.5
8. Opening a topic/narrative	0														
9. Concluding remarks	0														
10. Shared knowledge presumed by speaker	13.6			7	10.6					1	1.5			1	1.5
Unclassified	12.1			8	12.1										
Occurrences: 66	100.0			34	51.5			9	13.6	2	3.0	7	10.6	14	21.2

Table 7.37: Distribution of co-occurrence of *you know* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	13.5	2	0.8	5	2.0			9	3.7	1	0.4	10	4.1	6	2.5
2. Reported speech	4.5							6	2.5					5	2.0
3. Repairs	0.8											1	0.4	1	0.4
4. Emphatic lexis; key information	22.1	1	0.4	39	16.0	7	2.9	6	2.5					1	0.4
5. Exemplifications	2.9			6	2.5			1	0.4						
6. Clarifications; explanations	15.2	1	0.4	25	10.2	1	0.4	3	1.2					7	2.9
7. Contrasting or negative points	12.7	3	1.2	17	7.0	3	1.2	4	1.6	1	0.4			3	1.2
8. Opening a topic/narrative	1.2	2	0.8	1	0.4										
9. Concluding remarks	0														
10. Shared knowledge presumed by speaker	19.3	3	1.2	16	6.6	8	3.3	17	7.0					3	1.2
Unclassified	7.8			10	4.1	3	1.2	3	1.2					3	1.2
Occurrences: 244 out of 300 (random samples)	100.0	12	4.9	119	48.8	22	9.0	49	20.1	2	0.8	11	4.5	29	11.9

7.4.3 Contexts where Type B *I mean* tends to occur

The descriptive approach employed in the analysis of *you know* is also used in the analysis of Type B *I mean*. The collocation phenomena surrounding *I mean* are identified in order to discuss the functions of *I mean*. Tables 7.38 to 7.43 below present the six types of co-occurrence of *I mean* in relation to the distribution of positions in an utterance/turn.

Type B *I mean* is found to co-occur with 1) hesitation markers, pauses and restarts, 2) repairs, 3) clarifications and explanations, 4) elaborations, 5) exemplifications and 6) contrasting and negative points. These types of co-occurrence, except for elaborations, are shared with *you know* and have previously been discussed and exemplified. Due to the limited space, the overlapping types of co-occurrence are not illustrated again in this section. In Excerpt (7.4.27) *I mean* prefaces Items (2) and (3), which elaborate Item (1). *I mean* co-occurring with elaborations can serve as a cue for the listeners of the coming details.

(7.4.27)

P: Utterance-medial
E: Items (2) and (3)
elaborate Item (1)
F: To elaborate what
has been said

A: But I think to go to that... to go abroad early as soon... as possible is a very good... for... us.
B: I don't see that.... Eh... ... because... eh... ... first I think... eh... ... now you are just a high school graduates and you... ⁽¹⁾**you are not mature enough to judge right or wrong in the... western cities, [I mean].** ⁽²⁾**Maybe you'll pick up some very bad habits, maybe doing drugs** or... maybe you'll... ⁽³⁾**you'll become a homosexual... guys.** I think it's terrible. <A: No.> Something maybe you should finish your college in China and then I think... eh... ... when you are mature enough and you can go abroad.

(SECCL:C01-01-20)

It can be seen in Tables 7.38 to 7.43 that there is, in general, no huge difference in the distribution of the six types of co-occurrence between the two groups of speakers and across the two types of genre. The major types of co-occurrence of *I mean* occur with clarifications and explanations, representing about a third of the instances in each of the six sub-corpora. *I mean* co-occurring with elaborations is also frequent, accounting for 28% on average in the NSs' speech and 17% on average in the NNSs' speech.

The distinctions between the two groups of speakers are drawn in the proportion of the two categories of co-occurrence: 1) hesitation markers, pauses and restarts and 2) repairs. The proportion of the category of hesitation markers, pauses and restarts is higher in the NNSs' dialogues (28.4%) and in the NSs' unscripted monologues (35.7%). The category of repairs accounts for 3.3% on average in the NSs' speech, whereas it accounts for 19.6% in the NNSs' monologues and 11.5% in their dialogues.

Table 7.38: Distribution of co-occurrence of *I mean* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	9.8			2	3.9					1	2.0			2	3.9
2. Repairs	19.6			3	5.9					4	7.8			3	5.9
3. Clarifications; explanations	56.9			18	35.3			1	2.0	3	5.9	2	3.9	5	9.8
4. Elaborations	11.8			5	9.8									1	2.0
5. Exemplifications	0														
6. Contrasting and negative points	2.0			1	2.0										
Unclassified	0														
Occurrences: 51	100.0			29	56.9			1	2.0	8	15.7	2	3.9	11	21.6

Table 7.39: Distribution of co-occurrence of *I mean* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	28.4	6	2.5	21	8.6	5	2.1	12	4.9	5	2.1	9	3.7	11	4.5
2. Repairs	11.5	2	0.8	4	1.6	5	2.1	7	2.9					10	4.1
3. Clarifications; explanations	35.8	31	12.8	34	14.0	2	0.8	3	1.2					17	7.0
4. Elaborations	22.2	13	5.3	39	16.0			1	0.4					1	0.4
5. Exemplifications	0														
6. Contrasting and negative points	1.6	1	0.4	2	0.8	1	0.4								
Unclassified	0.4	1	0.4												
Occurrences: 243	100.0	54	22.2	100	41.2	13	5.3	23	9.5	5	2.1	9	3.7	39	16.0

Table 7.40: Distribution of co-occurrence of / mean as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	17.5			3	5.3			2	3.5	2	3.5	3	5.3		
2. Repairs	3.5			1	1.8									1	1.8
3. Clarifications; explanations	38.6			20	35.1									2	3.5
4. Elaborations	28.1			16	28.1										
5. Exemplifications	8.8			5	8.8										
6. Contrasting and negative points	3.5			2	3.5										
Unclassified	0														
Occurrences: 57	100.0			47	82.5			2	3.5	2	3.5	3	5.3	3	5.3

Table 7.41: Distribution of co-occurrence of / mean as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers, pauses; restarts	18.3	5	1.9	11	4.2			7	2.7	7	2.7	15	5.7	3	1.1
2. Repairs	1.1			2	0.8									1	0.4
3. Clarifications; explanations	41.4	25	9.5	72	27.4	5	1.9	3	1.1			1	0.4	3	1.1
4. Elaborations	30.8	23	8.7	52	19.8			4	1.5					2	0.8
5. Exemplifications	3.0	3	1.1	3	1.1			2	0.8						
6. Contrasting and negative points	1.5			1	0.4			3	1.1						
Unclassified	3.8	4	1.5	2	0.8	2	0.8	1	0.4					1	0.4
Occurrences: 263 out of 300 (random samples)	100.0	60	22.8	143	54.4	7	2.7	20	7.6	7	2.7	16	6.1	10	3.8

Table 7.42: Distribution of co-occurrence of / mean as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%				
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%			after an MA	%	after an MF	%
1. Hesitation markers; pauses; restarts	35.7							1	7.1			2	14.3	2	14.3
2. Repairs	7.1			1	7.1										
3. Clarifications; explanations	28.6			4	28.6										
4. Elaborations	21.4			3	21.4										
5. Exemplifications	0														
6. Contrasting and negative points	7.1				0.0									1	7.1
Unclassified	0														
Occurrences: 14	100.0			8	57.1			1	7.1			2	14.3	3	21.4

Table 7.43: Distribution of co-occurrence of / mean as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%				
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%			after an MA	%	after an MF	%
1. Hesitation markers, pauses; restarts	14.2	3	1.1	14	5.0			3	1.1	1	0.4	17	6.0	2	0.7
2. Repairs	1.4					1	0.4					2	0.7	1	0.4
3. Clarifications; explanations	34.9	21	7.5	64	22.8	9	3.2	1	0.4					3	1.1
4. Elaborations	32.4	16	5.7	74	26.3									1	0.4
5. Exemplifications	4.6	2	0.7	9	3.2			2	0.7						
6. Contrasting and negative points	6.8	4	1.4	14	5.0									1	0.4
Unclassified	5.7	6	2.1	8	2.8	2	0.7								
Occurrences: 281 out of 300 (random samples)	100.0	52	18.5	183	65.1	12	4.3	6	2.1	1	0.4	19	6.8	8	2.8

7.4.4 Contexts where Type B *you see* tends to occur

The process involved in the analyses of *you know* and *I mean* is repeated with the instances of *you see*. Tables 7.44 to 7.49 below show the eight types of co-occurrence of *you see* in relation to the distribution of positions in an utterance/turn.

Type B *you see* is found to co-occur with 1) hesitation markers, pauses and restarts, 2) emphatic lexis, 3) exemplifications, 4) explanations, justifications and conclusions, 5) indications of objects and places, 6) beginning of a new topic or new information, 7) shared knowledge presumed by the speaker and 8) for checking comprehension. Six of these types of co-occurrence are shared with *you know* and have previously been discussed and exemplified. The major types of co-occurrence of *you see* are found in explanations, justifications and conclusions in the speech of the NNSs and NSs. Interestingly, *you see* seldom – only twice – co-occurs in the NSs’ speech with hesitation markers, pauses and restarts, but in the NNSs’ dialogues it accounts for 10% in the 403 instances.

A distinction in the use of *you see* between the monologic genres and dialogic genres can be clearly made, with the proportions of the last two categories: 1) shared knowledge presumed by the speaker and 2) for checking comprehension. There are no instances of these two categories in the three sub-corpora of the monologic genres, while they represent about 20% on average in the dialogic genres. This finding is not surprising, as it appears to support my hypothesis that DMs contribute interaction in the presence of other speakers.

You see is rather different from *you know* in that it co-occurs with indications of objects and places and it is used for checking comprehension. In Excerpt (7.4.28) *you see* prefaces Item (1) *on the left*, which is an indication of place. This use is interpreted as a device to move the hearer’s attention to objects and places. It occurs rather frequently in the NSs’ speech, but not in the NNSs’ speech, probably because the NNSs talk about more abstract topics and give fewer chances to refer to objects and places.

(7.4.28)

P: Utterance-medial	um the things in the foreground stay in the foreground and the things in the
E: Item (1) refers to a place	background stay in the background. between turning this small oil sketch like that, into the full scale painting you see ⁽¹⁾ on the left Manet worked very hard
F: To gain attention	to introduce pictorial paradox and in doing so to remind us, as Courbet would have said that a painting is a made thing.

(MICASE: LEL320JU143)

The other particular category is *you see* occurring in the contexts for checking

comprehension. The identification of this category is based not on explicit linguistic evidence but on my interpretation of contextual information. In Excerpt (7.4.29), *you see* is used as a question. Items (1) and (2) are responses provided by Speaker B and Item (3) is a further explanation offered by Speaker A. In this case, it is reasonable to assume that Speaker A uses *you see* to check Speaker B's comprehension in order to know what to say next.

(7.4.29)

P: Utterance-final
E: Items (1) and (2) as responses to *you see*; Item (3) an explanation
F: To check comprehension

A: Yes, eh... you see eh... eh... you can train you... yourself and improve your ability in English and so I decide to enter it.
 B: Ah, well, it's good. It's very good.
 A: Yeah, but I... I... think it's a a... there is some problems |**you see**?
 B: ⁽¹⁾ **Oh?**
 A: And...
 B: ⁽²⁾ **What problems?**
 A: Yeah, eh... you see ⁽³⁾ **many top students will enter test... the contest.**

(SECCL: C97-01-11)

In Excerpt (7.4.30), Item (1) is Speaker 3's response to Speaker 4's *you see* for checking comprehension.

(7.4.30)

P: Utterance-medial
E: Item (1) as a response to *you see*
F: To check comprehension

S1: uh huh and to me i take that as a feeling i was thinking you know what, i think that's a s- i use it as a signal sometimes. as uh, and i think i could, think of it as, something helping me become aware, that i'm not, it's giving me some knowledge informa- it's giving me information, my feeling, |**you see**? so <OVERLAP1> so in that sense </OVERLAP1>
 S3: <OVERLAP1> ⁽¹⁾ **you haven't answered** </OVERLAP1> **my question have you, or have you?**
 S1: no yeah i'm i'm saying that in that sense, i could think it's a tool of cognition. it's a tool it's a way of gaining knowledge, about reality, but uh, apparently not.

(MICASE: SGR999MX115)

The instances of *you see* in Excerpts (7.4.29) and (7.4.30) can also be interpreted as a short form of *do you see?* moving towards the use of Type A *you see*. However, such instances are classified as Type B because it is very rare for the speaker to expect an answer. Thus, these instances of *you see* are more like DMs.

One possible interpretation of the use of *you see* is that the information surrounding *you see* is construed as shared and *you see* is used to impose the shared nature of the information, even of new information, as a means of constructing solidarity between the speakers. *You see* can be interpreted as a way among equals of expressing solidarity or as a way by the dominant speaker to impose the underlying assumption that *we share it*. This interpretation of the use of

you see can be put on all instances of the identified types of co-occurrence of *you see*, except those in the category of hesitation markers, pauses and restarts.

Table 7.44: Distribution of co-occurrence of *you see* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%			after an MA	%	after an MF
1. Hesitation markers; pauses; restarts	3.4			1	3.4									
2. Emphatic lexis	27.6			8	27.6									
3. Exemplifications	0													
4. Explanations; justifications; conclusions	41.4			12	41.4									
5. Indications of objects and places	10.3			3	10.3									
6. Beginning of a new topic/information	0			2	6.9									
7. Shared knowledge presumed by the speaker	0			1	3.4									
8. For checking comprehension	0													
Unclassified	6.9			1	3.4								1	3.4
Occurrences: 29	100.0			28	96.6								1	3.4

Table 7.45: Distribution of co-occurrence of *you see* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%			after an MA	%	after an MF
1. Hesitation markers; pauses; restarts	10.4	9	2.2	31	7.7						1	0.2	1	0.2
2. Emphatic lexis	17.9	22	5.5	42	10.4	3	0.7	2	0.5		2	0.5	1	0.2
3. Exemplifications	6.7	9	2.2	18	4.5									
4. Explanations; justifications; conclusions	39.5	54	13.4	102	25.3	2	0.5	1	0.2					
5. Indications of objects and places	0.2			1	0.2									
6. Beginning of a new topic/information	8.7	22	5.5	12	3.0			1	0.2					
7. Shared knowledge presumed by the speaker	10.9	18	4.5	24	6.0	1	0.2	1	0.2					
8. For checking comprehension	3.7	1	0.2	2	0.5	12	3.0							
Unclassified	2.0	3	0.7	3	0.7	2	0.5							
Occurrences: 403	100.0	138	34.2	235	58.3	20	5.0	5	1.2		3	0.7	2	0.5

Table 7.46: Distribution of co-occurrence of *you see* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	0														
2. Emphatic lexis	11.8			2	11.8										
3. Exemplifications	17.6			3	17.6										
4. Explanations; justifications; conclusions	47.1			8	47.1										
5. Indications of objects and places	23.5			4	23.5										
6. Beginning of a new topic/information	0														
7. Shared knowledge presumed by the speaker	0														
8. For checking comprehension	0														
Unclassified	0														
Occurrences: 17	100.0			17	100.0										

Table 7.47: Distribution of co-occurrence of *you see* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	0														
2. Emphatic lexis	5.1	1	2.6	1	2.6										
3. Exemplifications	5.1							1	2.6					1	2.6
4. Explanations; justifications; conclusions	38.5	7	17.9	6	15.4	1	2.6	1	2.6						
5. Indications of objects and places	23.1	4	10.3	5	12.8										
6. Beginning of a new topic/information	0														
7. Shared knowledge presumed by the speaker	2.6			1	2.6										
8. For checking comprehension	23.1			6	15.4	3	7.7								
Unclassified	2.6					1	2.6								
Occurrences: 39	100.0	12	30.8	19	48.7	5	12.8	2	5.1					1	2.6

Table 7.48: Distribution of co-occurrence of *you see* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%			after an MA	%	after an MF
1. Hesitation markers; pauses; restarts	7.7							1	7.7					
2. Emphatic lexis	7.7			1	7.7									
3. Exemplifications	0													
4. Explanations; justifications; conclusions	30.8			4	30.8									
5. Indications of objects and places	53.8			5	38.5			1	7.7					1 7.7
6. Beginning of a new topic/information	0													
7. Shared knowledge presumed by the speaker	0													
8. For checking comprehension	0													
Unclassified	0													
Occurrences: 13	100.0			10	76.9			2	15.4					1 7.7

Table 7.49: Distribution of co-occurrence of *you see* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%			after an MA	%	after an MF
1. Hesitation markers; pauses; restarts	2.1			1	1.0							1	1.0	
2. Emphatic lexis	15.5	3	3.1	11	11.3	1	1.0							
3. Exemplifications	2.1	1	1.0	1	1.0									
4. Explanations; justifications; conclusions	35.1	6	6.2	23	23.7	1	1.0	1	1.0					3 3.1
5. Indications of objects and places	22.7	1	1.0	19	19.6	2	2.1							
6. Beginning of a new topic/information	1.0	1	1.0											
7. Shared knowledge presumed by the speaker	3.1			3	3.1									
8. For checking comprehension	15.5			9	9.3	5	5.2							1 1.0
Unclassified	3.1	1	1.0	2	2.1									
Occurrences: 97	100.0	13	13.4	69	71.1	9	9.3	1	1.0			1	1.0	4 4.1

7.5 Chapter summary and conclusions

The NNSs' speech under investigation is unusual in that it is collected in a test-taking setting, where the speakers may try to be non-stop fluent in English. Due to this generic constraint, it is hypothesised that Type B *you know*, *I mean* and *you see* would be under-represented in the NNSs' speech. This hypothesis is supported by previous research on, for example, *you know* in Fung and Carter (2007) and *you see* in Prodromou (2008). However, the tests of statistical significance reveal that *you know* and *you see* are similarly represented in the monologic genres of the NNSs and NSs and over-represented in the dialogic genre of NNSs; *I mean* conversely, is significantly under-represented in the NNSs' dialogues, as opposed to the two dialogic genres of the NSs.

The hypothesis that there are more instances of these three Type B phrases in the dialogic genres than in the monologic genres is supported by the normalised frequencies across the six sub-corpora shown in Tables 7.1, 7.2 and 7.4. In addition, the results of the statistical significance tests indicate that these three DMs are under-represented in the monologic genres. It can be argued that, because *you know*, *I mean* and *you see* contain first and second pronouns, it is more likely that they occur more often in the presence of other speakers in the dialogic genres than in the monologic genres.

The descriptive approach to the analyses of *you know*, *I mean* and *you see* reveals their positions in an utterance/turn. In the cases of *I mean* and *you see*, there is no marked difference between the speech of the NNSs and NSs. In the case of *you know*, most of the instances (86.4% in the monologues and 91.9% in the dialogues) in the NNSs' speech are placed in an extra-clausal position, whereas about half the instances in the NSs' speech occur in an intra-clausal position. The NNSs' awareness of *you know* in an intra-clausal position may be raised to facilitate the process of comprehension for their communication with NSs.

On the basis of the collocation phenomena surrounding *you know*, *I mean* and *you see*, the contexts where they occur and their uses are often objectively identified. It is sometimes necessary to offer intuitive interpretations about their functions. On the whole, the Chinese NNSs use *you know* and *I mean* much as the NSs do. In the case of *you see*, two points may be worth noting here for the later discussion of their implications: 1) *you see* is not primarily used as a DM in the NSs' speech and 2) *you see* co-occurring with hesitation markers, pauses and restarts seems to be a non-native use. These differences between the Chinese NNSs and NSs are further discussed in Chapter 12.

CHAPTER 8: ANALYSIS OF *I THINK*

8.1 Introduction

This chapter begins with my hypotheses about the use of *I think* and my research questions, followed by a review of the literature, including the grammar aspects of *I think* with emphasis on its grammatical ambiguity, which causes difficulty in the distinction between its non-discourse use (Type A) and discourse use (Type B).

The same analytical procedure is employed, as in the previous chapters. The analysis presents the frequency information first and patterns of *I think*, showing an overall sense of the use of *I think* in the six sub-corpora under investigation. The major analysis is the discourse aspects of Type B *I think*, looking at its position in utterances/turns and the collocation phenomena surrounding it. Based on the identified co-occurrence, the functions of Type B *I think* are suggested.

8.1.1 Hypotheses and research questions

In my analyses in the preceding chapters, some DMs (e.g. *like* and *well*) are under-represented in the Chinese NNSs' speech and some DMs are over-represented (e.g. *oh*, *you know* and *you see*) as compared with the NS data. In this chapter, I hypothesise that Type A *I think* is frequently used in the Chinese NNSs' speech, as they are asked to give opinions, whereas Type B *I think* occurs more often than Type A in the NSs' speech. As one of the main hypotheses set out in Chapter 1, I also hypothesise here that there are more instances of Type B *I think* in the dialogic genres than in the monologic genres. With the aim of a thorough description of the use of *I think*, my hypotheses are tested within the framework of the core research questions addressed in Section 1.1.2 of Chapter 1.

8.1.2 Ways of distinguishing *I think* between non-discourse use (Type A) and discourse use (Type B)

Like the grammatical role of *you know*, *I mean* and *you see*, that of *I think* is ambiguous. *That* is often omitted in conversation by both the NSs and NNSs under investigation. *I think* can also be a standalone unit. Biber *et al.* (1999: 982-983) suggest that the comment clause *I think* is frequently used in conversations and it is used as a stance marker in turn-final position.

When *I think* is in clause-initial position, Biber *et al.*'s three criteria (1999: 1076-1078)

(see Section 3.3.2) for determining “utterance launchers” as DMs may be used, as exemplified below. In Excerpt (8.1.1) below, *that* can be added after the phrase *I think* and what follows, *he was very clever*, is a statement of personal opinion. *I think* in this case is a typical use of Type A.

(8.1.1)

Now, I want to introduce a very... good English teacher... of mine in secondary school. **I think** he was very clever, because in his class we never told to answer any questions or do any... kind... of dictations. Only thing we to do, we are asked to do is to study by, by ourselves.

(SECCL: B01-99-29)

In Excerpt (8.1.2), *I think* is followed by a non-declarative clause and that cannot be added. *I think* can be moved to final position and function as an interactive DM.

(8.1.2)

B: Oh, don't worry so much, just take it easy um and um study is one... is... just is one hand and um you you can seethere.. there are many activities in the university.

A: Yeah.

B: And you can... you can take part in some activities to improve you um...

A: But it's another problem for me, you know um I don't know whether I can do that.

B: Oh, don't worry, um **I think**, just take part some suitable activities and you can do... you can do it, so you can gain a lot from it.

A: But it will<we>-it will waste eh... very much time.

(SECCL: C00-11-18)

In addition to the above three criteria, Sinclair and Mauranen's *Linear Unit Grammar (LUG)* analysis (2006) is also used to investigate Type A and Type B *I think*. More detailed discussion is provided in Section 8.2.2 below.

In my analyses of *like*, *well*, *you know*, *I mean* and *you see*, my tagging of Type A words/phrases and Type B is compared with that in the ICE-GB corpus, which is the only corpus (under investigation) that tags DM. However, this comparison cannot be made in the investigation of *I think*, because Type B *I think* is not categorised as a DM in ICE-GB.

8.2 *I think* in the literature

The phrase *I think* is most problematic because its syntactical role is ambiguous and both Type A *I think* and Type B *I think* co-occur with personal opinions and evaluation. As a result, the discussions of Type A and the use of *LUG* are detailed and extensive, in order to present what Type B *I think* entails.

8.2.1 Grammatical aspect: Syntactical structure

The grammatical status of *I think* can be approached from two different angles: hierarchy and linearity. From these two angles, there are three possible ways of interpreting an utterance containing *I think*. Take the utterance *I think (that) he is wrong* for example. Using the conventional hierarchical analysis with *Subject-Verb-Object (S-V-O)* structure assists the interpretation of Type A *I think*. Using the extreme view of linear interpretation, in which *I think* is analysed as an “utterance launcher” for a clause (Biber *et al.* 1999: 1076-1078), assists the interpretation of Type B. Analysing the instance as two clauses leaves the question of Types A or B unresolved. One of the ways of resolving this is to ask whether *that* is retained or not, because the retention of *that* determines the instance as Type A and where *that* is omitted, it leaves the question open and the need for other criteria.

The omission and retention of *that* are an important reference for determining whether *I think* is a main clause or not. In terms of syntactical structure, *I think* is usually followed by a *that*-clause as direct object. In informal spoken English, *that* is very often omitted (Biber *et al.* 1999: 680, Carter and McCarthy 2006: 512). *That* tends to be omitted when the subjects in the main clause and that *that*-clause are co-referential and the subject in the *that*-clause is often a personal pronoun (Biber *et al.* 1999: 681, Carter and McCarthy 2006: 512). For example, *I think I could do well because I often did everything well.* (SECCL: B99-08-33) In the NNSs’ speech under investigation, there are 162 instances of *I think I* in the monologues and 574 instances in the dialogues. There are only 6 instances of *I think that I* in the monologues and dialogues respectively. *It*, used as a pronoun or a dummy subject, is one of the most frequent collocates immediately to the right. Other pronouns, such as *he*, *you* and *we*, are also used frequently as immediate right collocates of *I think* (see Tables 8.4 and 8.5 below). It can be concluded that *that* following *I think* is often omitted in the NNSs’ speech.

That tend to be retained by NSs in academic prose and be omitted in conversation (Biber *et al.* 1999: 680). In the NNSs’ monologues, there are 59 instances of *that* as the collocate immediately to the right of *I think* and in the dialogues, there are 371 instances. All of the instances are re-examined. In the monologues, out of 59 instances, 22 are used as the subject of the clause and 10 are used as determiner. The remaining 27 instances are *that*-clauses, accounting for about 2.6% of the 1,019 occurrences of collocates immediately to the right. In the dialogues, out of 371 instances, 127 are used as subject and 10 are used as determiner. The remaining 234 instances are *that*-clauses, representing about 2.5% of the 9,465 occurrences of

the collocates immediately to the right. It can be concluded that *that* is rarely retained in the NNSs' speech.

In the linear analysis, *I think that he is wrong* can be the structure of *utterance launcher* + *linker* + *S-V-C*. In the *LUG* analysis (Sinclair and Mauranen 2006), *I think* in clause-initial position can be categorised as an incomplete message (M-) element, followed by *that*, an text-oriented organisational (OT) element and *he is wrong*, a completion of message (+M) element. Without *that*, *I think he is wrong* can be analysed as *I think*, an interactive organisational (OI) element + a message-oriented (M) element. (More detailed analysis of *I think* with *LUG* is given in the next section.)

Where *I think* occurs in clause-medial and -final positions, it is always interpreted as Type B. However, there are comparatively low figures for these. *I think* can also be seen as a marker outside the structure of the preceding clause. For example, in Excerpt (8.2.1), *I think* is a comment tag, which is not part of the previous clause.

(8.2.1)

I will never <nevers> go to that restaurant again, "I said to myself. That's the... the terrible... the most terrible thing in my life and every time I think of it, I feel very uncomfortable..... Next time, I... I go to a restaurant, I will first have the meal, and then pay for the bill. **That's better, I think.**

(SECCL: B02-31-04)

8.2.2 *Linear Unit Grammar analysis of I think*

The phrase *I think* in the *LUG* analysis (Sinclair and Mauranen 2006) can be either an M-element or an OI element. As an M- element, *I think* is the use of Type A, carrying literal meanings. As an OI element, it is a Type B use, mainly contributing to the aspects of the interaction, such as initiating, maintaining and structuring the interaction and controlling the timings. Relevant discussion is made in Section 7.2.2 in the previous chapter.

The following excerpts are taken from the NNSs' speech and analysed with *LUG* (see Appendix 4 for a list of the labels in the *LUG* analysis). When the phrase *I think* is followed by *that*-clause, it is very likely that the instance of *I think* is an M- element, expressing a personal opinion, as in Excerpt (8.2.2) and/or an evaluative comment, as in Excerpt (8.2.3).

(8.2.2)

A: Oh, I... maybe I can get two jobs. One is working in the government, and the other is in the joint venture. I really do not know to decide which one to choose.

OI MF M M M- +M OT M- +M
M- +M- +M

B: Which one do you prefer most?

M

A: Oh, I don't know because I think that both of them have advantages and disadvantages.

OI M OT M- OT +M OT M

I want want to know your advice.

M- MF +M- +M

(SECCL: C99-21-33)

(8.2.3)

On that night he have the book with the beautiful paper and she willing to and she would like to give it to me as a present on my birthday. I was very excited. I think that Mary is a good girl and I remember all those things we just had come to with each other and at the hard time she come to with me.

MS M- +M OT M- OT +M-
+M- +M MS MS M M- OT +M
OT M- +M- +M- +M OT MS
M- +M

(SECCL: B00-82-30)

I think in Excerpt (8.2.4) is also assigned as an M- element according to the context, in which Speakers A and B are exchanging their opinions. The classification of this instance is not as objective as the previous two instances, because an intuitive judgement based on the contextual information (i.e. the discussion activity, in this case) has to be made.

(8.2.4)

A: I think the job working in the government is more colorful, because I will be... eh... able

M- +M- +M OT M- OI +M

to meet all kinds of people. But still, really, I can't decide... um...

OT OI OI M- OI

B: But I think in the joint <judge> venture, you can also meets a lot of new people,

OI M- MS +M- +M

interesting people, including some foreigners. I think there are some chances to go abroad.

M MS OI M- +M

I think that's good since you are English major.

OI M OT M

(SECCL: C99-35-07)

The first instance of *I think* in Excerpt (8.2.5) is assigned as an OI element, for what follows is a fact about the city of Shanghai, not a personal opinion. This instance can be left out without affecting the proposition in discourse. The second and third instances of *I think* are M- elements to express the speaker's opinions.

(8.2.5)

The first factor the first factor is that you can have a very good view of the city by means of
MF M- OT +M- +M MS OT
showing foreign guests around the city. ⁽¹⁾ **I think** shanghai is in the change of modern science
M- +M OI M MS
and modern technology and there used to be a lot of beautiful cities in Shanghai. And ⁽²⁾ **I think**
OT MS OT M- +M MS OI M-
it was really a good opportunity for me to tour around Shanghai city to see different
+M- +M- MS +M +M
landscapes and to see different people and have a talk with them and ⁽³⁾ **I think** this chance is
OT +M OT M- +M OI M- +M
great and quite unique for me.
OT +M

(SECCL: B98-21-30)

Two instances in Excerpt (8.2.6) are typical OI elements, which are used as fillers, serving the function of topical continuation or floor-holding; in particular, the NNSs in the present study were speaking under the pressure of an oral exam.

(8.2.6)

And ⁽¹⁾ **I think** what name shall I have. Shall I have true name? No, it's out of interest, interests.
OI OI M M M MA +M
It is..., it is not very interesting. So I thought and thought. So ⁽²⁾ **I think** in Chinese sweet means
MF M OT +M OI M OT M OI OI M M
"tian," and heart means "xin." So I think maybe I would call my call myself little sweetheart.
OT M OI OI M MA +M- +M
Oh, that means "xiao tian xin." It is very interesting.
OI M- +M M

(SECCL: B99-21-02)

The above analyses demonstrate how *LUG* is used to distinguish between Type A *I think* (M- element) and Type B (OI element). Although applying the *LUG* analysis does not make the distinction easier in ambiguous instances (e.g. *I think* in Excerpt (8.2.5) above), it offers a new model for looking at Types A and B. (See Section 12.2.6 of Chapter 12 for the usefulness and limitations of *LUG*.) Both the *LUG* analysis and Biber *et al.*'s three criteria (1999) are used in the present study for classifying the instances of *I think* into either Type A or Type B.

8.2.3 Previous studies of *I think*

The phrase *I think* is frequently used in spoken English (Stenström 1994: 59, O'Keeffe, McCarthy and Carter 2007: 65). It is one of the lexical items which Stenström (1994: 207) termed "interactional signals and discourse markers". This suggests that *I think* is typically used as a DM rather than in a reporting clause. In such cases, *I think* is categorised as an OI element in the *LUG* analysis (Sinclair and Mauranen 2006). As OI elements are characteristically absent or rare in written English, this may serve to explain why *I think* as a

DM occurs more often in spoken English.

Nevertheless, some researchers (e.g. Biber *et al.* (1999), O’Keeffe *et al.* (2007)) appear to infer that *I think* is commonly used with a *that*-clause (with or without *that*) in conversation. In addition, *think* is listed in first place of the common verbs followed by a complement *that*-clause (Biber *et al.* 1999: 662, 667-668). If *I think* is more often followed by a *that*-clause, it suggests that *I think* is frequently used as Type A.

It has been acknowledged by Sinclair and Mauranen (2006: 78) that it is difficult to make a distinction between *I think* as an M element (Type A) and that as an OI element (Type B). The ambiguous use of *I think* is discussed in Section 8.2.1 above. The difficulty in making a distinction probably accounts for the very few systematic analyses of *I think* as a DM. However, researchers in the previous studies have investigated the phrase *I think* without drawing a distinction between *I think* in a reporting clause and that as a DM.

The two studies by Simon-Vandenberg (2000) and Fortanet (2004) examine the phrase *I think* without distinguishing non-discourse and discourse uses. Simon-Vandenberg (2000) looks at *I think* in political interviews and casual conversations with respect to the syntax, intonation and collocation of *I think* and the proposition conveyed and interaction in discourse. She points out that frequency comparison across the two genres reveals only part of the way in which *I think* is used. In addition, the frequency information is different from previous studies, such as Aijmer’s data (1997) from the London-Lund Corpus. She concludes that *I think* is used with different functions in the two genres for analysis. In casual conversations, *I think* is primarily used as a hedge to express doubt, whereas in political interviews, *I think* is predominantly used as an expression of opinion to show feelings of certainty and authority rather than uncertainty and hesitation.

In the present study, the use of *I think* as an epistemic stance marker is categorised as Type A. This use is usually frequent when speakers are engaged in the expression of opinions and therefore more instances of Type A *I think* in the NNSs’ dialogues under investigation can be expected, because the speakers are asked to express their opinions on a given topic (see Appendix 1 for the topics for discussion in the NNSs’ speech).

A potential problem in Simon-Vandenberg’s (2000) study may be the use of one set of 100-line concordance samples each extracted from the Survey of English Usage (Svartvik and Quirk 1980) and her own corpus of radio political interviews. It is likely that this set of data provides a misleading distribution of functions. The present study uses three sets of 100-line

concordance samples and the distribution of possible functions shown in each of the three sets of random samples sometimes varies considerably (for more detailed discussion, see Chapter 3). Because of this, the proportions, on the basis of one set of 100-line concordance samples, of different functions may not show a representative distribution of the speech data.

The other study by Fortanet (2004) uses a corpus of five texts of lectures and five of discussion sections in order to compare the use of *I think* in the monologic and interactive discourse mode in MICASE. Six functions of *I think* in spoken academic English are identified: 1) to express opinion, 2) vagueness, 3) uncertainty, 4) to show politeness, 5) as an approximator and 6) as a hesitation marker. Fortanet argues that the type of activity is a key factor in the use of *I think* and finds variation across disciplines, which could be attributed to disciplinary differences and the common phenomenon of linguistic imitation among young speakers.

As in Simon-Vandenberg's study (2000), Fortanet's data (2004) show that opinion-based *I think* is most common. Fortanet (2004: 78) claims that in some cases of expressing opinion, *I think* seems to be associated with secondary functions, such as evaluation, vagueness and politeness.

Fortanet categorises *I think* in the intra-clausal position (e.g. *only if you have time he had he had I think uh two caravels [DS3]* (2004: 73)) as non-syntactic-integrated *I think*, which is referred as Type B *I think* in this thesis. This demonstrates the difficulty in describing the positions of DMs with conventional syntactical labels. In the *LUG* analysis, *I think* in this case can be described as occurring between an M- element and an OI element.

The two studies above have investigated Type A *I think* and Type B *I think* and reported that the epistemic stance use of *I think* is most common. Both studies indicate the difficulty of distinguishing the two uses of *I think*. In the present study, the *LUG* analysis is adopted to classify *I think* and the analysis focuses on Type B *I think*, which has the characteristics of DMs: 1) semantic and syntactical optional, 2) flexible in positions, 3) frequently prosodically independent, 4) offering connectivity in discourse and 5) multi-grammatical.

8.2.4 Previous studies of *I think* in the speech of Chinese NNSs

In addition to the small amount of research dedicated to the phrase *I think*, some studies discuss *I think* as one of the frequently-used chunks. Yang and Wei (2005: 40, 99) report that the frequency comparison of DMs across Chinese NNSs' and NSs' speech is statistically

significant. In the COLSEC corpus, *think* is placed in 9th position among the most frequent words, whereas in the spoken components of BNC and ICE it comes 35th and 33rd respectively. The researchers claim that in most cases, Chinese NNSs use *I think* as a “conversational filler” (Yang and Wei 2005: 40).

Xu and Xu (2007) investigate discourse management chunks in the COLSEC and ICE-GB and report that Chinese NNSs are unable to produce interpersonal chunks as varied as those of NSs. While British NSs prefer to use indirect language, Chinese NNSs tend to literally translate chunks in Chinese and use first-person perspective language, such as *I think*, *in my opinion*, *I want to say*, *it's my turn* and *I don't agree*. The researchers suggest that Chinese learners' use of “I-perspective” language would give listeners a bad impression of self-centredness (Xu and Xu 2007: 440).

These two studies seem to overlook the comparability of the corpora under investigation. Some of the Chinese NNS data are collected in the context where the speakers are asked to give opinions. These are similar to the data used in this thesis. When an interlocutor is given a question including the phrase *do you think*, s/he is likely to respond with *I think*, which is a typical use of Type A *I think* (see Table 8.3 in Section 8.3.1 for the frequencies of *do you think* across corpora). In the contexts where *do you think* is frequently used or where speakers are required to give personal opinions, it is not surprising to find a high incidence of *I think*. The frequency information in corpus data should be treated with due caution.

8.3 Frequency information in the speech of the non-native speakers and native speakers

8.3.1 Overall frequency of *I think*

The overall frequency of *I think* is shown in Table 8.1 below. There are 1,019 and 9,465 occurrences of *I think* in the NNSs' monologues and dialogues respectively. The raw frequencies are normed on a basis of 10,000 words. The normalised frequencies in Table 8.1 show that the sub-corpus of the NNSs' dialogues has a frequency more than five times greater than that in the monologues (159 vs. 30). Compared with the raw counts in the NSs' speech (shown in Figure 8.1), it seems that the NNS use the phrase *I think* much more frequently than the NSs. However, the instances include both non-discourse (Type A) and discourse (Type B) uses. The frequency information has to be viewed with a certain amount of caution, because

the types of activity across the six sub-corpora under investigation vary considerably.

The instances of *I think* are manually grouped into Types A and B (see Section 8.1.2 above for the ways of distinguishing *I think* between Types A and B). This classification reveals that *I think* is not primarily used as a DM. The proportions of Type B range from 7% to 25.3%. The raw frequencies of Type B are normed on a basis of 10,000 words and the normalised frequencies, ranging from 2.8 to 11.1 times, are shown in Table 8.1 and Figure 8.2. It is clear that there are many more instances in the NNSs' dialogues than in the NSs' speech. In terms of genre, there are apparently more instances of *I think* in the dialogic genres than in the monologic genres. This supports the hypothesis that the more interactive the genre and type of activity are, the more DMs occur.

Table 8.1: Frequency information of *I think* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percentage (%)	Normalised freq. of Type B per 10,000 words (times)**
SECCL: 1,143 monologues (Chinese NNSs)	336,303	1,019	30	30 out of 300 ^a	10.0	3.0
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	9,465	159	21 out of 300 ^b	7.0	11.1
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	89	7	20	22.5	1.5
MICASE: 48 transcripts of highly interactive (American NSs)	577,996	1,841	32	35 out of 300 ^c	11.7	3.7
ICE-GB: 70 unscripted monologues (British NSs)	153,646	170	11	43	25.3	2.8
ICE-GB: 90 private direct conversations (British NSs)	185,000	662	36	57 out of 300 ^d	19.0	6.8

* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B per 10,000 words are based on an extrapolation of the percentage of the Type B phrase.

a, b, c and d in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

For a frequency comparison of Type B *I think* between the two types of genre and between the speech of the NNSs and NSs, a log-likelihood (LL) test was made and the results of which are presented in Appendix 6. The LL scores are -200.12 between the monologues (Corpus A1) and dialogues (Corpus A2) in SECCL, -19.7 between the highly monologic discourse mode (Corpus B1) and the highly interactive one (Corpus B2) in MICASE and -28.63 between the unscripted monologues (Corpus C1) and private direct conversations

(Corpus C2) in ICE-GB. These LL scores indicate that the differences are highly statistically significant and that Type B *I think* is under-represented in the monologic genres. Between the two groups of speakers, there is a statistically significant difference in the dialogic genres between Corpora A2 and B2 (LL: +225.72, p-value: < 0.0001) and between Corpora A2 and C2 (LL: +28.27, p-value: < 0.0001), but the significance is not found in the monologic genres (LL: +9.8 between Corpora A1 and B1; LL: +0.2 between Corpora A1 and C1).

Figure 8.1: Normalised frequencies of *I think* across corpora

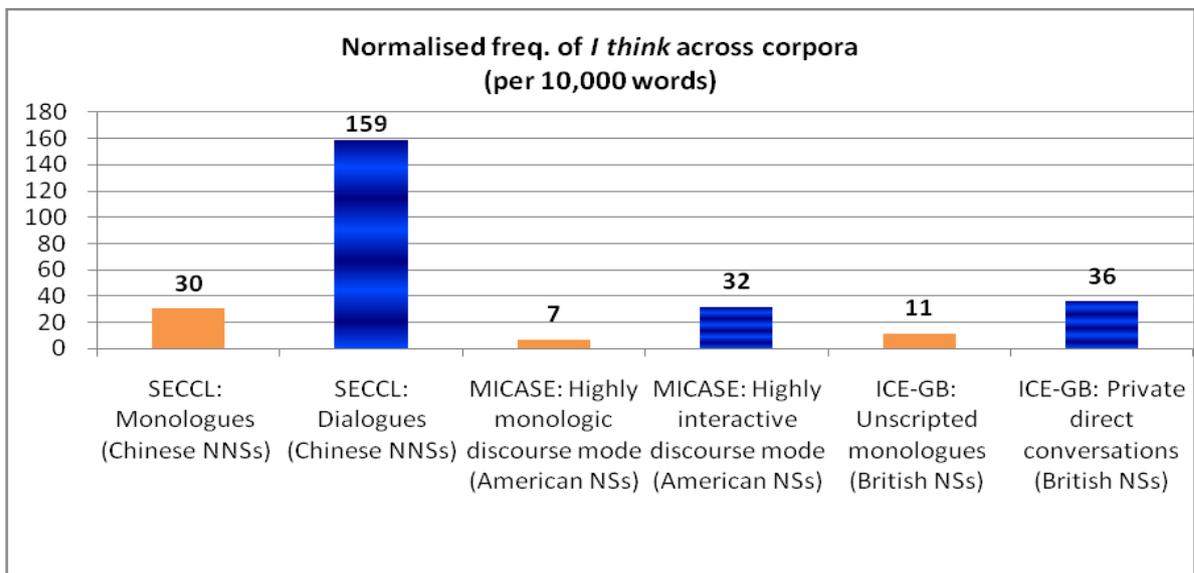
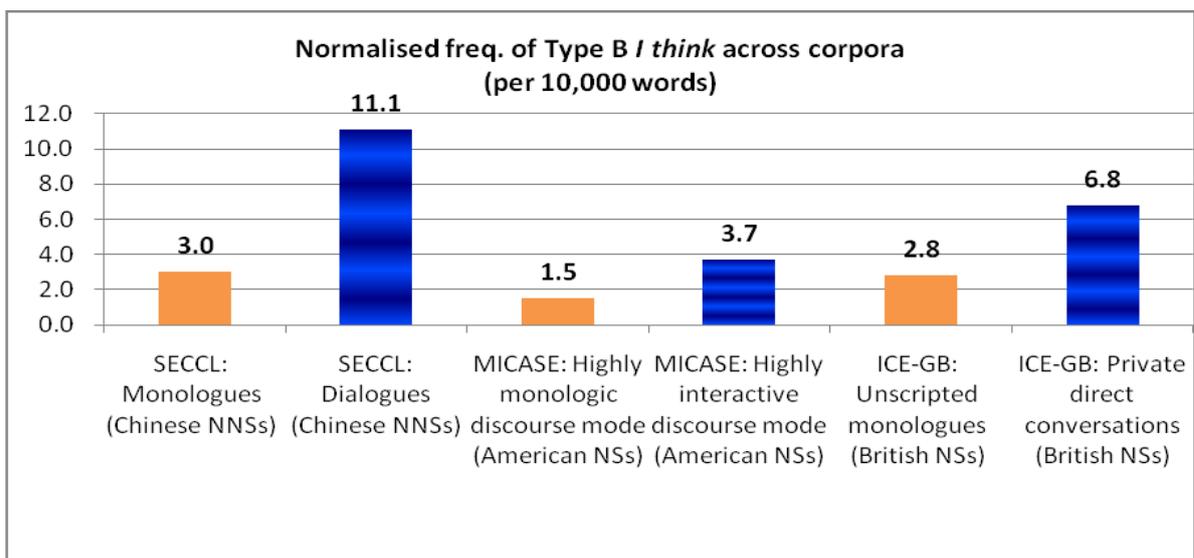


Figure 8.2: Normalised frequencies of Type B *I think* across corpora



Figures 8.1 and 8.2 show the contrasting levels of frequency of all instances of *I think* and Type B *I think*, in particular in the NNSs' dialogues, where a high percentage of the instances of *I think* is Type A (93%). This frequent use of Type A *I think* can be attributed to the type of activity in the NNSs' dialogues. The speakers are constantly exchanging opinions on the given topics and this leads to a high frequency of Type A *I think*. (For the topics discussed in the NNSs' speech, see Appendix 1.) To support this interpretation, a quick search was made for *do you think* in the six sub-corpora. Table 8.2 below shows that *do you think* is much more frequently used in the NNSs' dialogues. It is likely that the interlocutors respond to *do you think* with *I think*, thereby eliciting more instances of Type A *I think*.

Table 8.2: Frequencies of *do you think* across corpora

Corpus	Word counts (tokens)	Raw freq. (times)	Normalised freq. per 10,000 words (times)
SECCL: 1,143 monologues (Chinese NNSs)	336,303	15	0.4
SECCL: 1,143 dialogues (Chinese NNSs)	596,639	1,169	19.6
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	1	0.1
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	132	2.3
ICE-GB: 70 unscripted monologues (British NSs)	153,646	0	0.0
ICE-GB: 90 private direct conversations (British NSs)	185,000	54	2.9

The frequency information has revealed the overall use of *I think*. It is reasonable to conclude that *I think* is a frequently-used two-word chunk, but it is not primarily used as a DM by the NNSs and NSs under investigation. *I think* as a DM occurs more frequently in the dialogic genres than in the monologic genres under investigation. This resembles the use of other DMs (*oh, like, well, you know, I mean and you see*) discussed in the preceding chapters. In the next section, collocates of *I think* are used to further examine the use of Type A *I think* and Type B *I think*.

8.3.2 Collocates of *I think*

The patterns of *I think* in the six subsets of SECCL, MICASE and ICE-GB under investigation are presented in Tables 8.3 to 8.8. The two patterns of the NNSs' monologues and dialogues (see Tables 8.3 and 8.4) reveal similar collocates. The collocates to the left, *so*,

because and *but* and those to the right, *can* and *should*, shown in boldface, seem to indicate the use of Type A *I think* for expressing opinions. This corresponds to my manual classification of the random samples of *I think*. Most of the instances of *I think* in the NNSs' speech are Type A.

Table 8.3: Pattern of *I think* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (13)	very (16)	of (8)	and (32)	i think (300)	i (50)	is (34)	a (21)	i (21)
2	to (12)	to (13)	my (8)	so (27)		it (43)	i (21)	very (16)	very (16)
3	and (10)	i (11)	i (8)	because (13)		he (18)	was (21)	i (15)	a (16)
4	is (9)	the (10)	very (8)	but (13)		the (14)	s (15)	is (13)	is (14)
5	was (8)	my (9)	for (7)	eh (10)		that (13)	the (13)	can (7)	the (7)
6	his (8)	and (9)	and (7)	um (9)		maybe (12)	can (8)	my (7)	should (6)
7	the (7)	a (9)	the (7)	time (8)		eh (12)	should (7)	the (7)	to (6)
8	a (6)	so (8)	me (6)	me (8)		this (9)	will (7)	be (6)	for (5)
9	know (5)	me (7)	it (6)	then (6)		you (9)	this (6)	think (5)	all (5)
10	of (5)	in (7)	good (5)	2 (6)		um (8)	my (5)	unusual (5)	my (5)

Table 8.4: Pattern of *I think* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	a (16)	a (18)	a (34)	but (48)	i think (299)	the (35)	you (23)	a (15)	is (13)
2	the (13)	b (17)	b (28)	and (27)		eh (33)	is (16)	is (14)	a (11)
3	b (11)	the (14)	eh (14)	eh (20)		it (30)	s (15)	the (10)	to (10)
4	think (11)	i (11)	think (9)	b (17)		you (20)	i (14)	should (10)	the (8)
5	is (9)	yes (8)	um (9)	so (14)		i (18)	the (13)	you (9)	it (8)
6	i (8)	so (7)	so (9)	a (11)		um (16)	eh (13)	very (8)	in (7)
7	of (7)	very (6)	and (9)	i (11)		that (16)	should (10)	go (7)	eh (7)
8	so (7)	in (6)	yes (8)	yeah (11)		if (15)	can (10)	eh (7)	you (6)
9	you (7)	think (6)	you (7)	yes (10)		we (10)	a (9)	um (6)	abroad (6)
10	it (6)	of (5)	me (4)	because (9)		a (8)	will (7)	can (6)	um (6)

In Table 8.4, the collocates to the left, *a* and *b*, are mostly the identification of the two speakers in the dialogues. They indicate that *I think* is used in turn-initial position. This also relates to the relatively high frequency in this sub-corpus of *do you think*, shown in Table 8.2 above. It is possible that responses with *I think* are followed by questions containing *do you think*. Such collocates as *but*, *and*, *so* and *yeah* could be DMs co-occurring with *I think*.

In the NSs' speech in MICASE and ICE-GB, the patterns of *I think*, shown in Tables 8.5 to 8.8, also reveal the similar collocates to the left, *but*, *because* (*cuz*) and *so*, and, to the right, *should*. Another type of collocate immediately to the left, which is absent from the NNSs' patterns, is relative pronoun, such as *that*, *what*, *which* and *who* (shown in both boldface and italics), which might combine with *I think* to form relative clauses. This is one of the uses of

Type A *I think*. In addition, the collocate immediately to the right, *so*, in Tables 8.7 and 8.8 also suggest that *I think* is Type A.

Table 8.5: Pattern of *I think* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (8)	and (6)	of (4)	and (18)	i think (89)	it's (7)	are (4)	a (6)	of (4)
2	and (7)	the (4)	the (3)	that (10)		we (7)	is (4)	and (4)	a (4)
3	in (4)	other (3)	um (3)	um (5)		that (7)	uh (4)	to (3)	to (3)
4	of (3)	this (3)	than (2)	i (3)		it (4)	it (3)	of (3)	about (3)
5	a (2)	but (2)	a (2)	but (3)		is (4)	a (3)	was (2)	really (2)
6	to (2)	to (2)	and (2)	because (3)		you (3)	we (2)	it (2)	was (2)
7	um (2)	you (2)	office (2)	while (2)		um (3)	this (2)	that (2)	not (2)
8	that (2)	he (2)		what (2)		was (3)	wrong (2)	up (2)	in (2)
9	if (2)	of (2)		problem (2)		of (3)	would (2)	think (2)	and (2)
10	i (2)	i (2)		so (2)		i (3)	will (2)	do (2)	more (2)

Table 8.6: Pattern of *I think* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	is (7)	the (10)	uh (5)	and (30)	i think (170)	it (21)	s (21)	is (6)	to (9)
2	it (7)	of (7)	uhm (5)	but (13)		that (21)	the (12)	uh (6)	that (9)
3	the (7)	in (7)	s (5)	which (7)		the (10)	is (8)	a (6)	the (8)
4	and (7)	a (6)	is (4)	that (7)		in (9)	a (7)	the (5)	of (7)
5	in (6)	to (6)	that (4)	uh (6)		i (8)	of (6)	in (4)	a (5)
6	a (6)	it (5)	this (4)	uhm (4)		is (7)	may (5)	an (4)	in (5)
7	uh (6)	this (5)	and (4)	s (3)		there (7)	can (5)	to (4)	s (4)
8	s (4)	and (5)	to (4)	there (2)		this (6)	was (5)	s (4)	uh (3)
9	on (3)	s (4)	something (3)	work (2)		he (6)	it (5)	it (3)	it (3)
10	are (3)	that (4)	at (3)	who (2)		if (5)	i (4)	be (3)	and (3)

Table 8.7: Pattern of *I think* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (15)	i (11)	i (19)	so (18)	i think (255)	that (29)	the (15)	the (11)	of (11)
2	of (9)	the (11)	um (7)	but (14)		i (29)	was (11)	i (10)	i (11)
3	to (6)	think (7)	you (6)	and (14)		that's (23)	is (10)	to (8)	it (9)
4	the (6)	um (6)	and (6)	i (8)		you (17)	a (9)	of (7)	the (8)
5	it (3)	to (5)	that (5)	like (8)		it's (14)	i (9)	it (7)	that (8)
6	was (3)	mean (4)	it (5)	yeah (7)		it (13)	think (9)	just (6)	is (7)
7	that (3)	a (4)	know (5)	cuz (7)		we (9)	just (7)	a (6)	a (6)
8	think (3)	don't (4)	is (4)	um (7)		so (8)	are (6)	that (6)	like (4)
9	is (3)	and (3)	think (4)	think (6)		they (8)	really (6)	is (6)	to (4)
10	with (3)	of (2)	one (3)	mean (5)		this (7)	should (5)	um (4)	so (4)

Table 8.8: Pattern of *I think* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (21)	i (21)	i (20)	and (24)	i think (298)	i (37)	s (48)	i (16)	i (13)
2	s (14)	think (9)	the (14)	but (20)		that (36)	i (15)	the (13)	to (12)
3	to (7)	the (8)	s (10)	well (14)		it (28)	should (12)	that (10)	that (11)
4	a (7)	it (8)	it (10)	uhm (14)		the (15)	is (11)	just (8)	the (11)
5	it (7)	a (7)	uhm (9)	yeah (11)		she (15)	of (8)	uhm (7)	it (10)
6	yeah (6)	that (6)	to (7)	that (9)		so (13)	m (7)	it (7)	a (9)
7	the (6)	at (6)	in (6)	yes (8)		he (13)	that (6)	of (7)	think (8)
8	yes (5)	to (5)	know (5)	mean (8)		you (12)	yes (5)	s (6)	mean (7)
9	in (5)	actually (5)	yeah (5)	no (7)		they (11)	would (5)	have (6)	s (6)
10	uhm (5)	in (5)	a (5)	think (6)		we (9)	you (5)	a (6)	more (5)

The patterns shown in Tables 8.3 to 8.8 reveal that *I think* tends to be Type A and this supports the conclusion drawn from the manual classification of *I think* that it is not primarily used as a DM by the NNSs and NSs under investigation.

To inspect more closely the use of Type B, the patterns of Type B *I think* in the NNSs' and NSs' speech are presented. In the NNSs' speech (see Tables 8.9 and 8.10), *I think* often co-occurs with the hesitation markers *eh* and *um* and in the monologues, *I think* tends to be repeated. This suggests that the NNSs could use *I think* as a delaying device.

Table 8.9: Pattern of Type B *I think* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	is (3)	a (3)	very (3)	us (2)	i think (30)	<i>i</i> (6)	<i>i</i> (5)	<i>i</i> (3)	think (2)
2	his (3)	was (2)	teaching (2)	well (2)		eh (6)	what (2)	we (2)	is (2)
3	um (2)	very (2)	that (2)	and (2)		you (2)	will (2)	think (2)	a (2)
4	to (2)	save (2)	eh (2)	teacher (2)		um (2)	think (2)	be (2)	
5	he (2)		i (2)			that's (2)	my (2)		
6			me (2)				s (2)		

Table 8.10: Pattern of Type B *I think* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1		in (3)	and (2)	eh (3)	i think (19)	a (5)	i (4)	see (2)	the (2)
2		the (2)		and (2)		eh (4)	yes (2)	but (2)	eh (2)
3		a (2)				um (3)			
4		b (2)				is (3)			

In the NSs' speech, it is also found that Type B *I think* co-occurs with the hesitation markers *um* and *uh*. However, the small numbers of incidence do not seem to indicate any prominence.

Table 8.11: Pattern of Type B *I think* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (2)	and (4)		and (3)	i think (20)	um (3)	uh (2)	to (2)	
2				um (2)		<i>i</i> (2)	they (2)	about (2)	
3				uh (2)					

Table 8.12: Pattern of Type B *I think* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (4)	it (4)	s (4)	uh (3)	i think (43)	in (6)	the (7)	uh (3)	that (4)
2	the (3)	that (3)	to (3)	s (3)		the (4)	of (3)	seventeen (2)	of (3)
3	uh (2)	the (2)	uhm (2)			a (3)	a (3)	them (2)	in (3)
4	that (2)	s (2)	uh (2)			to (2)	s (2)	a (2)	and (2)
5	but (2)	in (2)				uh (2)	<i>i</i> (2)	good (2)	seventies (2)
6	are (2)	a (2)				for (2)		or (2)	<i>i</i> (2)
7						about (2)			

Table 8.13: Pattern of Type B *I think* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	it (2)	the (4)	<i>i</i> (3)	there's (2)	i think (30)	is (4)	kinda (2)	one (3)	of (3)
2		to (3)	in (2)	is (2)		<i>i</i> (3)	they (2)	the (2)	is (2)
3		um (2)	her (2)			there's (2)	a (2)		in (2)
4						and (2)	is (2)		

Table 8.14: Pattern of Type B *I think* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	at (3)	<i>i</i> (6)	the (8)	mean (3)	i think (55)	uh (4)	that (5)	<i>i</i> (6)	mean (5)
2	met (2)	in (4)	<i>i</i> (4)	uhm (3)		is (3)	he (3)	of (3)	<i>i</i> (4)
3	in (2)	they (2)	a (3)	that (3)		mm (3)	<i>i</i> (3)	you (3)	and (3)
4	uhm (2)	to (2)	or (2)	this (2)		yeah (3)	yes (3)	no (3)	was (2)
5	of (2)	stars (2)	in (2)	yeah (2)		and (3)	but (2)	two (2)	what (2)
6	if (2)	all (2)	had (2)	twice (2)		<i>i</i> (3)	articles (2)	uhm (2)	headmaster (2)
7	for (2)	once (2)	can (2)	do (2)		does (2)	you (2)	and (2)	for (2)
8	burton (2)			and (2)		so (2)	met (2)	just (2)	know (2)
9	<i>i</i> (2)			is (2)		you (2)	eighty (2)	purely (2)	
10	have (2)			people (2)		yes (2)	do (2)	the (2)	
11				known (2)		it (2)	know (2)		
12						isn't (2)			
13						feature (2)			
14						oh (2)			
15						nineteen (2)			

Type B *I think* does not seem to collocate with other DMs. In the pattern of the NNSs'

monologues (Table 8.9), possible DM collocations identified are *well I think* and *I think I think* and in the pattern of the NSs' private direct conversations (Table 8.14), *I mean I think* is identified.

The frequency information and collocates of *I think* are used as starting points of the major analysis. The frequencies of *I think* in the six sub-corpora reveal that the phrase *I think* is more often used by the NNSs, which may be a result of generic constraint rather than speakers' proficiency and their non-native variety of English, but *I think* is not primarily used as a DM among the NNSs and NSs under investigation. This conflicts with the proposition in the literature, discussed in Section 8.2, (e.g. Stenström's (1994)) that *I think* is typically used as a DM rather than in a reporting clause.

8.4 Discourse aspects of *I think*

In this section, the positions where Type B *I think* occurs in an utterance/turn are first described and then the linguistic items which *I think* tends to co-occur with are discussed. All the occurrences of Type B *I think* in the NSs' highly monologic discourse mode in MICASE (89) and the unscripted monologues in ICE-GB (170) were manually analysed, but the sheer number of occurrences in the NNSs' monologues (1,019) and dialogues (9,465), the NSs' highly interactive discourse mode in MICASE (1,841) and the private direct conversations in ICE-GB (662) was unamenable to manual analysis; therefore, three sets of 100-line concordance samples extracted from each of the four sub-corpora were manually examined (see Section 3.3.7 for the random sampling procedure).

8.4.1 Positions in an utterance/turn

In this section the positions of Type B *I think* in an utterance/turn are discussed. The distribution and percentages in the six sub-corpora under investigation are shown in Table 8.15 below.

There is a marked difference in the position in an utterance/turn of Type B *I think* across the two types of genre and between the two groups of speakers. In the NNSs' monologues, *I think* tends to occur in an extra-clausal position (63.3%), whereas in the NSs' monologues in MICASE (55%) and ICE-GB (83.7%), it tends to occur in an intra-clausal position. In the three sub-corpora of the dialogic genres, the most frequent position of *I think* is turn-medial.

Table 8.15: Distribution of the positions in an utterance/turn of Type B *I think*

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues			
	Random samples' (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)		
Positions in an utterance of Type B <i>I think</i>	30	100	20	100			43	100
Extra-clausal: utterance-initial	1	3.3	0	0.0			0	0.0
Extra-clausal: utterance-medial	15	50.0	63.3	9	45.0	45.0	7	16.3
Extra-clausal: utterance-final	3	10.0		0	0.0		0	0.0
Intra-clausal: after an M-	2	6.7		6	30.0		22	51.2
Intra-clausal: after an MA	0	0		0	0.0		0	0.0
Intra-clausal: after an MF	2	6.7	36.7	1	5.0	55.0	3	7.0
Intra-clausal: others	7	23.3		4	20.0		11	25.6
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations			
	Random samples' (times)	Percentage (%)	Random samples' (times)	Percentage (%)	Random samples (times)	Percentage (%)		
Positions in a turn of Type B <i>I think</i>	21	100	35	100			57	100
Extra-clausal: turn-initial	3	14.3		3	8.6		5	8.8
Extra-clausal: turn-medial	7	33.3	81.0	13	37.1	62.9	25	43.9
Extra-clausal: turn-final	7	33.3		6	17.1		16	28.1
Intra-clausal: after an M-	2	9.5		6	17.1		5	8.8
Intra-clausal: after an MA	2	9.5		1	2.9		1	1.8
Intra-clausal: after an MF	0	0.0	19.0	3	8.6	37.1	1	1.8
Intra-clausal: others	0	0.0		3	8.6		4	7.0

8.4.1.1 *I think* in extra-clausal position

Generally, in the speech of the NNSs and NSs under investigation, most of the instances of Type B *I think* are placed in extra-clausal turn-medial position, as exemplified in Excerpt (8.4.1):

(8.4.1)

P: Turn-medial

B:... eh... you should know, marks really are really important, but it_j's not the only important... **[I think]**, you could see in in .my classmates, there was some will who can also work well and study well, and those who did not take any organizations is not always the one who wonders the first scholarship.

(SECCL: C00-58-07)

It can be seen in Table 8.15 above, fewer instances of *I think* are placed in utterance/turn-initial or -final positions, except those in the sub-corpus of the private direct conversations in ICE-GB. In the work of Biber *et al.* (1999: 983), *I think* as a “comment clause” is usually placed in turn-final position, shown in Excerpt (8.4.2), and occurs more

often in British English than American English. This phenomenon is found in the sub-corpora of the British NSs' private direct conversations in ICE-GB and the American NSs' highly interactive discourse mode in MICASE. The proportion of *I think* in turn-final position in ICE-GB (28.1%) is larger than that in MICASE (17.1%).

(8.4.2)

P: Turn-final

A: There isn't room in their flat for that <,> They'd have to sell one |**I think**

B: Where do you live

C: Ladbroke Grove

A: No

(ICE-GB: S1A-017)

8.4.1.2 *I think* in intra-clausal position

In the sub-corpus of the unscripted monologues in ICE-GB, a higher percentage, 83.7%, of the occurrences of *I think* is placed in an intra-clausal position, occurring more often after an incomplete message (M-) element (51.2%), as in Excerpt (8.4.3).

(8.4.3)

P: M- + *I think* +
+M

A: But the reality of the second age for a woman <,> and the importance of the transition to the third <,> is |**I think**| a key element in the system

(ICE-GB: S2A-038)

8.4.2 Contexts where Type B *I think* tends to occur

The positions of Type B *I think* in an utterance/turn described in the last section are relevant to the present discussion of the contexts where *I think* tends to occur. Type B *I think* is found to co-occur with 1) hesitation markers, pauses and restarts, 2) personal opinions and evaluation, 3) factual information, 4) concluding remarks and 5) questions. Tables 8.16 to 8.21 at the end of this section illustrate the proportion of co-occurrence of *I think* in relation to their position in an utterance/turn in the six sub-corpora.

On average, approximately one-third of the instances of Type B *I think* in the NSs' speech and one-quarter in the NNSs' speech co-occur with personal opinions and evaluation, which is the main use of Type A *I think*. In a sense, compared with the phrases, *you know*, *I mean* and *you see*, discussed in Chapter 7, between Type A *I think* and Type B *I think* the distinction is less clear. The distinction depends more on the positions where *I think* is placed and whether there is a *that*-clause attached to it. As discussed below, Type B *I think* has the same functions as Type A.

8.4.2.1 *I think* co-occurring with hesitation markers, pauses and restarts

I think co-occurring with hesitation markers, pauses and restarts is more than three times as frequent in the NNSs' speech as in the NSs' speech. It accounts for 70% and 57.1% of the instances in the NNSs' monologues and dialogues respectively. In Excerpt (8.4.4), *I think* co-occurs with the hesitation markers, *er* and *um*, and pauses. In Excerpt (8.4.5), the utterance after *I think* repeats part of the previous utterance.

(8.4.4)

P: Utterance-medial
E: Hesitation markers
er and *um*; pauses
F: To search for
contents and lexical
words

Um... ener... Ier... I |**I think**| um... the piano is my favorite, is my favorite
game. (SECCL: 99-35-19)

(8.4.5)

P: MA + *I think* + +M
E: Item (2) repeats
part of Item (1) to
restart
F: To search for
contents and lexical
words

A:'en'do you think that so I think'er.... .in the .geology department. it is
very. .it is need not only the knowledge, but also the energy eh... 'so'the
male ⁽¹⁾the male'er...'is |**I think**| ⁽²⁾is better than the female. (SECCL: 02-61-05)

As argued in the use of *like* (see Section 4.4.2.1), even though it is not possible to draw definitive conclusions about the speaker's cognitive process from the use of DMs, it can be argued that when *I think* co-occurs with hesitation markers and pauses, it does so because the speaker is searching for content information or appropriate lexical expressions. This type of co-occurrence seems to suggest that the speakers are using *I think* as a filler while formulating what to say next. If this is true, it clarifies why this co-occurrence occurs more often in the NNSs' speech, as NNSs need more time to express their meaning in a foreign language. In De Cock's study (2007: 222), it is also found that French NNSs' *I think* is often surrounded by repetitions and hesitations.

8.4.2.2 *I think* co-occurring with personal opinions and evaluation

Type B *I think* co-occurring with personal opinions and evaluation is frequently used. In this category, there are some types of co-occurrence, such as positive evaluation, negative evaluation, mitigator *just*, vague language *sort of*, etc. The instances of *I think* in Excerpts (8.4.6) and (8.4.7) co-occur with a positive evaluation; here, the function of *I think* seems to

limit the evaluation to a personal opinion in order to avoid its being taken as a generally-accepted idea.

(8.4.6)

P: Turn-final

E: Item (1) a positive evaluation

F: To avoid being too assertive

A: Eh..., this morning I have received a lot of books, eh... .. so, eh..., I found that perhaps... the... studies in university... eh... also... receive a high pressure. What's your opinion?

B: Oh, I don't think so. ⁽¹⁾ **Study in university is also an interesting... and... attractive thing, |I think.**

(SECCL: C00-11-33)

(8.4.7)

P: M- + *I think* + OI

E: Item (1) a positive evaluation

F: To avoid being too assertive

A: ⁽¹⁾ **It's very tempting |I think|** uh to cut back investment <> whether you're in business <> uh because that's a way of propping up the share price and making sure that you're the cash flow is maintained

(ICE-GB: S2A-023)

In Excerpts (8.4.8) and (8.4.9), *I think* co-occurs with a negative evaluation. It can be argued that *I think* is used to reduce the impact of negative evaluation; this is Brown and Levinson's interpretation of *I think* (1987: 164, 171): a face-saving device in case of criticism.

(8.4.8)

P: M- + *I think* + +M

E: Negative evaluation

F: To soften the criticism

He taught our... he taught... taught us math. Um... his skill of teaching, |I think|, is not very good. The reason why I remembered... him, because I think he is a very good man.

(SECCL: 01-01-23)

(8.4.9)

P: M + *I think* + MS

E: Negative evaluation

F: To soften the criticism

A: yeah. don't even use chromatids in this class cuz it'll confuse you |i think| more than anything. just know, this a, chromosome, that's unreplicated, this is a chromosome, that is replicated.

(MICASE: DIS175JU081)

I think co-occurs with such indications of reducing commitment as the mitigator *just* and vague language *sort of*, as shown in Excerpts (8.4.10) and (8.4.11). The use of *I think* leads one to suppose that the opinions co-occurring with *I think* are meant to appear less assertive.

(8.4.10)

P: Turn-final

E: Mitigator *just*

F: To avoid being too assertive

S5: enhancers aren't t- are those transcription factors? are there just_ they need transcription, cuz they need activators <OVERLAP1> in order </OVERLAP1>

S4: <OVERLAP1> right enhancers are **just** something </OVERLAP1> that want the transcription factors |**i think**

S5: <OVERLAP1> enhancers are not_ aren't the (active) </OVERLAP1>

(MICASE: SGR175MU126)

(8.4.11)

P: MS + *I think* + MS

E: Vague language
sort of

F: To avoid being too assertive

A: And above that you've got a European sky with blue and white <,> merging together |**I think**| in a sort of wet technique <,> The painter is again a Hindu

(ICE-GB: S2A-059)

This category is highly represented in the NSs' speech with 5 (25%) out of 20 instances in the highly monologic discourse mode in MICASE, 15 (42.9%) out of 35 instances in the highly interactive discourse mode in MICASE, 18 (41.9%) out of 43 instances in the unscripted monologues in ICE-GB and 18 (31.6%) out of 57 instances in the private direct conversations in ICE-GB. Among these instances, a high percentage of the instances (34.9%) in the unscripted monologues in ICE-GB occurs in an intra-clausal position, as shown in Excerpt (8.4.12). *I think* is placed after the key information, *Noah*. It seems that *I think* is used to draw attention to the subject. This use is not found in the NNSs' speech.

(8.4.12)

P: M- + *I think* + +M

E: Following the subject

F: To focus on key information

A: For some of us it might be <,,> the second sighting <,> of the airship <,,> Now Noah <,> |**I think**| had quite good reasons for inviting his company onto the ark <,> two by two

(ICE-GB: S2A-040)

The instances of Type B *I think* in this category are not particularly distinct from those of Type A, except the flexible positions of Type B. In Biber *et al.*'s work on the grammar of spoken and written English (1999: 197), *I think* is categorised not as a DM, but as a stance marker, an "insert" in speech. This is probably because the researchers found it unnecessary to make a clear distinction.

Based on this type of co-occurrence, it is interpreted that the function of *I think* as a DM is either to indicate that the co-occurring opinion/evaluation is personal or to mitigate personal opinions and evaluation, acting as a softener to reduce the impact of negative evaluation and to avoid being too assertive with positive evaluation. *I think* as a DM can be pragmatically used as a hedge, which is similar to the referents of the semantic term *epistemic modality*

(Coates 2003: 331), for the avoidance of threatening face.

8.4.2.3 *I think* co-occurring with factual information

It has often been noted that expressions which are associated with uncertainty do not necessarily indicate actual uncertainty, but are used to avoid appearing too assertive (Coates 2003). In Excerpt (8.4.13), *I think* in turn-final position, co-occurring with factual information, seems to either mark genuine uncertainty about the fact, *the head is darker than the body*, or simulate uncertainty in order not to appear too assertive. Speaker SU-f is a third-year and above undergraduate. It is difficult to argue for the (un)certainty of the speaker about the fact. If this speaker is a lecturer, it is rather impossible that the speaker does not know the bird. In this case, *I think* can be possibly used to downplay the authority.

(8.4.13)

P: Turn-final

E: Item (1) a fact

F: Either to express uncertainty or to appear less assertive

SU-m: <OVERLAP1> they have that on the computer, (but do they have it here?)
</OVERLAP1>

SU-f: <OVERLAP1> i know cuz you know Francie'll be like wrong. um that's not that. </OVERLAP1>

SU-f: <OVERLAP1> that one that just flew off </OVERLAP1> had a white tail band. white

SU-f: i have seen some white.

SU-f: i saw some white on <OVERLAP1> its tail. </OVERLAP1>

SU-f: <OVERLAP1> i think ⁽¹⁾ **the head** </OVERLAP1> **is darker than the body**, |i **think**.

SU-f: (does coffee come out)

SU-f: okay

(MICASE: LAB175SU026)

In Excerpt (8.4.14), the speakers are probably looking at photos. *I think* by Speaker 3 follows the fact *that was us that was* and Speaker 4 gives a positive response *oh yeah I think it was*. Similar to the instance in Excerpt (8.4.13), it is difficult to be precise whether or not the speaker is certain about the fact, but it seems that *I think* is used either to mark uncertainty or to sound less assertive.

(8.4.14)

P: Turn-final

E: Item (1) a fact

F: Either to express uncertainty or to appear less assertive

S3: got it <EVENT WHO="SS" DESC="LAUGH"></EVENT> (this isn't even eight-
was it la-) were you guys the group that had an eighties day? no

S1: <OVERLAP1> (xx) </OVERLAP1>

S6: <OVERLAP2> must be Tuesday </OVERLAP2> we had an eighties theme day.

S4: <OVERLAP2> (really) </OVERLAP2> <OVERLAP1> everything </OVERLAP1>
was coming back to the eighties <OVERLAP1> (xx) </OVERLAP1>

S3: <OVERLAP1> oh wait no </OVERLAP1> ⁽¹⁾ **that was us. that was. |i think.**

S4: oh yeah i think it was.

(MICASE: SGR200JU125)

In Excerpts (8.4.15) and (8.4.16), *I think* co-occurs with numerical information. In Excerpt (8.4.15), *I think* prefaces the year *nineteen eighty-two* and there is no vague language surrounding this information; therefore, it is reasonable to assume that *I think* is primarily used to be less assertive. In contrast, in Excerpt (8.4.16), *I think* co-occurs with frequency information, *once or twice*, which indicates the speaker is not sure about how many times s/he met the headmaster. In this case, it is possible that *I think* is used to express uncertainty.

(8.4.15)

P: Turn-medial

E: Item (1) is numerical information

F: Primarily to appear less assertive

F: I mean I seem I don't read books for pleasure at all I mean |**I think**| ⁽¹⁾

nineteen eighty-two was last time I read a book

(ICE-GB: S1A-013)

(8.4.16)

P: Turn-medial

E: Item (1) is numerical information

F: To express uncertainty

A: I met ⁽¹⁾ **once or twice** |**I think**| I met the headmaster there when he came to
some <,> you know when people come from schools to check out the place
once a year

(ICE-GB: S1A-033)

Another type co-occurrence of factual information relates to places, as shown in Excerpts (8.4.17) and (8.4.18). There is not enough contextual information to speculate that *I think* is used to mark uncertainty or to simulate it in order to reduce the level of commitment to the factual information.

(8.4.17)

P: M- + *I think* + +M
E: Item (1) is information relating to places
F: Either to express uncertainty or to reduce commitment

S1: oh i'd love to try <OVERLAP1> a chunk. </OVERLAP1>

SU-f: <OVERLAP1> yeah it </OVERLAP1> tastes good. it's very good. (xx) tasty.
<EVENT DESC="AUDIO DISTURBANCE"></EVENT>

S1: rubbish over onto the other boat cuz the rubbish bag will be, |i think|⁽¹⁾ on there.

(MICASE: LAB175SU032)

(8.4.18)

P: Turn-final
E: Item (1) is information relating to places
F: Either to express uncertainty or to reduce commitment

A: She's a student⁽¹⁾ **at Saint Martin's |I think**

D: Is she

A: Uhm

(ICE-GB: S1A-020)

I think co-occurring with factual information is seldom used by the NNSs under investigation; only two instances occurring in their monologues. By contrast, this type of co-occurrence is most often used by the NSs, representing 45% in the highly monologic discourse mode in MICASE, 28.6% in the highly interactive mode in MICASE, 41.9% in the unscripted monologues in ICE-GB and 50.9% in the private direct conversations in ICE-GB.

8.4.2.4 *I think* prefacing concluding remarks

The two types of co-occurrence discussed in this section and the next section are least often used. *I think* is found to preface a concluding remark, as shown in Excerpts (8.4.19) and (8.4.20). *That* in Excerpt (8.4.19) and *they* in Excerpt (8.4.20) encapsulate the ideas that the speakers have conveyed. My interpretation of this use is that *I think* may have been used here to reduce the impact of imposing personal conclusions on others.

(8.4.19)

P: Utterance-medial
E: Item (1) is a concluding remark
F: To reduce the impact of imposing

But after all, as Mrs Brown was very easy to get angry, we were... we have to hand in our homework and pay attention... pay more attention in classroom when he was teaching us... **|I think|⁽¹⁾ that's what we did.**

(SECCL: B01-01-16)

(8.4.20)

P: Utterance-medial
E: Item (1) is a
concluding remark
F: To reduce the
impact of imposing

S2: so, those are the values um, those are the tricky problems, some of them are pretty nitty-gritty some of them are pretty lofty, um, but they come um floating, across our desk um on a regular basis and |i think|, um ⁽¹⁾ **they are, finally convincing me in trying to put together this talk that, a provost does address, something valuable, sometimes** <EVENT DESC="LAUGH" WHO="SS"></EVENT> so, thank you <EVENT DUR=":14" WHO="SS" DESC="APPLAUSE"></EVENT>

(MICASE: COL999MX036)

I think prefacing concluding remarks occurs 4 times (7.8%) in the 51 instances in the NNSs' speech, 3 times (5.5%) in the 55 instances in the two sub-corpora of MICASE and there is no instance in ICE-GB.

8.4.2.5 *I think* prefacing questions

The other least often used co-occurrence is with questions. *I think* is found to preface a question, as exemplified in Excerpt (8.4.21). In Excerpt (8.4.22), *I think* prefaces an indirect question. It seems that the speakers in these two cases use *I think* to sound less imposing before raising a question. There are only 4 instances of this in the NSs' speech and only one instance (shown in Excerpt (8.4.22)) in the NNSs' speech.

(8.4.21)

P: Turn-medial
E: Item (1) is a
question
F: To sound less
imposing

S9: well there you go,
S7: well no no no wait a minute |i think|, ⁽¹⁾ **are you messing with my cerebellum, now?**
S8: (this is) the cerebellum yeah
S7: yeah

(MICASE: LAB500SU044)

(8.4.22)

P: Turn-initial
E: Item (1) is an
indirect question
F: To sound less
imposing

B: Yes, that's also very important. But you know that if we want to take our future's job, we must get, get some advice <advices> or experience before we attend the job. So I think to take part-time job, I mean to, the job is related to our study, we can get much experience and to get preparation for the future's job.
A: Oh, |I think|, ⁽¹⁾ **I want to know if there are the other, some other advantages about part-time job.**
B: Yes, I think there's <re> another advantage.

(SECCL: C96-13-13)

8.4.2.6 Summary of the contexts where Type B *I think* tends to occur

Tables 8.16 to 8.21 below illustrate the distribution of the identified types of co-occurrence in

relation to the positions in an utterance/turn of Type B *I think*. There is a marked difference in the use of *I think* as a DM between the NNSs and NSs. In terms of the position in an utterance/turn of *I think*, in the NNSs' speech *I think* seldom occurs in intra-clausal position, whereas in the NSs' monologues more than half the instances of *I think* occur in an intra-clausal position. In terms of the types of co-occurrence, most of the instances of *I think* in the NNSs' speech (63.5% of instances on average) co-occur with hesitation markers, pauses and restarts. This could suggest that the NNSs use *I think* as a filler in their speech. The NSs also use *I think* in this way, but much less often (18% of instances on average).

In the NSs' speech, the frequent types of co-occurrence are personal opinions and evaluation and factual information. They are highly represented with more than two thirds in MICASE and four fifths in ICE-GB. When *I think* co-occurs with personal opinions and evaluation, it is interpreted as a hedge to avoid being too assertive with a positive evaluation and to soften a negative one. To some extent, this use is similar to the use of Type A *I think*, thereby downplaying the distinction between Types A and B. This is probably one of the reasons why the previous studies (e.g. Simon-Vandenberg (2000) and Fortanet (2004) discussed in Section 8.2.3) make no distinction between Types A and B of *I think*.

The other type of co-occurrence, which is frequently observed in the NSs' speech but seldom in the NNSs' speech, is *I think* co-occurring with factual information. It is probably used either to mark uncertainty or to sound less assertive.

Two more types of co-occurrence, which are infrequently used by both groups of the speakers, are concluding remarks and questions. Type B *I think* may be used to reduce the impact of imposing personal conclusions on others and to sound less imposing when raising a question.

Table 8.16: Distribution of co-occurrence of *I think* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	70.0	1	3.3	9	30.0	2	6.7	1	3.3			2	6.7	6	20.0
2. Personal opinions & evaluation	13.3			1	3.3	1	3.3	1	3.3					1	3.3
3. Factual information	6.7			2	6.7										
4. Concluding remarks	10.0			3	10.0										
5. Questions	0														
Unclassified	0														
Occurrences: 30 out of 300 (random samples)	100.0	1	3.3	15	50.0	3	10.0	2	6.7			2	6.7	7	23.3

Table 8.17: Distribution of co-occurrence of *I think* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	57.1	2	9.5	6	28.6	1	4.8	1	4.8	2	9.5				
2. Personal opinions & evaluation	33.3			1	4.8	5	23.8	1	4.8						
3. Factual information	0														
4. Concluding remarks	4.8					1	4.8								
5. Questions	4.8	1	4.8												
Unclassified	0														
Occurrences: 21 out of 300 (random samples)	100.0	3	14.3	7	33.3	7	33.3	2	9.5	2	9.5				

Table 8.18: Distribution of co-occurrence of *I think* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%				
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%			after an MA	%	after an MF	%
1. Hesitation markers; pauses; restarts	25.0			1	5.0			1	5.0			1	5.0	2	10.0
2. Personal opinions & evaluation	25.0			3	15.0			2	10.0						
3. Factual information	45.0			4	20.0			3	15.0					2	10.0
4. Concluding remarks	5.0			1	5.0										
5. Questions	0														
Unclassified	0														
Occurrences: 20	100.0			9	45.0			6	30.0			1	5.0	4	20.0

Table 8.19: Distribution of co-occurrence of *I think* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%				
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%			after an MA	%	after an MF	%
1. Hesitation markers; pauses; restarts	20.0	1	2.9	2	5.7					1	2.9	3	8.6		
2. Personal opinions & evaluation	42.9	2	5.7	4	11.4	1	2.9	5	14.3					3	8.6
3. Factual information	28.6			5	14.3	4	11.4	1	2.9						
4. Concluding remarks	2.9					1	2.9								
5. Questions	5.7			2	5.7										
Unclassified	0														
Occurrences: 35 out of 300 (random samples)	100.0	3	8.6	13	37.1	6	17.1	6	17.1	1	2.9	3	8.6	3	8.6

Table 8.20: Distribution of co-occurrence of *I think* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	16.3							1	2.3			3	7.0	3	7.0
2. Personal opinions & evaluation	41.9			1	2.3			15	34.9					2	4.7
3. Factual information	41.9			6	14.0			6	14.0					6	14.0
4. Concluding remarks	0														
5. Questions	0														
Unclassified	0														
Occurrences: 43	100.0			7	16.3			22	51.2			3	7.0	11	25.6

Table 8.21: Distribution of co-occurrence of *I think* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Hesitation markers; pauses; restarts	14.0	1	1.8	4	7.0					1	1.8	1	1.8	1	1.8
2. Personal opinions & evaluation	31.6	1	1.8	7	12.3	8	14.0	2	3.5						
3. Factual information	50.9	3	5.3	12	21.1	8	14.0	3	5.3					3	5.3
4. Concluding remarks	0														
5. Questions	3.5			2	3.5										
Unclassified	0														
Occurrences: 57 out of 300 random samples)	100.0	5	8.8	25	43.9	16	28.1	5	8.8	1	1.8	1	1.8	4	7.0

8.5 Further investigation

8.5.1 Importance of the type of activity

I think is not primarily used as a DM, in particular in the NNSs' speech. About 90% of the instances of the phrase *I think* are used to express personal opinions and follow the *S-V-O* structure. As noted above, the nature of the two types of genres is likely to influence speakers' linguistic choices. In the NNSs' monologues, the speakers mainly talk about personal experiences and past events, while in the dialogues the two speakers are role-playing, discussing topics and exchanging opinions (see Appendix 1 for the rubrics set up in the oral examinations). Given these two different contexts, it is to be expected that the present tense *I think* will occur more frequently in the dialogues than the monologues, and vice versa for the past tense *I thought*. Investigation reveals that the frequency of *I think* in the dialogues (158) is more than five times that in the monologues (30) and that the frequency of *I thought* in the monologues (9) is more than eleven times that in the dialogues (0.8). Moreover, the incidence of *I think* and *I thought* in the monologues (30 vs. 9) is close, whereas in the dialogues (158 vs. 0.8) it is less close. This quick comparison shows that the type of genre (monologic and dialogic) and type of activity (narrative and discussion) are key factors in using the phrase *I think*.

Table 8.22: Frequencies of *I think* and *I thought* in the non-native speakers' speech

Corpus	Corpus size (tokens)	Raw freq. of <i>I think</i> (times)	Normalised freq. per 10,000 words (times)
SECCL: Monologues	336,303	1,015	30
SECCL: Dialogues	596,639	9,412	158
Corpus	Corpus size (tokens)	Raw freq. of <i>I thought</i> (times)	Normalised freq. per 10,000 words (times)
SECCL: Monologues	336,303	290	9
SECCL: Dialogues	596,639	49	0.8

8.5.2 Alternatives to *I think* used by the non-native speakers

It is found that the phrase *I think* is over-represented in the NNSs' speech. Some similar studies (e.g. Xu and Xu (2007)) conclude that Chinese NNSs "overuse" *I think*. Nevertheless, it seems that the influence of the constraints relating to genre has been overlooked. One contributing factor in the NNSs' use of *I think* in this study is that most of the exam questions, in particular in the dialogues, ask the NNSs to express and discuss their opinions (see Appendix 1 for the topics for discussion). Furthermore, the use of *do you think* to raise

questions is frequent, thereby leading to many responses beginning with *I think* (see Table 8.2 above).

In addition to *I think*, many other “modal-like expressions” (Hunston 2011) can be used to express personal opinions, ideas or suggestions; for instance, *in my opinion*, *it seems to me*, *I believe*, *maybe* and *possibly*. Table 8.23 below illustrates that the five alternatives to *I think*, except for *maybe*, are rarely used by the Chinese NNSs. In the dialogues, 65 out of 241 (27%) occurrences of *in my opinion* co-occur with *I think*, where either *in my opinion* or *I think* is redundant. It seems to be true that the Chinese NNSs tend to rely heavily on using *I think* instead of other expressions. Although *maybe* is frequently used, the occurrences include those implying uncertainty. They need to be manually examined to analyse the use of *maybe* to make a suggestion.

Table 8.23: Frequencies of alternatives to *I think* used by the non-native speakers

Phrases	SECCL: Monologues	SECCL: Dialogues	Remarks
<i>I think</i>	1,015	9,412	
<i>in my opinion</i>	19	241	1 instance in monologues and 65 instances in the dialogues co-occur with <i>I think</i> .
<i>it seems to me</i>	1	3	
<i>I believe</i>	27	45	
<i>maybe</i>	340	2,387	including those imply uncertainty
<i>possibly</i>	3	1	

8.6 Chapter summary and conclusions

The phrase *I think* is not a central DM. Some studies (e.g. Biber *et al.* (1999), Simon-Vandenberg (2000) and Fortanet (2004)) investigate *I think* without making a distinction between its non-discourse and discourse uses. This chapter uses an innovative approach to analysing the phrase *I think*. The three criteria of Biber *et al.* (1999: 1076-1078) for determining “utterance launchers” as DMs, the *LUG* analysis (Sinclair and Mauranen 2006) as well as the co-texts of *I think* are of use in the manual classification of *I think* as either non-discourse (Type A) or discourse (Type B) use. As I hypothesised, Type A *I think* is frequently used in the Chinese NNSs’ speech examined in this thesis, since they were being asked to give opinions. I hypothesised that in the NSs’ speech, Type B *I think* occurs more often than Type A, but there is evidence to the contrary. Less than 20% on average of the occurrences of *I think* use of Type B. Like the NNSs, the NSs also frequently use Type A *I think*.

On the basis of the collocation phenomena, the functions of Type B *I think* are discussed and summarised in Section 8.4.2.6 above. There is a marked difference in the use of Type B *I think* between the two groups of speakers. This can be attributed to the genre and type of activity. The frequency comparison of *I think* and *I thought* shows that the type of genre (monologic and dialogic) and type of activity (narrative and pair discussion) are key factors in using the phrase *I think* in the NNSs' speech.

Both Type A *I think* and Type B *I think* are over-represented in the speech of the NNSs under investigation. Some possible reasons for this are: 1) the use of *do you think* leading to the response beginning with *I think*; 2) the context offering chances to exchange opinions; and 3) the NNSs' preference for using *I think* over other options for expressing epistemic stance. A frequency comparison between some alternatives to *I think* (see Table 8.23) reveals that the NNSs have a strong preference for using *I think* over other possible expressions. Two implications can be considered: 1) the over-representation of Type A *I think* and Type B *I think* in the Chinese NNSs' speech is a feature of a non-native variety of English and 2) the NNSs overuse *I think* at the expense of other modal-like possibilities. The former requires that NNSs become more tolerant and inclusive of the versions of English that they hear around them. The latter requires pedagogical interventions for the Chinese NNSs. These two implications are discussed below in Chapter 12.

CHAPTER 9: ANALYSIS OF NOW

9.1 Introduction

This chapter begins with my hypotheses in the use of *now*, followed by a review of the literature. As presented in the preceding chapters, the frequency information and patterns of *now* are introduced first, giving an overall sense of the use of *now*. The major part of the analysis is the discourse aspects of Type B *now*, looking at its position in utterances/turns and the collocation phenomena surrounding *now*. The identification of co-occurrence leads to the interpretations of the functions of *now* as a DM.

Now, as an adverb, is usually acquired by NNSs at the early stage of learning English and it is one of the common adverbs in spoken English. In this chapter, I expect to identify a larger proportion of the non-discourse use (Type A) of *now* in the Chinese NNSs' speech and find out how the discourse use (Type B) of *now* is made by the NNSs as opposed to the NSs.

In the analysis of *like* in Chapter 4, it is found that the Chinese NNSs are more likely to use Type A *like* than the NSs are. It seems that the NNSs do not employ Type B *like* as the NSs do. In this chapter, I investigate the word *now*, which is similar to the words *like* and *well* in Chapters 4 and 6, in that there is usually a clear-cut distinction between Type A and Type B and in that the NNSs are probably more familiar with the uses of Type A word than those of Type B. Therefore, I hypothesise that in the NNSs' speech, Type A *now* is the more often used, whereas in the NSs' speech, Type B *now* is predominantly used. If this is so, it is hypothesised that the uses of Type B *now* made by the NNSs are not as varied as those by the NSs. The chapter aims to find how similar or different is the use of *now* in the speech of the NNSs and NSs. My hypotheses are tested within the framework of the core research questions addressed in this thesis (see Section 1.1.2 of Chapter 1).

As noted above, the distinction between Type A *now* and Type B *now* is not always obvious. Like the phrases discussed in the previous two chapters, *now* in clausal-initial position is problematic. The ways of distinguishing *now* between Types A and B are exemplified below.

Type A *now*, as a deictic item, refers to the time of the utterance it occurs in, not the time of the proposition. The instances of *now* in Excerpts (9.1.1) and (9.1.2) are typical uses of Type A.

(9.1.1)

.....there's uh been a lot of work done with. um, soybean, fifty percent, of the, soy bean and corn crop grown in the world right **now** is genetically modified in some way.

(MICASE: LEL405JG078)

(9.1.2)

..... so make sure you, don't confuse, the experiment on twenty-one, the experiment i'm **now** about to go over. in this experiment, they took some some mouse melanoma cells were taken and injected into the tail vein of the mice,

(MICASE: LEL175SU106)

Type B *now*, as opposed to the temporal meaning of Type A *now*, performs discursual functions. There are three ways, according to the literature, of distinguishing between Type A *now* and Type B *now*. First, the use of prosodic information is found to be of importance. Type B *now* is usually a single tone unit and is followed by a brief pause (Schiffrin 1987: 231, Hirschberg and Litman 1993: 509, Aijmer 2002: 59). However, the prosodic transcription is not available in all of the three corpora under investigation. Therefore, the prosody of *now* is not taken as a criterion for the classification of Types A and B in the present study. Second, the position of *now* can be analysed. Hirschberg and Litman's study (1993) finds that a high proportion of Type B *now* is placed in a clausal-initial, with Type A *now* in a non-initial position. Third, the distinction can be made with reference to lexical collocates. For example, Type A *now* and the adverb *then* are not expected to co-occur (Schiffrin 1987: 230-232). When *now* co-occurs with another *now*, it is very likely that only one of them is an adverb. The first instance of *now* in Excerpt (9.1.3) below is Type B, because the second instance is an adverb.

(9.1.3)

..... But I didn't mind. I I was eager to learn learn learn it, so I kept it up. In those those, |**now**| I feel happy **now**, because learned to, I have learned to ride a bicycle. I can ride a bicycle around the ring road. That's all.

(SECCL: B99-35-08)

Now co-occurring with particles, such as *well*, *then* and *look*, is also treated as Type B (Aijmer 2002: 61).

The instances of *now*, in general, can be divided into Type A and Type B. In some cases, the positions, lexical collocates and *Linear Unit Grammar (LUG)* (Sinclair and Mauranen 2006) are used to make classification easier. However, there are some cases in which *now* is for discourse use as well as time reference. For instance, *now* in Excerpt (9.1.4) is ambiguous. It could be the use of Type A, referring to *at this time* and it could also be the use of Type B,

co-occurring with listing items and serving as a boundary marker for the organisation of the discourse. The second possibility seems to be more obvious and salient for the context, in this case a lecture; hence, it is classified as Type B.

(9.1.4)

P: Utterance-medial

E: Items (1) and (2) are listing items

F: To separate units and draw attention to the following point

..... you'll get ⁽¹⁾ **two phases** supernatant, and the precipitate uh form. so **[now]** we come to very ⁽²⁾ **first part**, first part of the lab that's the background. some terminology and stuff to get ready for. and i will say this, uh we're not aiming this week to get that far in this lab.

(MICASE: LEL200JU105)

My manual tagging of DMs is compared with that in the ICE-GB corpus. (This comparison is also made in the investigations of *like*, *well*, *you know*, *I mean* and *you see* and the results are shown in Appendix 5.) The comparison shows very similar results. This finding, which is also noted in the previous chapters, adds credibility to my manual examination process, as does the tagging in ICE-GB.

9.2 Previous studies of *now*

9.2.1 Grammatical aspect: Word classes

Type A *now* is categorised as an adverb, conjunction and pronoun (*Collins COBUILD Advanced Learner's English Dictionary* 2006: 978). In the speech of the NNSs and NSs under investigation, almost all the instances of Type A *now* are adverbs. *Now* as a conjunction or pronoun is seldom used.

Type B *now* does not belong to any word class, but rather to a generally-accepted category, *discourse marker*. This category has been widely used (e.g. Biber *et al.* (1999) and Carter and McCarthy (2006)).

9.2.2 Grammatical aspect: Syntactical structure

Type A *now* in the NNSs' and NSs' speech occurs mainly as an adverb, which means the present time or immediately. The syntactical structures of *now* as an adverb are flexible. *Now* can be placed at the beginning or the end of a clause. It can also be inserted between the subject and the verb of a clause.

Type B *now* is syntactically optional, but unlike other DMs occurring in various

positions, *now* generally occurs between two textual segments. In Example (9.2.1) below, *now* is placed between Items (1) and (2).

(9.2.1)

..... i've got a bunch of observations over here and they vary. and here's another bunch of observations and ⁽¹⁾they vary too. |now| ⁽²⁾part of the reason why they vary is because they came from different farms.

(MICASE: OFC575MU046)

Now occurring in intra-clausal positions is described in the *LUG* analysis (Sinclair and Mauranen 2006). Further details are provided in the next section (see Section 2.5 of Chapter 2, where the *LUG* analysis is dealt with in detail).

9.2.3 Linear Unit Grammar analysis of *now*

Now as an adverb in Example (9.2.2) below (the same as Example (9.1.1)) is part of an M element in the *LUG* analysis (Sinclair and Mauranen 2006) (see Appendix 4 for a list of the labels in *LUG*). In this case, *now* makes a propositional contribution.

(9.2.2)

.....there's uh been a lot of work done with. um, soybean, fifty percent, of the, soy bean and corn crop
M- OI +M- +M- OI +M M- OT M-
grown in the world right now is genetically modified in some way.

MS MS +M MS

(MICASE: LEL405JG078)

In Example (9.2.3) (the same as Example (9.2.1)), *now* as a DM is an O element, which does not augment knowledge but makes the discourse flow. It is further categorised as an OI element rather than an OT element, because this instance of *now* does not create cohesion at the textual level.

(9.2.3)

.....i've got a bunch of observations over here and they vary. and here's another bunch of observations
M- +M MS OT M OI M- +M
and they vary too. now part of the reason why they vary is because they came from different farms

OT M OI M- +M- OT +M

(MICASE: OFC575MU046)

9.2.4 Previous studies of *now* as a discourse marker

Now is one of the peripheral DMs. To my knowledge, little research has been done on *now*. However, two comprehensive studies of *now* have been made by Schiffrin (1987) and Aijmer

(2002). Schiffrin (1987) investigates *now* and *then* together, looking at them from their deictic meanings to the discourse uses. Aijmer (2002) sees *now* as a topic-changer and emphasises its textual and affective functions in discourse. In addition to these two studies, the use of Type B *now* is discussed in reference books, such as the *Longman Grammar of Spoken and Written English* (Biber *et al.* 1999) and *Cambridge Grammar of English* (Carter and McCarthy 2006).

The above studies reach agreement on two major functions of *now* as a DM. First, *now* is used to (re)open or close the discourse or topic (Biber *et al.* 1999: 1088, Carter and McCarthy 2006: 214). In a way, this use of *now* is similar to *well*, but they are not interchangeable. *Now* tends to be used in formal types of activity, for example, radio discussion (Aijmer 2002: 71-72).

Second, *now* is commonly taken as a boundary marker, indicating various kinds of shift. The common one is a shift of topic (Aijmer 2002: 62, Carter and McCarthy 2006: 112). The change is found to be justifications, explanations and elaborations of the topic preceding *now* (Aijmer 2002: 73). This use of *now* functions as a “frame” between two discursual segments (Sinclair and Coulthard 1975: 22). In addition to marking a shift of topic, *now* marks a change of “footing”¹⁷ (Goffman 1981: 124-159). *Now* is used to negotiate the floor or preface a turn-holding device, such as *wait a moment*, *let me finish* and *listen to me* (Schiffrin 1987: 241, Aijmer 2002: 93-94). *Now* also marks a shift of mood (e.g. from declarative to interrogative) and a shift of mode (from narrative mode to evaluative mode) (Schiffrin 1987: 240-241).

Other functions of *now* reported in the literature are to mark a sequence of events or actions¹⁸ (Schiffrin 1987: 240, Aijmer 2002: 83), to indicate emphasis and disagreement (Aijmer 2002: 92) and to mark comparative sub-topics in the discourse (Schiffrin 1987: 233).

A recent study of *now* by Fraser (2009) argues that *now* is an attention marker preceding topic orientation markers. Their four uses are 1) “return to a prior topic”, 2) “continuation with the present topic”, 3) “digression from the present topic” and 4) “introduction of a new topic”. These four broadly cover the functions of *now* as a DM.

¹⁷ The notion of *footing* is first introduced by Goffman (1981:127), who defines it as “significant shifts in alignment of speaker to hearers”. Shifts of footing are common in spontaneous speech. Shifting from reporting our current self, i.e. “addressing self”, to others can be marked para-linguistically or by code-switching.

¹⁸ Aijmer (2002) points out that this use of *now* can also be taken to present *now* as an adverb, but since it has the function of organising the discourse, it is taken as a discourse marker.

9.3 Frequency information in the speech of the non-native speakers and native speakers

9.3.1 Overall frequency of *now*

The overall frequency of *now* is shown in Table 9.1 below. In the NNSs' monologues and dialogues, there are 475 and 1,163 occurrences of *now* respectively. These raw counts are normed on a basis of 10,000 words. The normalised frequencies show that *now* occurs slightly more often in the dialogues than in the monologues (19.5 vs. 14.1 times per 10,000 words). In the NSs' speech, it is the other way around. There are more instances of *now* in the monologic genres than in the dialogic genres in MICASE and ICE-GB.

As with the word *like* in Chapter 4, the distinction between Type A *now* and Type B *now*, in most cases, can be drawn without difficulty. (A dividing line, however, between Type A and Type B in the cases of *you know*, *I mean*, *you see* and *I think* is more difficult to establish. This is discussed in Chapters 7 and 8.)

The instances of *now* are manually grouped into Types A and B. This classification reveals that *now* is not primarily used as a DM. The proportions of Type B in the six sub-corpora, shown in Table 9.1, range from 17.7% to 55.3% of all the instances of *now*.

Table 9.1: Frequency information of *now* in the non-native speakers' and native speakers' speech

Corpus	Word counts (tokens)	Raw freq. (times)*	Normalised freq. per 10,000 words (times)	Raw freq. of Type B (times)	Percentage (%)	Normalised freq. of Type B per 10,000 words (times)**
SECL: 1,143 monologues (Chinese NNSs)	336,303	475	14.1	65 out of 300 ^a	21.7	3.1
SECL: 1,143 dialogues (Chinese NNSs)	596,639	1,163	19.5	53 out of 300 ^b	17.7	3.5
MICASE: 13 transcripts of highly monologic discourse mode (American NSs)	134,096	367	27.4	203	55.3	15.1
MICASE: 48 transcripts of highly interactive discourse mode (American NSs)	577,996	1,175	20.3	145 out of 300 ^c	48.3	9.8
ICE-GB: 70 unscripted monologues (British NSs)	153,646	620	40.4	93 out of 300 ^d	31.0	12.5
ICE-GB: 90 private direct conversations (British NSs)	185,000	372	20.1	108	29.0	5.8

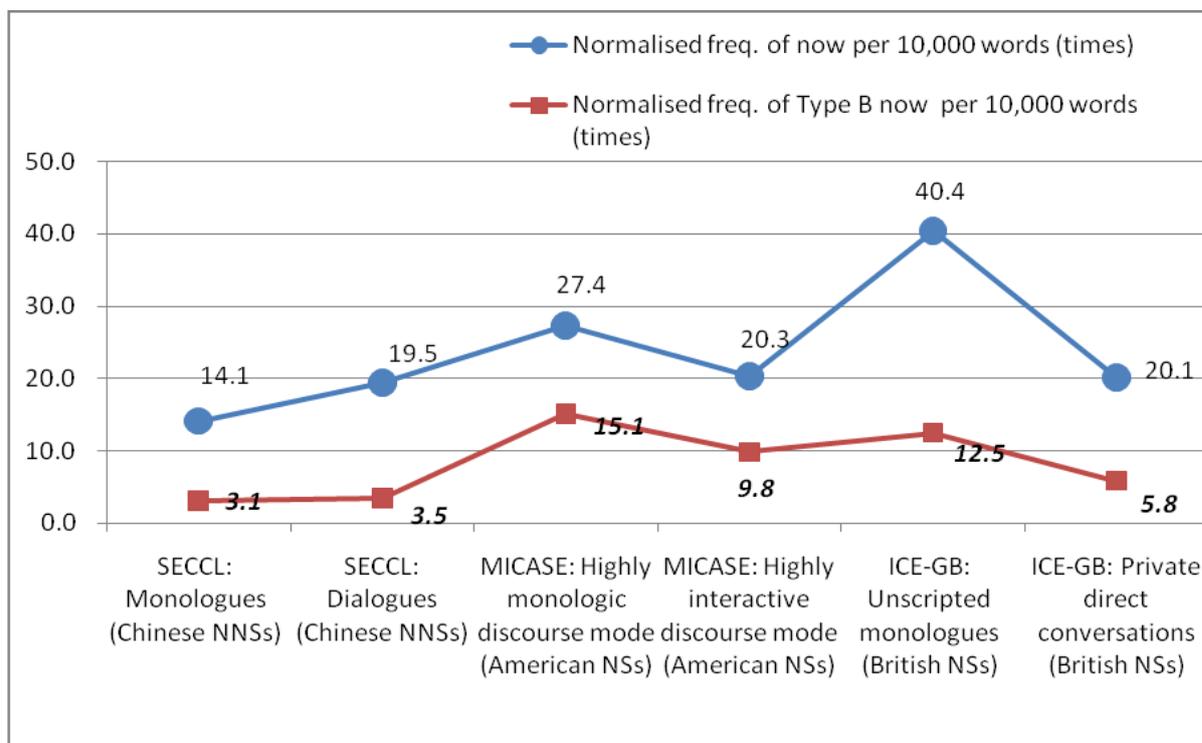
* Three sets of 100-line concordance samples are used for analysis when the incidence is over 400 times.

** For the cases of random sampling, the normalised frequencies of Type B per 10,000 words are based on an extrapolation of the percentages of the Type B word.

a, b, c and d in superscript: The number 300 is the total of the three sets of 100-line concordance samples, which reveal a similar distribution of Types A and B.

The raw frequencies of *now* are normed on a basis of 10,000 words and the normalised frequencies, ranging from 14.1 to 40.4 times per 10,000 words across the six sub-corpora, are shown in Table 9.1 above. The same normalisation is used on the incidence of Type B *now*, which ranges from 3.1 to 15.1 times. Figure 9.1 below shows the comparison of normalised frequencies of *now* across sub-corpora. It is evident that, in the NSs' speech, there are more instances of *now* in the monologic genres than in the dialogic genres. In the NNSs' speech, the frequency of *now* in the dialogues is slightly higher than that in the monologues. This frequency comparison seems to indicate that the case of *now* does not support my hypothesis that the more interactive the genres or types of activity are, the more instances of Type B words/phrases occur.

Figure 9.1: Comparison of normalised frequencies of *now* across corpora



The results of the test of statistical significance (see Appendix 6) indicate that there is a statistically significant relationship between the two types of genre in MICASE (LL: +25.97, p-value: < 0.0001) and in ICE-GB (LL: +42.02, p-value: < 0.0001). The positive LL scores indicate the over-representation in the sub-corpora of the monologic genres. This supports the finding discussed above that there are more instances of Type B *now* in the monologic genres than in the dialogic genres, even though this is not true in SECCL, where the difference between the two types of genre is not significant (LL: -1). Between the two groups of speakers, statistical significance lies in the comparisons in both the monologic genres and the dialogic genres (LL: -187.75, p-value: < 0.0001 between Corpora A1 and B1; LL: -141.14, p-value: < 0.0001 between Corpora A1 and C1; LL: -187.78, p-value: < 0.0001 between Corpora A2 and B2; LL: -18.34, p-value: < 0.0001 between Corpora A2 and C2). The test results of negative LL scores indicate that Type B *now* is under-represented in the speech of Chinese NNSs.

Surprisingly, the case of *now* does not support my hypothesis that the more interactive the genre or type of activity is, the more DMs occur. Type B *now* is rather different from the other DMs under investigation in this thesis.

9.3.2 Collocates of *now*

The patterns of *now* in the six subsets of SECCL, MICASE and ICE-GB under investigation are shown Tables 9.2 to 9.7. The two patterns of the NNSs' monologues and dialogues (see Tables 9.2 and 9.3) reveal similar collocates. The collocates immediately to the left, *just*, *from*, *till*, *til*, *even*, *are*, *until* shown in boldface, are indications of the use of Type A *now*. In both patterns, the first pronoun to the right, *I*, is a strong collocate. Further investigation reveals that *I* is used with such verbs and modal verbs as *have*, *am*, *can* and *know*. These can be attributed to the instruction to the NNSs to talk about their personal experiences and opinions (see Appendix 1 for topics for the NNSs' monologues and dialogues).

Table 9.2: Pattern of *now* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (14)	i (17)	and (16)	and (28)	now (295)	i (109)	i (39)	i (17)	i (19)
2	i (14)	the (11)	um (16)	just (16)		eh (15)	still (17)	a (12)	to (15)
3	to (10)	to (10)	i (14)	eh (15)		and (15)	have (16)	the (10)	my (12)
4	in (10)	my (9)	the (7)	2 (13)		when (11)	am (12)	um (7)	the (11)
5	very (8)	so (8)	eh (6)	from (12)		we (10)	is (11)	still (7)	a (8)
6	a (8)	in (7)	he (6)	till (12)		on (9)	you (10)	eh (7)	in (8)
7	my (7)	a (7)	me (6)	but (11)		she (9)	can (9)	it (7)	is (6)
8	the (7)	and (7)	that (6)	til (10)		he (8)	at (8)	you (7)	this (6)
9	not (6)	that (7)	years (5)	even (9)		you (7)	the (7)	that (6)	you (6)
10	of (5)	was (5)	a (5)	until (8)		but (7)	are (6)	in (6)	now (6)

Table 9.3: Pattern of *now* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (20)	you (18)	you (28)	but (28)	now (300)	i (70)	have (25)	<i>i</i> (31)	you (13)
2	a (17)	i (16)	a (17)	just (22)		you (29)	i (17)	you (13)	i (12)
3	you (15)	a (15)	b (13)	and (14)		b (17)	are (13)	a (12)	the (11)
4	um (10)	are (12)	i (12)	know (11)		a (15)	you (13)	the (10)	to (10)
5	but (9)	b (11)	eh (9)	are (11)		we (13)	eh (8)	is (9)	can (8)
6	and (9)	the (9)	the (8)	eh (10)		the (11)	we (8)	have (9)	is (8)
7	of (9)	is (9)	think (7)	think (9)		because (9)	know (7)	think (7)	have (8)
8	to (8)	to (7)	and (6)	um (8)		eh (9)	the (7)	eh (7)	m (7)
9	b (7)	and (7)	college (5)	from (6)		so (7)	think (7)	um (7)	think (7)
10	what (6)	in (7)	it (5)	you (6)		and (6)	and (6)	in (6)	eh (6)

In the American NSs' speech in MICASE (see Tables 9.4 and 9.5), the first collocate immediately to the left is *right*. Further investigation confirms that most of the instances of *now* in *right now* are Type A, whereas the immediately left collocates, *but*, *okay* and *so* co-occur with Type B *now*.

Table 9.4: Pattern of *now* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (16)	the (28)	the (17)	right (14)	now (367)	the (24)	the (22)	to (17)	the (25)
2	to (12)	of (17)	of (12)	but (14)		in (22)	is (12)	the (14)	to (12)
3	in (12)	in (13)	a (8)	and (13)		this (14)	a (12)	that (12)	of (11)
4	of (10)	to (13)	this (7)	okay (12)		you (14)	we (10)	you (10)	that (9)
5	is (9)	a (11)	that (7)	so (8)		we (13)	to (8)	a (10)	a (8)
6	that (8)	that (9)	and (6)	you (7)		and (10)	of (8)	of (9)	at (7)
7	and (7)	uh (8)	cells (5)	are (6)		um (9)	i (8)	in (6)	and (6)
8	a (6)	and (8)	okay (5)	cells (6)		i (9)	have (7)	is (6)	on (6)
9	for (4)	these (6)	you (4)	is (5)		that (7)	you (7)	have (5)	in (6)
10	cells (4)	as (5)	they (4)	that (4)		if (7)	what (6)	if (5)	about (5)

Table 9.5: Pattern of *now* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	one (8)	the (9)	the (17)	right (31)	now (279)	you (21)	you (14)	the (13)	the (13)
2	the (7)	be (6)	a (5)	so (13)		i (19)	i (10)	to (9)	to (10)
3	a (6)	to (6)	is (5)	and (12)		that (13)	know (9)	that (8)	this (9)
4	you (6)	and (6)	um (4)	but (9)		what (11)	the (8)	you (7)	you (9)
5	it (5)	is (5)	it (4)	this (7)		we (11)	is (8)	know (6)	of (8)
6	and (4)	of (5)	like (4)	you (6)		it's (10)	this (7)	this (6)	a (7)
7	have (4)	in (4)	so (4)	um (6)		is (9)	have (7)	is (5)	do (5)
8	so (4)	we (4)	point (4)	now (5)		the (8)	can (7)	be (5)	about (5)
9	to (4)	so (4)	this (4)	like (5)		if (7)	about (6)	do (5)	is (4)
10	gonna (3)	i (4)	for (4)	square (5)		i'm (7)	just (5)	take (5)	it (4)

The collocates immediately to the left of *now* in Tables 9.6 and 9.7 (the British NSs' speech) are rather different from those in Tables 9.4 and 9.5 (the American NSs' speech). The collocates shown in boldface, *s* (the contraction of *is*), *is*, *back*, *by* and *right* suggest that *now* is Type A. In Table 9.7, about half the instances of the collocate immediately to the left *right* are DM *right*, which is part of the DM collocation *right now*. In Excerpt (9.3.1) below, both *right* and *now* are Type B, which are distinct from Type A *now* in *right now*.

(9.3.1)

B: Uhm how long do they go on for

C: Six days

B: **Right| Now|** you're going out t to France the end of that time is it or a month later

C: July the fifteenth

(ICE-GB: S1A-011)

Table 9.6: Pattern of *now* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (17)	the (18)	and (19)	and (27)	now (300)	the (22)	the (36)	the (17)	the (25)
2	and (13)	in (13)	the (18)	s (12)		as (9)	s (16)	to (15)	of (12)
3	in (8)	and (12)	it (13)	uhm (10)		i (9)	to (12)	that (8)	to (11)
4	to (8)	of (9)	to (10)	is (8)		that (8)	it (10)	and (8)	and (8)
5	that (7)	to (8)	he (9)	back (6)		to (8)	a (7)	a (8)	on (7)
6	uh (6)	this (6)	i (6)	are (6)		why (7)	is (5)	s (6)	a (7)
7	it (5)	but (5)	s (6)	it (5)		and (7)	are (5)	who (5)	uh (6)
8	two (5)	a (5)	in (5)	they (4)		in (6)	that (5)	of (5)	s (6)
9	of (4)	start (4)	a (4)	things (4)		with (6)	into (4)	uh (4)	is (6)
10	three (4)	it (4)	back (4)	by (4)		this (6)	ve (4)	think (4)	it (6)

Table 9.7: Pattern of *now* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	i (24)	s (13)	it (15)	it (21)	now (369)	i (54)	s (41)	i (17)	to (13)
2	s (14)	you (10)	a (12)	uhm (16)		you (17)	i (17)	you (16)	you (13)
3	to (11)	and (9)	the (10)	right (11)		it (16)	you (14)	that (10)	i (13)
4	it (9)	i (9)	i (9)	ok (10)		and (14)	ve (12)	s (10)	the (11)
5	you (9)	it (9)	of (9)	and (9)		yes (13)	it (12)	to (10)	s (9)
6	a (7)	that (8)	s (7)	well (9)		that (12)	the (10)	the (10)	that (9)
7	of (6)	on (7)	no (7)	is (7)		what (11)	m (9)	got (9)	it (7)
8	we (6)	m (7)	to (7)	s (7)		we (10)	think (9)	just (8)	of (6)
9	that (6)	a (7)	right (6)	yes (6)		he (9)	re (9)	it (7)	a (6)
10	oh (6)	uhm (7)	that (6)	you (6)		the (9)	was (7)	not (7)	and (6)

The patterns shown in Tables 9.2 to 9.7 reveal frequent collocates of *now*. To further investigate the use of Type B, the patterns of Type B *now* in the speech of the NNSs and NSs are produced and shown in Tables 9.8 to 9.13.

In the pattern of the NNSs' monologues (Table 9.8), there are no very frequent collocates, except the collocate 2, which refers to the beginning of a monologue. Utterance-initial *now* accounts for about 20% of the instances. This use is further discussed in the next section. The frequencies of collocates to the right/left are probably too low to be important.

Table 9.8: Pattern of Type B *now* in the non-native speakers' monologues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (5)	oral (3)	the (4)	2 (13)	now (64)	i (22)	i (9)	i (5)	to (5)
2	i (4)	after (3)	english (3)	eh (7)		let's (4)	d (3)	it (4)	in (4)
3	his (3)	i (2)	2 (3)	and (6)		eh (4)	study (3)	like (3)	now (4)
4	and (3)	time (2)	that (2)	because (3)		study (3)	you (2)	to (3)	is (3)
5	in (2)	quick (2)	very (2)	university (2)		um (3)	here (2)	english (3)	you (3)
6	eh (2)	hello (2)	uncle (2)	well (2)		there (2)	we (2)	the (2)	the (2)
7	aunt (2)	better (2)	eh (2)	you (2)		when (2)	ll (2)	this (2)	this (2)
8	a (2)	apologized (2)	decision (2)	um (2)		though (2)	continue (2)	want (2)	will (2)
9	do (2)	enter (2)	than (2)	but (2)		the (2)	can (2)	was (2)	but (2)
10	could (2)	cake (2)	for (2)	much (2)		do (2)	begin (2)	tell (2)	has (2)

In Table 9.9, the boldface collocates to the left, *a* and *b*, mostly refer to the identification of the two speakers in the dialogues. This use of *now* in the dialogues for opening a turn is discussed as a type of co-occurrence in the next section.

Further investigation in the patterns of the NNSs' dialogues identified such DM collocation as *but now* and *you know now*.

Table 9.9: Pattern of Type B *now* in the non-native speakers' dialogues in SECCL

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	of (4)	a (5)	a (9)	but (7)	now (53)	i (12)	have (6)	i (6)	i (4)
2	b (4)	the (3)	you (8)	eh (7)		you (6)	i (3)	is (4)	offered (3)
3	to (2)	yeah (3)	i (5)	think (5)		eh (4)	the (3)	um (3)	so (3)
4	with (2)	you (3)	b (3)	know (4)		our (3)	tuition (3)	been (3)	the (2)
5	yeah (2)	i (2)	um (2)	b (3)		when (3)	this (2)	you (2)	sophomores (2)
6	follow (2)	b (2)	yes (2)	um (3)		the (2)	they (2)	are (2)	you (2)
7	and (2)	and (2)	but (2)	yes (2)		we (2)	feel (2)	a (2)	very (2)
8	than (2)	dorms (2)	tape (2)	you (2)		first (2)	before (2)		can (2)
9		boys (2)	think (2)	see (2)		because (2)	see (2)		in (2)
10				because (2)		in (2)	our (2)		have (2)

In the patterns of the NSs' speech, shown in Tables 9.10 to 9.13, the collocates to the left *okay*, *but*, *so*, *and*, *well*, *right*, *yes* and *yeah* are mostly DMs. Although these DM collocations are not found in the NNSs' speech, where *but now* and *you know now* are identified, they show a preference for the order of DM collocations. *Now* in a DM collocation is usually preceded by other DMs.

Table 9.10: Pattern of Type B *now* in the native speakers' highly monologic discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	the (13)	the (17)	the (12)	okay (12)	now (203)	the (17)	is (10)	to (10)	the (15)
2	is (8)	of (11)	of (9)	but (10)		in (13)	the (8)	the (7)	to (10)
3	to (8)	a (9)	okay (5)	so (5)		you (13)	i (7)	that (7)	at (5)
4	of (7)	in (9)	a (5)	and (5)		this (12)	we (7)	you (7)	we (4)
5	in (6)	to (7)	that (4)	cells (4)		if (7)	are (5)	of (6)	a (4)
6	and (4)	that (5)	like (3)	metastase (3)		i (7)	first (4)	a (6)	that (4)
7	on (3)	and (4)	with (3)	this (3)		we (7)	might (4)	in (4)	of (4)
8	a (3)	one (4)	or (3)	you (3)		what (6)	you (4)	is (4)	in (3)
9	what (3)	uh (4)	for (3)	cell (3)		uh (5)	what (4)	say (4)	on (3)
10	for (3)	as (4)	an (3)	protease (2)		um (5)	of (4)	back (2)	it (3)

Table 9.11: Pattern of Type B *now* in the native speakers' highly interactive discourse mode in MICASE

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	one (6)	the (5)	the (8)	so (12)	now (134)	you (12)	you (8)	the (8)	you (8)
2	the (5)	is (3)	a (5)	but (7)		i (11)	the (6)	you (6)	the (8)
3	it (4)	so (3)	point (3)	um (6)		what (10)	is (5)	that (5)	to (6)
4	so (3)	be (3)	this (3)	and (5)		the (8)	can (5)	this (5)	about (5)
5	an (2)	and (3)	is (3)	square (5)		we (5)	about (5)	know (3)	is (4)
6	than (2)	above (2)	of (2)	okay (4)		this (4)	know (4)	is (3)	have (4)
7	two (2)	there's (2)	um (2)	well (3)		is (4)	do (4)	to (3)	of (3)
8	all (2)	was (2)	trees (2)	fine (2)		now (3)	this (4)	be (3)	i (3)
9	oh (2)	that's (2)	oh (2)	this (2)		that (3)	i (3)	a (3)	do (3)
10	cannot (2)	interesting (2)	okay (2)	yep (2)		we're (3)	going (3)	you're (3)	this (3)

Table 9.12: Pattern of Type B *now* in the native speakers' unscripted monologues in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	and (7)	in (8)	the (8)	uhm (9)	now (93)	the (9)	s (8)	to (5)	the (8)
2	not (3)	of (4)	to (4)	and (5)		why (7)	are (5)	i (4)	to (5)
3	it (3)	the (3)	it (3)	things (4)		i (6)	m (3)	this (3)	is (4)
4	to (3)	quite (3)	this (3)	road (2)		this (6)	that (3)	it (3)	s (4)
5	the (3)	s (2)	different (3)	three (2)		that (5)	should (3)	of (3)	a (4)
6	do (3)	sending (2)	writers (2)	right (2)		as (4)	just (3)	did (3)	uh (3)
7	these (2)	this (2)	with (2)	institution (2)		there (4)	you (3)	that (3)	digress (3)
8	as (2)	start (2)	twelve (2)	endocrinologists (2)		what (4)	the (3)	the (2)	this (2)
9	study (2)	to (2)	these (2)	october (2)		uh (4)	is (3)	was (2)	will (2)
10	two (2)	away (2)	and (2)	mouse (2)		if (3)	ve (2)	uh (2)	was (2)

Table 9.13: Pattern of Type B *now* in the native speakers' private direct conversations in ICE-GB

N	L4	L3	L2	L1	Centre	R1	R2	R3	R4
1	s (4)	on (4)	of (5)	uhm (11)	now (105)	i (16)	s (16)	you (8)	you (7)
2	of (3)	i (3)	right (5)	ok (7)		you (10)	ve (7)	that (6)	to (6)
3	where (3)	it (3)	uhm (4)	right (6)		that (8)	you (4)	got (4)	the (3)
4	to (3)	the (3)	it (3)	yes (6)		what (7)	the (4)	the (4)	know (3)
5	i (3)	and (3)	see (2)	well (5)		the (6)	are (3)	they (3)	a (3)
6	as (3)	of (2)	s (2)	and (3)		then (5)	m (3)	i (3)	this (2)
7	for (2)	that (2)	yeah (2)	yeah (3)		this (4)	would (3)	going (3)	not (2)
8	you (2)	can't (2)	you (2)	mm (3)		where (4)	was (2)	to (3)	up (2)
9	not (2)	a (2)	that (2)	it (3)		uhm (3)	uhm (2)	uh (2)	was (2)
10	know (2)	in (2)	there (2)	you (2)		they (3)	that (2)	this (2)	two (2)

9.4 Discourse aspects of *now*

The positions in an utterance/turn where Type B *now* occurs are first described and the linguistic items which Type B *now* tends to co-occur with are discussed. The investigation of the position and collocation phenomena leads to my interpretations of the uses of Type B *now* in the speech of the Chinese NNSs and American and British NSs.

All the occurrences of *now* in the NSs' highly monologic discourse mode in MICASE (367) and the private direct conversations in ICE-GB (372) were manually analysed and classified into Types A and B for further examination, but the incidence in the other four sub-corpora was unmanageable for manual analysis; therefore, as pointed out in the chapter of methodology (see Section 3.3.7), in the cases where the instances were more than 400, three sets of 100-line concordance samples were used.

9.4.1 Positions in an utterance/turn

In this section, the positions of Type B *now* in an utterance/turn are discussed. The distribution and proportion in the six sub-corpora under investigation are shown in Table 9.14 below.

In all the six sub-corpora, a large proportion of Type B *now* occurs in an extra-clausal position. Most of the instances appear in extra-clausal utterance/turn-medial position. There is no marked difference in the distribution of the positions of Type B *now* in an utterance/turn across the two types of genre and between the two groups of speakers. In the three monologic genres, the proportion of *now* in utterance-initial position in the NNSs' monologues is higher, because each monologue is treated as a single utterance. There are 1,143 monologues in the NNS data, as opposed to 13 and 70 texts in MICASE and ICE-GB respectively and therefore

there are more opportunities in the NNSs' monologues for *now* to occur in an utterance-initial position.

Table 9.14: Distribution of the positions in an utterance/turn of Type B *now*

Corpus	SECCL (NNSs): Monologues		MICASE (NSs): Highly monologic discourse mode		ICE-GB (NSs): Unscripted monologues	
	Random samples (times)	Percentage (%)	Raw freq. (times)	Percentage (%)	Raw freq. (times)	Percentage (%)
Positions in an utterance of Type B <i>now</i>	65	100	203	100	93	100
Extra-clausal: utterance-initial	15	23.1	1	0.5	0	0.0
Extra-clausal: utterance-medial	43	66.2	195	96.1	90	96.8
Extra-clausal: utterance-final	1	1.5	0	0.0	0	0.0
Intra-clausal: after an M-	2	3.1	3	1.5	2	2.2
Intra-clausal: after an MA	0	0.0	0	0.0	0	0.0
Intra-clausal: after an MF	2	3.1	0	0.0	0	0.0
Intra-clausal: others	2	3.1	4	2.0	1	1.1
Unclassified	0	0.0	0	0.0	0	0.0
Corpus	SECCL (NNSs): Dialogues		MICASE (NSs): Highly interactive discourse mode		ICE-GB (NSs): Private direct conversations	
	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)	Random samples (times)	Percentage (%)
Positions in a turn of Type B <i>now</i>	53	100	145	100	108	100
Extra-clausal: turn-initial	18	34.0	43	29.7	40	37.0
Extra-clausal: turn-medial	26	49.1	95	65.5	63	58.3
Extra-clausal: turn-final	0	0.0	0	0.0	2	1.9
Intra-clausal: after an M-	2	3.8	2	1.4	1	0.9
Intra-clausal: after an MA	0	0.0	0	0.0	0	0.0
Intra-clausal: after an MF	3	5.7	2	1.4	2	1.9
Intra-clausal: others	4	7.5	2	1.4	0	0.0
Unclassified	0	0.0	1	0.7	0	0.0

9.4.1.1 *Now* in extra-clausal position

In the NNSs' and NSs' speech under investigation, Type B *now* is predominantly placed in extra-clausal utterance/turn-medial position, as in Excerpt (9.4.1).

(9.4.1)

P: Utterance-medial my mother smiled<smelled> and told me: "I knew you... you are... I know you are you are... you are eager to... to need reference books, so I sell some wheat and um go to the town to buy these books for you, you must keep studying... um... we are poor now, but in the future, we will be better. [**Now**], at that time, I didn't know... I didn't know what to say, I just... I just tell... I just told myself, you must working... .."

(SECCL: B00-11-18)

9.4.1.2 *Now* in intra-clausal position

Across the six sub-corpora, a very low proportion, 6.8% on average, of the occurrences of Type B *now* is placed in an intra-clausal position, most frequently occurring after an M-element, as in Excerpt (9.4.2).

(9.4.2)

P: M- + *now* + +M remember in that article you read about, two, mouse cancer cell lines yeah, turn to the page in the course pack yeah right uh, this is |now| on page twenty-two... remember you read about, two cancer cell populations in mice.

(MICASE: LEL175SU106)

9.4.2 Contexts where Type B *now* tends to occur

The positions of Type B *now* in an utterance/turn described in the preceding section are referred to in the present discussion of the contexts where *now* tends to occur. The divisions in the discussion of *now* range from the types of co-occurrence for the purpose of the organisation of discourse to the micro-textual linguistic items. Some related types of co-occurrence are put in one category due to the low incidence of *now* as a DM.

Most instances of Type B *now* mark a start of change in topic. Sometimes it occurs at the opening or closing of a topic and sometimes it co-occurs with a question or a list of items, etc. This section could have discussed the use of Type B *now* in two broad categories, one related to topic shifting and the other not. However, in order to be consistent with previous analyses of other DMs, each type of co-occurrence is treated separately. In the following discussion, some types of co-occurrence suggest the same function of *now*.

Type B *now* is found to co-occur with the following eight categories: 1) opening/closing of topic and concluding remarks, 2) changes of topic and viewpoint, 3) questions, 4) elaborations, explanations and exemplifications, 5) listing items and sequences of events, 6) contrasting items, 7) emphatic lexical items and structure and 8) indications of location and object.

The instances in ambiguous contexts, with no linguistic evidence and insufficient contextual information remain unclassified in my analysis. A few instances of *now* are found to co-occur with hesitation markers and pauses, but the instances are too few to form a category; thus, these are grouped into the category of unclassified. Tables 9.15 to 9.20 at the end of this section illustrate the distribution of types of co-occurrence of Type B *now* in

relation to positions in an utterance/turn in the six sub-corpora.

9.4.2.1 *Now* occurring at the opening/closing of topic and concluding remarks

Now is used to (re)open or close the discourse or a topic (Biber *et al.* 1999: 1088, Carter and McCarthy 2006: 214). In Excerpt (9.4.3) below, the speaker uses *now* to begin a new topic.

(9.4.3)

P: Utterance-initial
E: Nothing relevant preceded; Item (1) is the beginning of a topic
F: To mark the opening of a topic

Task 2

Now⁽¹⁾ let me introduce an unusual teacher of mine to you. Miss Ben is my teacher in senior school.

(SECCL: B01-100-32)

This type of co-occurrence is frequent in the NNSs' speech, but not in the NSs' speech. It represents 33.8 and 13.2% in the NNSs' monologues and dialogues respectively, while in the NS corpora, it occurs occasionally, with only a few instances. It can be attributed to the number of texts in the corpus. In the NNS data, there are 1,143 texts in each sub-corpus, whereas in the NS data, there are fewer than 100 each. It is reasonable to suppose that there are more chances to begin a topic in the NNSs' monologues and dialogues than in the NSs' speech under investigation.

9.4.2.2 *Now* occurring at a shift of (sub)topic and viewpoint

It is found that *now* marks a shift of (sub)topic and viewpoint. In Excerpt (9.4.4), *now* prefaces Item (1), which is clearly irrelevant to the previous discourse. The speaker uses *now* to make a digression from the lecture to the film clip.

(9.4.4)

P: Utterance-medial
E: Item (1) separates the preceding discourse from the following discourse
F: To mark a shift of topic

..... although classically we've thought of you know the distance between uh uh between him and me is ten feet. well, it's ten feet in this frame of reference, but in other frames of reference it's something else. distance does not have, a definite value. it depends upon the frame of reference from which it's measured. same thing is true of timing. |**now**|, uh ⁽¹⁾ **i'm going to show you a brief little film clip**. it takes about five minutes. uh and uh, let me just before we uh, before i show it to you, uh let me just mention that in your text book,

(MICASE: LEL485JU097)

Now occurring at a transition is most frequent in the NS data, accounting for 46.3% and

29% of the instances of Type B *now* in the American NSs' highly monologic and highly interactive discourse modes in MICASE and 48.4% and 44.4% respectively in the British NSs' unscripted monologues and private direct conversations in ICE-GB. It is also frequent in the NNS data, with lower percentages, 29.2% and 13.2% of the instances of Type B *now* in the monologues and dialogues respectively. This type of co-occurrence suggests that *now* is used to mark a shift of topic, which has been discussed as a major function of *now* in the literature (see Section 9.2.4 above).

9.4.2.3 *Now* prefacing a question

It is found that *now* is used to preface a question when the question is also a change of (sub)topic. In the example below, Speaker A uses *now* before raising a question. This use suggests that *now* marks the boundary between the preceding utterance and the following question. Also, the use of *now* seems to make the speaker sound more confident and may draw the listener's attention, while the use of *well* could make a question sound more indirect. This use of *now* is often by speakers in authority, for example, Speaker A in Excerpt (9.4.5) is a doctor.

- (9.4.5)
- | | |
|--|--|
| <p>P: Turn-medial</p> <p>E: <i>Now</i> prefacing the question, Item (1)</p> <p>F: To mark a boundary and shift of topic</p> | <p>A: I'm going to give you a prescription to clear up the infection <> then you need to have your teeth extremely thoroughly cleaned <> as soon as possible Now ⁽¹⁾ when are you going off to Japan</p> <p>B: Right Well actually not for another month</p> |
|--|--|
- (ICE-GB: S1A-087)

Now prefacing a question is one of the least frequently-used types of co-occurrence in the NNS data, representing only 3.1% and 9.4% in the monologues and dialogues respectively. By contrast, it is frequent in the NS data, accounting for about 17% on average.

9.4.2.4 *Now* co-occurring with elaborations, explanations and exemplifications

All the instances of *now* co-occurring with elaborations, explanations and exemplifications are grouped into one category, since *now* with these types of co-occurrence seems to mark a boundary as well as serving as a continuer in discourse. In Excerpt (9.4.6), Item (2) is an elaboration of Item (1) and in Excerpt (9.4.7), Item (2) provides grounds for the speaker's opinion, Item (1). These two instances of *now*, placed between Items (1) and (2), seem to

serve as a continuer and mark the boundary.

(9.4.6)

P: Turn-medial
E: Item (2) elaborates
Item (1)
F: As a continuer to
elaborate

..... it appears that both of these families of proteases they do somewhat different things and ⁽¹⁾ **both of them are needed in order to get efficient, invasion by cancer cells through the stroma.** |now| ⁽²⁾ **once, these enzymes have allowed the cancer cells to digest a path through the stroma the, cancer cells can migrate away from the primary site of origin, until they encounter a blood vessel,**

(MICASE: LEL175SU106)

(9.4.7)

P: Turn-medial
E: Item (2) explains
Item (1)
F: As a continuer to
provide an
explanation

Task 3

A: I think ⁽¹⁾ **I think that of my friend should go to abroad,** |now| ⁽²⁾ **first the Chinese education system is different from the other country** so if you go abroad you can receive a different kind of education and meanwhile you can receive much new information you see our country is a develop developing country and when you come to a developed country.

(SECCL: C01-67-20)

Similar to the previous type of co-occurrence, this is one of the least frequently-used types of co-occurrence of *now* in the NNS data, but in the NS data, it represents slightly more than a quarter on average.

9.4.2.5 *Now* co-occurring with listing items and sequence of events

The four types of co-occurrence in this section and the following sections are not frequent in the NNSs' and NSs' speech. *Now* co-occurring with listing items and sequence of events seems to act as a device to separate the parts of the discourse. In Excerpt (9.4.8), *now* co-occurring with listing items can mark discursal segments and probably draw attention to the following point.

(9.4.8)

P: Utterance-medial
E: Item (1) is the first
item of the two
listing items
F: To separate units
and draw attention
to the following
point

..... you can <> bi combine those into a list and use that for sending mail <> Uhm <> |now| there are always two types of these things ⁽¹⁾ **One is common** <> for instance I already mentioned postgraduates <> uhm or those that are private the ones where you may <> uhm as I've said having ongoing conversations with people <>

(ICE-GB: S2A-028)

9.4.2.6 *Now* co-occurring with contrasting items

Similar to the use of the previous type of co-occurrence, *now* co-occurring with contrasting items, as shown in Excerpt (9.4.9), seems to draw the listener's attention. This use of *now* is

also a movement of topic.

(9.4.9)

P: Turn-medial

E: Items (1) and (2) are contrasting items

F: To draw attention to the following point

..... it demonstrates the pattern of what happens to hair, in the nineteenth century, the pattern of long hair to hair up on the head, at sixteen and then, a- at eighteen to ⁽¹⁾ **the hair up but also the corsetted body**, um which is a feature, of, adolescent growth and development in that time period. |**now**| if we jump ahead to the nineteen twenties, young women look quite different this is ⁽²⁾ **the bob**, you will all have seen pictures like this this is an anonymous American girl, uh F Scott Fitzgerald wrote about the bob and everybody seems to have done it it had a lot of symbolic value,

(MICASE: COL605MX039)

9.4.2.7 *Now* co-occurring with emphatic lexical items and structure

Now co-occurring with emphatic lexical items and structure is frequent only in the NNSs' dialogues, representing 26.4%. In the NS sub-corpora, it accounts for less than 5% each. In Excerpt (9.4.10), *now* co-occurs with the emphatic lexical item *very*. This can be interpreted to indicate that *now* is used to emphasise the following statement and to sound confident.

(9.4.10)

P: Turn-medial

E: Emphatic lexis
very

F: To emphasise the statement and sound confident

A: Um, my, one of my friends said... eh... .. he can during his part-time job, he enjoyed... eh... .. how... eh... .. to talk with many kinds of people and he at this let him know a lot of, a lot of rules in the society and he said he... eh... .. |**now**| he's known that it is **very** hard to earn money, and he said he was very grateful to her... eh... .. to his parents.

(SECCL: C96-13-14)

9.4.2.8 *Now* co-occurring with indications of location/object

It is found that *now* co-occurring with indications of location and object. In Excerpt (9.4.11) below, *now* precedes Item (1) *on page twenty-two*, which is the location to which the speaker is trying to draw attention.

(9.4.11)

P: M- + *now* + +M

E: Item (1) is an indication of location

F: To draw attention to the location

.....so i will pretty quickly, go over them with you and make sure you completely understand them. remember in that article you read about, two, mouse cancer cell lines yeah, turn to the page in the course pack yeah right uh, this is |**now**| ⁽¹⁾ **on page twenty-two**... remember you read about, two cancer cell populations in mice.

(MICASE: LEL175SU106)

Some instances of *now* co-occurring with indications of location and object are ambiguous. For instance, in Excerpt (9.4.12), *now* can be treated as an adverb. However, since it can be omitted without changing the proposition of the utterance, it is treated as a DM.

(9.4.12)

P: M- + *now* + +M

E: Item (1) is an indication of location

F: To draw attention to the location

..... some of the other genes that the Agrobacterium are expressing at that time, will coat the D, T-D-N-A so your T-D-N-A is |now|⁽¹⁾ in here, it's being coated by these vir genes,

(MICASE: LES405JG078)

9.4.2.9 Problematic and unclassified instances of *now*

As noted above, it is possible that more than one type of co-occurrence can be observed. In these cases, *now* is grouped in the category of stronger evidence. In Excerpt (9.4.13), *now* occurs at the shift of topic and it also co-occurs with emphatic lexis, *very*. Because the two speakers are about to end their conversation, the co-occurrence of shift of topic seems to be stronger and therefore this instance of *now* is categorised as a shift of (sub)topic and of viewpoint.

(9.4.13)

P: Turn-initial

E: Item (1) is a new topic; emphatic lexis *very*

F: Primarily for changing topic

A: If he stay in China. And the condition is not good. How... how can he improve himself?

B: I think in our China we have... we still have some famous professor <A: Professors.> I this area, and I think these professors can help Rob, do you think so? And... eh... the condition I don't think it is necessary thing for him, because the confidence and something else is very necessary.

A: But you know the condition is also very important.

B: I think we can still do a good job without a very good condition.

A: En... OK, OK. You are right maybe.

B: Thank you.

A: Eh... |**now**|⁽¹⁾ **I have a very important appointment.**

B: Appointment or a date?

A: Of course an appointment.

B: OK.

A: I must go now.

(SECCL: C01-01-22)

In the six sub-corpora under investigation, 10 instances of *now* in the NNSs' speech and 21 in the NSs' speech are not classified, because there is mostly not enough linguistic information to interpret the use of *now*.

In addition, a few instances of *now* co-occurring with hesitation markers and pauses are grouped in this category because they are too small in number of occurrences to form a separate category.

9.4.2.10 Summary of the contexts where *now* tends to occur

The types of co-occurrence which involve Type B *now* are discussed above and Tables 9.15 to 9.20 below illustrate the distribution of the positions in an utterance/turn of Type B *now* in the six sub-corpora under investigation.

As identified in the literature, one of the major functions of Type B *now* is segmenting the discourse, serving as a boundary marker. This function is supported by the first five categories of co-occurrence: 1) opening/closing of topic and concluding remarks, 2) shifts of (sub)topic and viewpoint, 3) questions, 4) elaborations, explanations and exemplifications and 5) listing items and sequence of events. *Now* co-occurring with these linguistic items indicates a clear boundary in discourse.

The instances of *now* in the categories of contrasting items, emphatic lexical items and structure and indications of location and object are interpreted as markers for drawing listeners' attention. This finding supports Fraser's (2009) argument that *now* is an attention marker.

Due to the variations in the types of activity in the six sub-corpora, it is anticipated that the distribution of the identified co-occurrence of *now* varies to some extent across sub-corpora. In the NNSs' monologues, the most frequent type of co-occurrence is opening/closing of topic and concluding remarks; in the dialogues, the most frequent one is emphatic lexical items and structure. These two types of co-occurrence are rarely found in the NSs' speech, in which the category of shifts of (sub)topic and viewpoint is highly represented.

Table 9.15: Distribution of co-occurrence of *now* as a discourse marker in SECCL: Monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Opening/closing of topic; concluding remarks	33.8	15	23.1	7	10.8										
2. Shifts of (sub)topic/viewpoint	29.2			18	27.7						1	1.5			
3. Questions	3.1			2	3.1										
4. Elaborations; explanations; exemplifications	10.8			6	9.2		1	1.5							
5. Listing items; sequence of events	0														
6. Contrasting items	4.6			3	4.6										
7. Emphatic lexical items and structure	9.2			3	4.6		1	1.5					2	3.1	
8. Indications of location/object	3.1			2	3.1										
Unclassified	6.2			2	3.1	1	1.5				1	1.5			
Occurrences: 65 out of 300 (random samples)	100.0	15	23.1	43	66.2	1	1.5	2	3.1		2	3.1	2	3.1	

Table 9.16: Distribution of co-occurrence of *now* as a discourse marker in SECCL: Dialogues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions									
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Opening/closing of topic; concluding remarks	13.2	4	7.5	3	5.7										
2. Shifts of (sub)topic/viewpoint	13.2	4	7.5	3	5.7										
3. Questions	9.4	3	5.7	2	3.8										
4. Elaborations; explanations; exemplifications	13.2	2	3.8	5	9.4										
5. Listing items; sequence of events	11.3	1	1.9	4	7.5								1	1.9	
6. Contrasting items	1.9		0.0	1	1.9										
7. Emphatic lexical items and structure	26.4	3	5.7	7	13.2		2	3.8					2	3.8	
8. Indications of location/object	0				0.0										
Unclassified	11.3	1	1.9	1	1.9						3	5.7	1	1.9	
Occurrences: 53 out of 300 (random samples)	100.0	18	34.0	26	49.1		2	3.8			3	5.7	4	7.5	

Table 9.17: Distribution of co-occurrence of *now* as a discourse marker in MICASE: Highly monologic discourse mode

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Utterance- initial	%	Utterance- medial	%	Utterance- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Opening/closing of topic; concluding remarks	2.0	1	0.5	3	1.5										
2. Shifts of (sub)topic/viewpoint	46.3			94	46.3										
3. Questions	8.9			17	8.4		1	0.5							
4. Elaborations; explanations; exemplifications	26.6			53	26.1								1	0.5	
5. Listing items; sequence of events	4.4			8	3.9		1	0.5							
6. Contrasting items	3.0			5	2.5								1	0.5	
7. Emphatic lexical items and structure	3.9			7	3.4								1	0.5	
8. Indications of location/object	4.4			8	3.9		1	0.5							
Unclassified	0.5													1	0.5
Occurrences: 203	100.0	1	0.5	195	96.1		3	1.5					4	2.0	

Table 9.18: Distribution of co-occurrence of *now* as a discourse marker in MICASE: Highly interactive discourse mode

Co-occurrence	%	Extra-clausal positions						Intra-clausal positions							
		Turn- initial	%	Turn- medial	%	Turn- final	%	after an M-	%	after an MA	%	after an MF	%	others	%
1. Opening/closing of topic; concluding remarks	0														
2. Shifts of (sub)topic/viewpoint	29.0	12	8.3	30	20.7										
3. Questions	26.2	16	11.0	22	15.2										
4. Elaborations; explanations; exemplifications	22.8	6	4.1	26	17.9								1	0.7	
5. Listing items; sequence of events	2.1	2	1.4	1	0.7										
6. Contrasting items	1.4		0.0	2	1.4										
7. Emphatic lexical items and structure	2.8	1	0.7	3	2.1										
8. Indications of location/object	9.7	3	2.1	8	5.5		2	1.4					1	0.7	
Unclassified	6.2	4	2.8	3	2.1						2	1.4			
Occurrences: 145 out of 300 (random samples)	100.0	44	30.3	95	65.5		2	1.4			2	1.4	2	1.4	

Table 9.19: Distribution of co-occurrence of *now* as a discourse marker in ICE-GB: Unscripted monologues

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Utterance-initial	%	Utterance-medial	%	Utterance-final	%	after an M-	%			after an MA	%	after an MF
1. Opening/closing of topic; concluding remarks	1.1			1	1.1									
2. Shifts of (sub)topic/viewpoint	48.4			45	48.4									
3. Questions	14.0			12	12.9			1	1.1					
4. Elaborations; explanations; exemplifications	19.4			18	19.4									
5. Listing items; sequence of events	2.2			2	2.2									
6. Contrasting items	3.2			2	2.2								1	1.1
7. Emphatic lexical items and structure	4.3			4	4.3									
8. Indications of location/object	5.4			5	5.4									
Unclassified	2.2			1	1.1			1	1.1					
Occurrences: 93	100.0			90	96.8			2	2.2				1	1.1

Table 9.20: Distribution of co-occurrence of *now* as a discourse marker in ICE-GB: Private direct conversations

Co-occurrence	%	Extra-clausal positions				Intra-clausal positions				others	%			
		Turn-initial	%	Turn-medial	%	Turn-final	%	after an M-	%			after an MA	%	after an MF
1. Opening/closing of topic; concluding remarks	2.8	3	2.8											
2. Shifts of (sub)topic/viewpoint	44.4	17	15.7	29	26.9						2	1.9		
3. Questions	19.4	8	7.4	13	12.0									
4. Elaborations; explanations; exemplifications	16.7	6	5.6	12	11.1									
5. Listing items; sequence of events	0.9			1	0.9									
6. Contrasting items	1.9			2	1.9									
7. Emphatic lexical items and structure	4.6	1	0.9	4	3.7									
8. Indications of location/object	0.9	1	0.9											
Unclassified	8.3	4	3.7	2	1.9	2	1.9	1	0.9					
Occurrences: 108 out of 300 (random samples)	100.0	40	37.0	63	58.3	2	1.9	1	0.9			2	1.9	

9.5 Chapter summary and conclusions

The frequency information and collocates of *now* are the starting points of the present research. The frequencies of *now* in the six sub-corpora reveal that there are more instances of Type B *now* in the monologic genres than in the dialogic genres, which makes Type B *now* a rather different DM. The analyses of the other DMs under investigation in this thesis show evidence for the hypothesis that the more interactive the genre or type of activity is, the more DMs occur.

Now is one of the peripheral DMs. The manual classification of Type A *now* and Type B *now* reveals that *now* is not primarily used as a DM by the NNSs and NSs under investigation. Manual examination found that the use of *now* is constrained by the types of activity. In the sub-corpus of the unscripted monologues in ICE-GB, *now* as an adverb is very common, in particular in sports commentaries, which spontaneously report what is happening in the field, thereby offering many opportunities for the use of Type A *now*.

The uses of *now* as a DM are discussed on the basis of its collocation phenomena in relation to the positions in an utterance/turn. In general, Type B *now* has a strong preference for extra-clausal positions. Most of the instances appear in extra-clausal utterance/turn-medial position in the six sub-corpora. There is no marked difference in the distribution of the positions in an utterance/turn of Type B *now* across the two types of genre and between the two groups of speakers. In contrast, there are marked differences in the distribution of the types of co-occurrence of *now* across sub-corpora. This can be attributed to the variations in the types of activity. For example, the NNSs' monologues are mainly narratives, in which there are more opportunities for using *now* to open the topic at the beginning of the discourse. It can be concluded here that the use of *now* depends on the context.

Another factor in the use of *now* as a DM is the role of the speaker, in particular in the classroom setting. In Sinclair and Coulthard's term (1975), *now* as a "frame" is normally used by speakers in power. This can be the reason why Type B *now* is highly represented in the sub-corpus of the NSs' monologic discourse mode in MICASE, which mainly consists of lectures, in which the speakers are mostly faculty members.

Similar conclusions are drawn in Aijmer's study (2002: 69). Based on the London-Lund Corpus, her study indicates that there is a link between the frequency of *now* and text types. *Now* tends to occur more often in formal types of activity (e.g. debates, interviews and public lectures) than everyday conversations. The roles of the speakers also influence the use of *now*

as a DM. Aijmer (2002: 95) concludes that *now* is a “marker of subjective modality”, for its close link to the speaker.

In conclusion, *now* as a peripheral DM has distinctive features. Unlike other DMs discussed in previous chapters, it occurs more often in the monologic genres than in the dialogic genres. Unlike such DMs as *like*, *well* and *you know*, *now* co-occurring with hesitation markers and pauses is rarely found. It is reasonable to assume that *now* is not used to signal that the speaker is searching for content or lexis and for holding/gaining the floor.

CHAPTER 10: TEXT-BASED ANALYSIS OF DISCOURSE MARKERS IN THE SPEECH OF THE NATIVE SPEAKERS

10.1 Introduction

The previous chapters report the investigation of corpora, from searching for relevant items to scrutinising their co-texts. This bottom-up approach finds common ground in the different contexts where DMs occur. Essentially, the corpus study brings evidence of the typical co-occurrence and distribution of DMs for analysis in the six sub-corpora under investigation.

The corpus-based investigation of the DMs shows that the Chinese NNSs and the NSs reveal different aspects of their use. For example, the DMs under investigation in the NSs' speech occur with a wider range of co-occurrence than is found in the NNSs' speech. However, this does not indicate that the NNSs overuse or underuse the DMs, because the contexts in the six sub-corpora vary widely and they are not comparable in terms of frequency. The corpus methodologies used in this study are unable to give adequate explanations for the under- or over-representation of DMs in a particular corpus.

Therefore, text-based analysis is employed to examine the DMs from a complementary standpoint in relation to larger linguistic contexts as well as situational contexts, if these are available. This chapter further explores the NSs' speech and the next chapter focuses on the NNSs' speech. Excerpts are particularly selected from the subsets of SECCL, MICASE and ICE-GB and analysed so that the contexts where DMs occur can be examined and all the DMs occurring in one text can be studied together, with the aim of seeing the connection between DMs and contexts.

The corpus analysis reveals that the type of activity and the degree of interactivity are key factors in using DMs. In addition, the initial investigation indicates that the DMs are unevenly distributed across texts in a corpus. High users and non-users of the DMs can be identified. In this chapter, a more qualitative text-based analysis seeks to explain how and why DMs occur more in one text than another.

10.2 Texts for analysis

Table 10.1 below lists the texts selected from the subsets of ICE-GB and MICASE. These eight excerpts represent two broad categories: monologic and dialogic genres. Excerpts 10.1, 10.2, 10.5 and 10.6 are monologic genres and the remainder are dialogic. In the next section,

the use of DMs is discussed in relation to the type of activity, participants and settings.

Table 10.1: Texts of the native speaker speech for analysis

Excerpt number	Type of activity	Participant	Setting	Source
10.1	Spontaneous rugby commentary on TV	A commentator	TV	ICE-GB: S2A-004 (Unscripted speech)
10.2	Interview	A photojournalist	Private home	ICE-GB: S2A-050 (Unscripted speech)
10.3	Private direct conversation	3 unknown speakers	N/A	ICE-GB: S1A-014 (Private direct conversation)
10.4	Private direct conversation	3 unknown speakers	Private flat	ICE-GB: S1A-038 (Private direct conversation)
10.5	Lecture on Physics	A senior faculty member and students	Classroom	MICASE: LEL485JU097 (Highly monologic discourse mode)
10.6	Lecture on Renaissance to Modern Art History	A senior faculty member and students	Classroom	MICASE: LEL320JU143 (Highly monologic discourse mode)
10.7	Anthropology of American Cities, discussion in office hours	An instructor (senior graduate) and student (senior undergraduate)	Coffee house	MICASE: OFC115SU060 (Highly interactive discourse mode)
10.8	Discussion by a Biochemistry study group	A peer group leader (senior undergraduate) and student (senior undergraduate)	Learning centre on campus	MICASE: SGR175SU123 (Highly interactive discourse mode)

10.3 Text-based analysis of discourse markers in the native speakers' speech

10.3.1 Unscripted speeches in the ICE-GB corpus

The types of activity in the sub-corpus of the unscripted monologues extracted from the ICE-GB corpus are various (for details of the texts, see Appendix 3). The unscripted monologues consist of demonstrations, legal presentations, spontaneous commentaries and unscripted speeches, in which the audience is unknown and/or unseen. Two texts are particularly selected to demonstrate how the frequency and use of DMs are affected by the type of activity. Excerpt 10.1 is chosen as an example of a monologue without frequent use of DMs. It also illustrates the distinctive features of spontaneous sports commentaries. The second text, Excerpt 10.2, is chosen for analysis on the basis of its having most occurrences of the DMs *oh*, *like*, *now* and *you see*.

Excerpt 10.1 below is a rugby sports commentary, which has only a few instances of DMs. The two instances of *oh* co-occur with emotional expressions and evaluations. There are not many other DMs in the whole text from which this excerpt is extracted. It can reasonably be argued that the speaker may give priority to fluency and may not be able to use DMs in

such a fast-paced activity because a commentary is expected to be purely descriptive. The engagement of listeners and the interaction between the speaker and listeners are not considered important, because listeners have a high motive for listening. They listen to the sports commentaries of their own accord. Furthermore, it is barely feasible to interact with listeners when the speaker is being broadcast. These features would explain the speaker's selective use of DMs.

DMs	Excerpt 10.1: NSs' monologue, unscripted speech; spontaneous commentary; Rugby league (ICE-GB)	Remarks
(1) and (2) co-occurring with emotions and evaluations	<p>..... Great Britain moving in very very quickly but that's a good run by Roach to Maninga <unclear-words> <,,> ⁽¹⁾ Oh good run <unclear-words> one man there <,> Good tackle by Schofield <,,> superb play by the Australians now that they want to move it fast from left to right they do <,> Elias <,> to Belcher <,> this is a good run from the Kangaroos <,> ⁽²⁾ Oh he's held on too long <,> held on too long <,,> and the fifth tackle <,,> Elias <,> doing a lot of kicking in midfield <,,> It's a good one It's beaten Gregory <,,></p>	(ICE-GB: S2A-004)

Excerpt 10.2 is an interview with frequent use of DMs *you know* and *you see*. The two instances of *you see* co-occur with key points and they are used on the early stage of the account. DM *you know* is constantly used by the speaker. The use of some instances (DMs (3), (6), (7) and (8)) is straightforward, co-occurring with emphatic lexis, reported speech and vague language. However, the use of some instances is found to be impossible to classify. It is difficult to say why these instances of *you know* and *so* (DMs (2), (9), (10), (11), (13) and (14)) are used. Considering the type of activity, interview, it can be clearly seen that the speaker as an interviewee uses DMs in his narrative account. However, the ICE-GB corpus, from which Excerpt 10.2 is extracted, does not provide enough contextual information for further discussion of the connection between the DMs and the context. There are no utterances from the interviewer and the first fifteen minutes of the recordings have not been used, in order to eliminate the self-conscious speech in the non-surreptitious recording method (Nelson 1996: 30-31). The utterances from the interviewer in Excerpt 10.2 might have helped to tell whether the interviewee responded to questions with DMs and if the interviewer used DMs to give feedback and acknowledge what was being said.

It seems that the speaker in Excerpt 10.2 produces the DMs with unidentified use to get the listener(s) involved in his account, although neither verbal nor non-verbal responses are provided in the corpus. The speaker is probably looking for a balance between fluency and the

engagement of listener(s), since in interviews there is a need for feedback from interviewers. Another possibility is that the speech is unscripted, so that the speaker uses DMs as delaying devices to formulate what to say next.

DMs	Excerpt 10.2: NSs' monologue, unscripted speech; at a private home in London (ICE-GB)	Remarks
	[A: a photojournalist] A: I went to the Israeli war in nineteen sixty-seven And what I did is I was in Cyprus I'd been to Egypt I came back to London on the Sunday <,> jumped on a plane and went to Cyprus on the Monday cos I'd heard the Israelis were bringing a plane to lift the press in ⁽¹⁾ you see for the Six Day War <,> They did bring the plane <,> When I arrived in the early hours of the morning in Israel from Cyprus I was really excited and I wanted to go And then the battle of Jerusalem was shaping up and there was the Sinai was virtually won ⁽²⁾ you know And I said to Murray Sayle I'm going to go to Jerusalem cos I thought ⁽³⁾ you know the Holy City is the most important aspect of the war Everybody else went south <,> Murray went south and I went north with a Sunday Times reporter There was a big team of us ⁽⁴⁾ you see <,> And I arrived in the early hours of the morning I'd got my pass from the Israelis <,> had a quick breakfast jumped in a car and went to Jerusalem with Colin Simpson of the Sunday Times <,> And when we got to Jerusalem the uhm Israeli paratroopers were marshalling They were grouping for the assault on the city They were below the old medieval walls <,> uhm near the Lions' Gate That was their objective <,> And uhm <,> ⁽⁵⁾ you know <,> they didn't know who I was what I was doing there Why are you here Who are you ⁽⁶⁾ You know And I said I'm from the press ⁽⁷⁾ Oh press They were kind of pleased ⁽⁸⁾ you know Cos once you get to the front line the soldiers are pleased to see you <,> because they want to know that the world knows their commitment ⁽⁹⁾ you know They want it to They want to know that it's not just confined to them and their imminent death or injury and so on They They want to know that other people know about their circumstance <,> But that's not what the the hierarchy wants <,> They don't care about that it seems <,> ⁽¹⁰⁾ So anyway we we got the were grouping there and I was about in the tomb of David <,> which all sounds amazing in the twentieth century ⁽¹¹⁾ you know <,> And uhm <,> the one thing I never needed in my life even though I got the ⁽¹²⁾ you know a lot of respect for some of them and not for a lot of them was a reporter <,> I used to get rid of reporters ⁽¹³⁾ you know cos I didn't want to feel as if I had to pay lip service to their <unclear-word> ⁽¹⁴⁾ so and I had my own work to do and my own drive	
(1) co-occurring with a key point		
(2) unclassified		
(3) co-occurring with emphatic lexis, <i>most</i>		
(4) co-occurring with a key point		
(5) co-occurring with hesitation marker, <i>um</i> and pauses		
(6), (7) co-occurring with reported speech		
(8) co-occurring with vague language, <i>kind of</i>		
(9), (10), (11), (13), (14) unclassified		
(12) co-occurring with a repair		
	(ICE-GB: S2A-050)	

Excerpts 10.1 and 10.2 are two distinctive kinds of monologue. The *discourse identities*¹⁹ (Zimmerman 1998: 90) of the speakers in Excerpts 10.1 and 10.2 are different. The sports commentator in Excerpt 10.1 is primarily the current speaker. There is almost no opportunity

¹⁹ Zimmerman (1998: 90) defines *discourse identities* thus: "Discourse identities are integral to the moment-by-moment organization of the interaction. Participants assume discourse identities as they engage in the various sequentially organized activities: current speaker, listener, story teller, story recipient, questioner, answerers, repair initiator, and so on".

for him to switch to the role of listener. The speaker in Excerpt 10.2, most of the time, is the current speaker and story-teller and sometimes his role may change to that of a respondent. In both cases, the speakers are aware of being recorded. The speaker in Excerpt 10.2, being an interviewee, probably prepared himself for the interview. Although it is unscripted, the speaker knows quite well what he is going to say. By contrast, the speaker in Excerpt 10.1 is giving a live sports commentary and has little idea of what is going to happen. However, the language used by a sports commentator is likely to be repetitive and formulaic. Additionally, while the interviewee in Excerpt 10.2 may expect feedback from listeners and feel driven to engage them at the expense of fluency, the commentator in Excerpt 10.1 does not anticipate receiving responses from listeners and speaks under strict time constraints. The characteristic of the identity of commentator is through spoken language to give listeners a simulated sense of being on the spot. These can explain why the speakers in Excerpts 10.1 and 10.2 employ different DMs surrounded by different kinds of co-occurrence and why the speaker in Excerpt 10.1 uses fewer DMs.

10.3.2 Private direct conversations in the ICE-GB corpus

The conversation component in the ICE-GB corpus includes private direct conversations between acquaintances, friends and colleagues. Excerpts 10.3 and 10.4 are extracted for analysis from two private direct conversations which include most occurrences of DMs *oh*, *like*, *now* and *you see*. They are chosen to demonstrate why there are more DMs in the dialogic genres than in the monologic genres. The DMs and their co-occurrence in Excerpts 10.3 and 10.4 are listed below, in the left-hand column and their uses are discussed in detail in the previous relevant chapters. I first discuss the speakers' identities in the discourse and comment on their use of *oh*, pointing towards the different uses in conversation from those in Excerpts 10.1 and 10.2 above.

The relationship of the speakers is unknown, but it is stated in the ICE-GB corpus that the speakers are acquaintances, friends or colleagues; hence, it is supposed that the relationship between them tends to be symmetrical. In the conversation in Excerpt 10.3, it can be inferred that the speakers have a mutual friend, Yanka. In Excerpt 10.4, Speaker A's utterance, *Is that why you came back in such a bad mood*, implies that the three speakers are likely to be flatmates. In both cases, the topics chosen are casual. The speakers have no

*situated identities*²⁰. The speakers' *discourse identities* constantly change from speakers to listeners and from inquirers to answerers. Their *transportable identities*²¹ are probably more influential in these everyday conversations. However, the limited information about the speakers provided by the corpus makes it harder to discuss this point further.

The uses of *oh* in the conversations, Excerpts 10.3 and 10.4, differ from those in the monologues, Excerpts 10.1 and 10.2 in the previous section. *Oh* co-occurs with emotions and evaluations in Excerpt 10.1 and with reported speech in Excerpt 10.2. In Excerpts 10.3 and 10.4, there are three speakers in each conversation and the floor is constantly negotiated and re-negotiated between speakers as the conversation goes along. It is evident that *oh* occurs turn-initially, prefacing responses in the private direct conversations. *Oh* with expletives, such as DMs (6) *oh my god* and (8) *oh god* in Excerpt 10.4, is more likely to occur in informal conversations. Nevertheless, the use of *oh* as a (preface to a) response in turn-initial position is context-sensitive and is often avoided in interviews or classroom settings. For instance, using *oh* as a (preface to a) response by either instructors or students in the office hour conversation is rather inappropriate. A conclusion is drawn here that genre and type of activity are factors in using DMs.

In Excerpt 10.4, *well* (DM (3)) prefacing a *dispreferred* response has not yet been discussed in the previous corpus investigation. According to the *preference system*, the *preferred* responses, i.e. acceptances and agreements, are usually given without hesitation and elaboration, while the *dispreferred* responses are the reverse (Pomerantz 1984: 72, Cameron 2001: 97). A similar type of co-occurrence identified in the analysis of *well* (Chapter 6) is disagreement and negative evaluation, which may be seen as *dispreferred* responses. In such cases, the use of prefatory *well* makes a *dispreferred* response less direct and less face-threatening. This could be seen as a way of following the politeness principles proposed by such researchers as Brown and Levinson (1987).

²⁰ Zimmerman (1998: 90) defines *situated identities* as follows: "Situated identities come into play within the precincts of particular types of situation. Indeed, such situations are effectively brought into being and sustained by participants engaging in activities and respecting agendas that display an orientation to, and an alignment of, particular identity sets, for example, in the case of emergency telephone calls, citizen-complainant and call-taker".

²¹ Zimmerman's definition (1998: 90) of *transportable identities* is: "[T]ransportable identities travel with individuals across situations and are potentially relevant in and for any situation and in and for any spate of interaction. They are latent identities that 'tag along' with individuals as they move through their daily routines..."

DMs	Excerpt 10.3: NSs' private direct conversation, reproduced from the Survey of English Usage (ICE-GB)	Remarks
(1) prefacing a new topic	A: I didn't see the news <,> this morning I just caught the end ⁽¹⁾ you know some Yugoslav fighting<,,>	Underlined: Yanka, a mutual friend
(2) prefacing the response to new information	B: ⁽²⁾ oh Talking about Yugoslavs I told you about that poster I gave to Vlad <,> He denied all knowledge of it A: Yeah Whose was it then B: I don't know Must be a Yugoslavian in the building But I don't know <,,>	
(3) as a response	A: ⁽³⁾ Oh <,,> B: I think perhaps <u>Yanka</u> 's been over there <,,> C: Which part's he from Do you know A: He's from Belgrade <,,> I can't remember where she's from <,> the other bit C: Presumably Presumably he's done his national service A: Yeah he did Yeah Yeah <,,> C: I can remember going there and being amazed how pimply the <,> the conscripts were <,> Incredibly young B: That's because they're young Yes It's the same in the Soviet Union C: When I went through Romania <,,> there were guards there as well They could be even younger there <,> It was about four years ago <,,> Baggy uniforms which looked like they had sort of one size or something to fit every recruit <,,> Old fashioned guns and all the stuff that <unclear-words> <,,> Mind you there were <unclear-words> that go with us B: Mhm Unfriendly C: Mhm A: Why B: Probably wouldn't 've been <unclear-words> C: I don't know really You get a visa for four days You catch a train You timed it all so that you think you'll arrive <,,> to catch your train on it and <unclear-words> And you arrive quite early in the morning and then march off this train at the border to get ⁽⁴⁾ you know You have to change your money You have to change about eight pounds for the day It's a lot of a lot of money But they make sure they sort of they seem to make sure that the processes take long enough for you to miss your train So you have to wait for hours <,,> A sort of manip ⁽⁵⁾ Well this was in Timisoara actually where we had to wait	
(4) as a replacement of something unsaid	A: In where C: Timisoara where we had to B: <unclear-words> to dilettante	
(5) co-occurring with a restart	C: And sort of riots started there ⁽⁶⁾ But ⁽⁷⁾ I mean my only recollection of it is sleeping in a wood for about four or five hours Rather idyllic	
(6), (7) co-occurring with elaborations	B: <unclear-words> C: ⁽⁸⁾ But it was just it was on the way to Bucharest and <,,> to the Black Sea ⁽⁹⁾ But uhm it was very uh	
(8), (9) co-occurring with emphatic lexis, <i>just</i> and <i>very</i>	B: Whereabouts in the Black Sea	

(ICE-GB: S1A-014)

DMs	Excerpt 10.4: NSs' private direct conversation, at a private flat (ICE-GB)	Remarks
(1) unclassified	A: These are peanutty <,> B: There you go <,> A: It's a very nice shirt by the way ⁽¹⁾ well B: It's a very old shirt A: It's very nice B: Thank you <,> A: What shit I'm writing <,>	
(2) co-occurring with negative comments	C: Are we supposed to be talking constructively B: You're supposed to be conversing A: ⁽²⁾ Well I couldn't care less what the hell we talk about Dani come on you're not doing anything start talking yeah <,>	
(3) prefacing a dispreferred response	C: ⁽³⁾ Well then you'll have to give me some sort of topic to talk about A: ⁽⁴⁾ Oh for God 's sake woman How was your day <,> C: <laugh> A: <laugh>	
(4), (6), (8) expressing emotions	C: It was OK How was yours A: ⁽⁵⁾ Oh it was all right <,> ⁽⁶⁾ Oh my god I went to have dinner with this girl called Kate who's on my course ⁽⁷⁾ I mean she's going to be the next Margaret Thatcher without <,> without a doubt C: <laugh>	
(5) as a response to a question	A: ⁽⁸⁾ Oh god She's British Canadian and she was just going on about how she'll have no problem getting a job no problem And she knew an ex-professor because she did some MSC in shipping trade and finance whatever and she's there saying she's going to get no problem ah She was making me sick and she was going to use her contacts here there and everywhere and it's like ⁽⁹⁾ oh Kate shut up <,> Silly girl <,> C: No no She sounds quite sensible actually	
(7) co-occurring with elaborations	A: ⁽¹⁰⁾ Oh Dani You pretend You try so hard at being this <,> ⁽¹¹⁾ like <,> ⁽¹²⁾ oh Must be the Jewish in you <,> C: Being what Being what	
(9) co-occurring with reported speech	A: This kind of a mercenary person <,> C: Maybe I am <,> A: ⁽¹³⁾ Well you've missed the yuppie boat <,> ⁽¹⁴⁾ Oh typical bloody yuppie <,> That's just so passé in the nineties <,> Can I count this as viewing banking as providing <,> ⁽¹⁵⁾ Well not much is being said here is it <,> Come on woman Say something <,>	
(10) as a preface to the response	C: I can never think of anything to say when I'm being being under stress ⁽¹⁶⁾ I mean come on <,> Never usually say much anyway You talk too much <,> A: ⁽¹⁷⁾ So how was baby-sitting <,>	
(11) co-occurring with pauses	C: I can't stand baby-sitting <,> Baby-sitting is just the pits honestly <,> The child cried constantly She wouldn't go to sleep she wouldn't eat she wouldn't uck she wouldn't do anything <,> She just crawled around the whole time being totally destructive pulling things off tables and just really <,> annoying me really <,> A: <u>Is that why you came back in such a bad mood</u> <,>	
(12) indicating a cognitive process has been done	
(13) unclassified		
(14) expressing emotions		
(15) co-occurring with negative comments		
(16) unclassified		
(17) prefacing a question		

(ICE-GB: S1A-038)

Underlined:
Implying
the
speakers
are
flatmates

10.3.3 Texts of highly monologic discourse mode in the MICASE corpus

Unlike the ICE-GB corpus, MICASE offers more contextual information, such as the roles of speakers, the relationship between speakers and settings. This information helps to interpret the uses of DMs.

In general, there are more instances of DMs in the sub-corpus of the NSs' highly interactive discourse mode than those in the highly monologic discourse mode in MICASE. This section looks at two excerpts selected from the highly monologic discourse mode in MICASE to analyse the use of the DMs and to discuss possible reasons why there are fewer instances of DMs in the highly monologic discourse mode than in the highly interactive discourse mode.

The sub-corpus of the highly monologic discourse mode in MICASE consists of 13 texts and 10 of them are lectures (see Appendix 2 for the fact sheet about this sub-corpus). Excerpts 10.5 and 10.6 are chosen because there are fewer instances of the DMs *oh, like, now* and *you see*. The corpus methodologies are used to identify the co-occurrence of DMs. The co-occurring linguistic items or phenomena are shown in the left-hand column of Excerpts 10.5 and 10.6.

In Excerpts 10.5 and 10.6, both speakers are constructing their institutional identities, senior faculty members. As they hold authority over the listeners in the classroom, the relationship between the speakers and listeners is primarily asymmetrical. The speakers are lecturing and maintaining the floor all the time.

As noted earlier, DMs facilitate interactional interpretations. They are inherently interactive. Therefore, it is not surprising to find fewer instances of DMs in the texts of the highly monologic discourse mode than those of the highly interactive discourse mode. Although there are fewer instances of DMs in monologues, they are still of use. In lectures, for example, DMs are used as boundary markers. Both lectures in Excerpts 10.5 and 10.6 begin with the DM *okay* and then a recap of the previous lectures.

In these two lengthy lectures, DMs as boundary markers are infrequently used, but it can be identified that the lectures are clearly signposted with longer sentential markers to indicate boundaries. In Excerpt 10.5, Item (A), *today we're going to talk about the theory of relativity*, Item (B) *today, we'll probably be talking mostly about the uh uh the space and time aspects of relativity and the basic ideas* and Item (C) *let me just give you a a forty-five second sketch of Einstein's early life*, are explicit lexical boundary items. In Excerpt 10.6, there is one explicit lexical boundary item *and now we're getting into some tricky territory*. Arguably, in lectures, using these explicit boundary markers is better than a simple DM for the process of understanding among the students.

DMs can contribute to interaction in discourse. The lectures in Excerpts 10.5 and 10.6

are among the activities in which interaction is feasible. However, the speakers, if they do not employ DMs, they use other devices to signal listeners' engagement, thereby increasing the degree of interactivity in the discourse. Five devices in written discourse are discussed by Hyland (2009); they are used to analyse the engagement strategies in Excerpts 10.5 and 10.6. Tables 10.2 and 10.3 below list the examples of the devices used in Excerpts 10.5 and 10.6.

The first type of device, interrogatives, is an explicit engagement feature, given that questions invite listeners to respond orally or mentally. This device is not used by the lecturer in Excerpt 10.5, but the second type of device, pronouns (circled in the excerpts), are most heavily used, the inclusive pronoun(s), *we (our)*, in particular. The pronoun, *we*, is used to bring in four different referents, discourse participants, people in general, *you* (i.e. the listeners) in discourse and *I* (i.e. the speaker). The use of *we* referring to all discourse participants and people in general seems to express solidarity; the use of *we* to replace the self-centred *I* and the use of the exclusive *you* help to position the speaker and listeners in the same group.

Evidently, the use of first pronoun *I (me)* is relatively scarce, with 4 instances, compared to the 17 instances of *we (our)*. It seems that the speaker avoids using *I* to point out his superior academic status and power. It is argued that the speaker uses *we* to construct an academic community with students as members.

The 8 instances of *you*, in this lecturing context, co-occur with the speaker's suggestions, reminders and assumptions about the listeners' competence, indicating the listeners' identity as students. These may alert the listeners to pay more attention to the ongoing utterance.

The third type of device, directives, can be seen as constructing power differences in the classroom. In Example 3-1, the co-occurrence of *I think, for you* and the *to*-clause suggest the speaker's authority. This instance of directives seems to show the authority of the speaker as a lecturer over the listeners as students.

The fourth device, references to shared knowledge, seems to address the listeners as if distinctions of power from knowledge and academic status do not exist and makes it easier to engage the listeners through (an assumption of) shared knowledge or experience. Examples 4-1, 4-2 and 4-3 are appeals to the listeners on the basis of a shared knowledge of physics. The speaker uses this device to claim disciplinary membership with the listeners. In particular, Type A *you know* in Example 4-2, *as you, i'm sure, know*, shows that the speaker seems to make reference to shared knowledge in order to create interaction. It is also possible that this

instance of *you know* makes the relationship between the speaker and the listeners more symmetrical, because the speaker assumes the listeners have equal knowledge.

The last device, asides and interruptions to the current argument, is used to step out and offer a personal comment, focusing on the interaction between speaker and listener(s) rather than the development of the proposition. Example 5-1 interrupts the lecture and Example 5-2 includes the speaker's evaluation, *interestingly enough*. The speaker uses these two instances to leave lecturing on the content subject and take up a commenting role, addressing himself to the listeners.

DMs	Excerpt 10.5: NSs' highly monologic discourse mode (MICASE)	Remarks
(1) boundary marker to open the utterance	<p><TITLE>Intro to Physics Lecture</TITLE> <WHO="S1" NSS="NS" ROLE="SF" SEX="M" AGE="4"> uh, ⁽¹⁾ okay uh, now that we've studied optics, um, we're, in a position to talk about, the two, theories which pretty much revolutionized physics in the twentieth century.</p>	
(2) co-occurring with exemplifications	<p>uh, what we've been studying up to now, we've we've pretty much worked out by the year nineteen hundred, and it's fine, for uh figuring out things like ⁽²⁾ you know what somebody's eyeglass prescription would be or how a generator works or so on. uh,</p>	
(3) co-occurring with hesitation marker, uh	<p>⁽³⁾ but uh, the theory of relativity, and quantum theory, together, have_ were both developed in the last hundred years and they really revolutionized the way we look at reality.</p>	
(4) indicating a self-repair	<p>uh, to some extent also they've they've had a, ⁽⁴⁾ well not just to some extent, to a large extent, they've had an influence on some of the_ they've made possible some of the modern inventions for example, uh our knowledge of how the atom the nucleus works would have been impossible without knowing, relativity and quantum theory. uh computers, transistors uh the invention of those, depend completely on quantum mechanics. without quantum mechanics we'd never, have um uh been able to build a computer. uh, in the time we have remaining we can't um um, go on to great detail, in these theories</p>	Device 3-1
(5) boundary marker for continuing	<p>but ³⁻¹ i think it's worth, uh it it it's very worthwhile for you to at least have a taste of what they're like. uh to get some idea sort of a working knowledge, of what the basic ideas are.</p> <p>⁽⁵⁾ and , ^(A) today we're going to talk about the theory of relativity and then again some more uh tomorrow. uh and then uh after that we'll be talking pretty much about quantum theory and the structure of the atom. um, relativity um... the reason, we think of it as being revolutionary, is the theory of relativity has changed the way we think, of space and time.</p>	Items (A) explicit lexical boundary items
(6) continuing	<p>⁴⁻¹ you may have heard of space-time, relativity and so forth. e- essentially what the nature of the distance between two points is with the time intervals, what that is. uh also uh, the theory, pointed out something that had not been suspected before, and that is that mass and energy are, just different aspects of the same thing. um, that's the famous equation, E equals M-C-squared.</p>	Device 4-1
(7) prefacing a concluding remark	<p>^(B) today, we'll probably be talking mostly about the uh uh the space and time aspects of relativity and the basic ideas, we won't have time to say too much about the equivalence of mass and energy.</p> <p>⁵⁻¹ you have a couple of homework problems for tomorrow, uh which are concerned with that ⁽⁶⁾ and ⁽⁷⁾ so i think after you have a chance to read the book and and try them uh, probably uh what i have to say perhaps will make a</p>	Items (B) explicit lexical boundary items

(8) emphasis	<p><i>little more sense.</i> um uh but <u>we</u> probably won't get there uh the mass and energy part of it today. um,</p> <p>⁴⁻² <u>as <u>you</u>, i'm sure, know, the author of the theory of relativity was Albert Einstein.</u> and actually Einstein um, was one of the leading, um developers of quantum theory as well. in fact, it might be just, as_ for a sense of historical perspective,</p> <p>^(C) let <u>me</u> just give <u>you</u> a a forty-five second sketch of Einstein's early life. he um got his PhD in Germany a little bit before the year nineteen hundred. but he couldn't get an academic job at a university. in those days instead of having slews of professors, each university had a professor, ⁽⁸⁾ <u>you know</u> of physics or something. and he couldn't get, couldn't get a job. so he ended up um as a patent ics- inspector in Switzerland. essentially what he did, um was uh people would send in patent applications and Einstein would look at the application and see if it made scientific sense. uh but, while he did this he continued his interest in, theoretical physics. and in the year nineteen-oh-five, when Einstein was twenty-six years old, he published five papers. and, of those five papers one was probably nothing special, but two of them, contained the entire special theory of relativity. a th- another one, um... suggested to, scientists a way to determine Avogadro's number, with precision which had never before been dreamed of. um, in those days Avogadro's number, the number of molecules in a mole, wasn't known to much better than a factor of two.</p>	Device 4-2
(9) prefacing a concluding remark	<p>⁴⁻³ <u>but Einstein wrote a paper uh, which involved which <u>you</u> may have heard in in chemistry. Brownian motion,</u> if <u>you</u> haven't fine, uh but essentially he suggested by studying this phenomenon, uh one could measure Avogadro's number, with great precision. and it turned out ten or fifteen years after he made this suggestion somebody else did it and won a Nobel Prize for physics just for doing that, for taking Einstein's suggestion. the remaining paper, was one in which Einstein invented the photon, the basic quantum (of) electromagnetic radiation, which was one of this was one of the key steps, uh in the development of quantum theory.</p> <p>⁽⁹⁾ <u>so</u> ⁵⁻² <u>in the year nineteen-oh-five, Einstein did quite a bit. and interestingly enough, um the he won the Nobel Prize for physics.</u> uh and he won the prize for his invention of the photon.</p>	Device 4-3
	(MICASE: LEL485JU097)	Device 5-2

Table 10.2: Devices signalling listeners' engagement in Excerpt 10.5

Devices	Examples in Excerpt 10.5
1. interrogatives	(Not found)
2. pronouns	<ul style="list-style-type: none"> ▪ <i>we</i>: referring to all discourse participants <ul style="list-style-type: none"> <i>uh, what [we] 've been studying up to now, ...</i> <i>... [we] 've [we] 've pretty much worked out by the year nineteen hundred...</i> <i>...in the time [we] have remaining...</i> ▪ <i>we</i>: referring to people in general <ul style="list-style-type: none"> <i>...they really revolutionized the way [we] look at reality.</i> <i>... [our] knowledge of how the atom the nucleus works would have been impossible without knowing, relativity and quantum theory.</i> <i>without quantum mechanics [we] 'd never, have um uh been able to build a computer.</i> <i>...the reason, [we] think of it as being revolutionary, is the theory of relativity has changed the way [we] think, of space and time...</i> ▪ <i>we</i>: referring to you in the discourse <ul style="list-style-type: none"> <i>...now that [we] 've studied optics...</i> ▪ <i>we</i>: referring to I in the discourse <ul style="list-style-type: none"> <i>... [we] 're, in a position to talk about, the two, theories...</i> <i>... [we] can't um um, go on to great detail, ...</i> <i>today [we] 're going to talk about the theory of relativity...</i> <i>... [we] 'll be talking pretty much about quantum theory...</i> <i>today, [we] 'll probably be talking mostly about the uh uh the space and time aspects of relativity and the basic ideas,</i> <i>[we] won't have time to say too much about the equivalence of mass and energy.</i> <i>but [we] probably won't get there uh the mass and energy part of it today.</i> ▪ <i>I</i>: <ul style="list-style-type: none"> <i>... [i] think it's worth, uh it it it's very worthwhile for you to at least have a taste of what they're like.</i> <i>so [i i] think after you have a chance to read the book...</i> <i>...what [i] have to say perhaps will make a little more sense.</i> <i>let [me] just give you a a forty-five second sketch of Einstein's early life.</i> ▪ <i>you</i>: <ul style="list-style-type: none"> <i>...i think it's worth, uh it it it's very worthwhile for [you] to at least have a taste of what they're like.</i> <i>[you] may have heard of space-time, relativity and so forth.</i> <i>[you] have a couple of homework problems for tomorrow,</i> <i>so [i i] think after [you] have a chance to read the book...</i> <i>as [you], i'm sure, know, the author of the theory of relativity was Albert Einstein.</i>

	<i>let me just give you a forty-five second sketch of Einstein's early life.</i>
	<i>but Einstein wrote a paper uh, which involved _ which you may have heard in in chemistry, Brownian motion, if you haven't fine, ...</i>
3. directives	³⁻¹ <u><i>i think it's worth, uh it it it's very worthwhile for you to at least have a taste of what they're like.</i></u>
4. references to shared knowledge	⁴⁻¹ <u><i>you may have heard of space-time, relativity and so forth.</i></u> ⁴⁻² <u><i>as you, i'm sure, know, the author of the theory of relativity was Albert Einstein.</i></u> ⁴⁻³ <u><i>but Einstein wrote a paper uh, which involved _ which you may have heard in in chemistry, Brownian motion</i></u>
5. asides addressed to the listeners	⁵⁻¹ <u><i>you have a couple of homework problems for tomorrow, uh which are concerned with that ⁽⁶⁾ and ⁽⁷⁾ so i i think after you have a chance to read the book and and try them uh, probably uh what i have to say perhaps will make a little more sense.</i></u> ⁵⁻² <u><i>in the year nineteen-oh-five, Einstein did quite a bit. and interestingly enough, um the he won the Nobel Prize for physics.</i></u>

Similarly, the speaker in Excerpt 10.6 does not employ many DMs in the lecture. She rather uses various signals of engagement. Two instances of interrogatives, Examples 1-1 and 1-2, are used to engage the listeners in responding mentally. As in Excerpt 10.5, the inclusive pronoun, *we*, is relatively frequently used. The 2 instances of *we* referring to all the discourse participants demonstrate solidarity between the lecturer and students. The 6 instances of *we* referring to the speaker herself contribute to building the sense of membership, whereas the first pronoun, *I*, is only used once. (See Table 10.3 for a list of the devices signalling listeners' engagement in Excerpt 10.6.)

Unlike the co-occurrence of *you* in Excerpt 10.5, the 6 out of 8 instances of *you* in Excerpt 10.6 are used to direct the listeners' attention to paintings, which are the subject matter in the lecture.

The third type of device, consisting of directives, and the last one, asides addressed to the listeners, are not used by the speaker. There is one instance of engagement strategy, appeals to shared knowledge, as shown in Example 4-1.

DMs	Excerpt 10.6: NSs' highly monologic discourse mode (MICASE)	Remarks
	<TITLE>Renaissance to Modern Art History Lecture</TITLE>	
	<WHO="S1" NSS="NS" ROLE="SF" SEX="F" AGE="3">	
(1) boundary marker to open the utterance	(1) okay we ended the class last time talking about Courbet's painting The Real Allegory . and we talked about that paradox how could you have a real, allegory. and we talked about Courbet, in the category of realism and this is an ism that, really was used by artists at the time (we 've) discussed Courbet issuing a realist manifesto, um at the time, of his exhibition in eighteen fifty-five.	
(2) co-occurring with emphatic lexis, <i>very</i>	(2) now we discussed Courbet as v- being very self-conscious about what he was doing with art self-consciously a modern artist. um setting himself up, against the past in many ways. he's rejecting artistic institutions he's challenging artistic institutions by exhibiting outside the salon system. he's creating a persona around himself, of the artist as outsider the artist as revolutionary the artist as radical, the artist as somewhat bohemian but the artist also aligned with movements of liberation and with the working class. when he talks about real, it means on several levels at once. he talks about the need to draw his subject matter from real things that's the most obvious.	
(3) co-occurring with emphatic lexis, <i>most</i> and negation	(3) you know he's not going to paint an angel if he hasn't seen one he's going to base his subject matter, on things in his own world that he could have seen. not ancient history, not literature, real means not ideal he's not going to romanticize mythologize glamorize um, real also means not imaginary. his art isn't going to be about fantasies, it's not going to appeal to the imagination it's supposed to be about the here and now. moreover for Courbet, real tends to be drawn from lower social classes. there's no reason why poor people are supposed to be more real, um than the artist's own friends,	
(4) co-occurring with a restart	or, you could make (4) you know supposedly people who buy art are real too. but at this time in the eighteen thirties and forties, um, he's working with an idea that contemporary middle-class and upper-class society is false, and that the real, the real people are the working people.	
(5) as a boundary marker	so ¹⁻¹ <u>is there other ways, um that realism signifies_ realism means art without illusions.</u>	Device 1-1
(6) co-occurring with negation	(5) okay ^(A) and now we 're getting into some tricky territory, because the real allegory as we discussed, (6) you know wasn't based on anything he could have seen exactly. an allegorical figure is supposed to be a figure that stands for an ideal. now we 've seen how he deflates the idea of an allegory making the real allegory in this painting by taking the figure of a nude and saying no she doesn't stand for truth or beauty, if she's in this painting it must be because she's an artist's model and here you get a bunch of discarded clothing on the floor. usually when you have, (7) you know nymph and shepherd and- shepherd in a landscape or, naked figure of truth you don't get truth's underwear, there in front of you . alright and so this is one way he makes an allegorical figure real and yet this is paradoxical because if she really is an artist's model which is one way that you would see a nude in a contemporary nineteenth century Parisian context.	Item (A) explicit lexical boundary items
(7) co-occurring with key points, <i>nymph</i> and <i>shepherd</i>	¹⁻² <u>if she is an artist's model why isn't he painting her?</u> um he's making a painting of a landscape here. so these are more ways he's bringing paradox into this work suggesting that it's not exactly a realist picture but nor is it an allegorical one. it's his version of a modern real allegory fit for the nineteenth century. <PAUSE DUR=":08" /> but there's another way that Courbet's art can be described as real also, and that has to do with his painting technique and that's why i 've brought along the Burial at Ornans from eighteen forty-nine on the left and a close-up of it on the right. and	Device 1-2
	⁴⁻¹ <u>you'll recall some of the comments that people made at the time about how Courbet applied his paint about his technique for painting.</u> um he was criticized for not having any of the conventional technique.	Device 4-1
	(MICASE: LEL320JU143)	

Table 10.3: Devices signalling listeners' engagement in Excerpt 10.6

Devices	Examples in Excerpt 10.6
1. interrogatives	¹⁻¹ <u>is there other ways, um that realism signifies realism means art without illusions.</u> ¹⁻² <u>if she is an artist's model why isn't he painting her?</u>
2. pronouns	<ul style="list-style-type: none"> ▪ we: referring to all discourse participants ...and now we're getting into some tricky territory, ... now we've seen how he deflates the idea of an allegory making the real allegory... ▪ we: referring to people in general (Not found) ▪ we: referring to <i>you</i> in the discourse (Not found) ▪ we: referring to <i>I</i> in the discourse ... we ended the class last time talking about Courbet's painting <i>The Real Allegory</i>. ... we talked about that paradox how could you have a real, allegory. ... we talked about Courbet, in the category of realism ... this is an ism that, really was used by artists at the time (we've) discussed Courbet issuing a realist manifesto, now we discussed Courbet as v- being very self-conscious about ... because the real allegory as we discussed, ⁽⁶⁾ you know wasn't based on anything he could have seen exactly. ▪ I: ...that has to do with his painting technique and that's why I've brought along the <i>Burial at Ornans</i> from eighteen forty-nine on the left... ▪ you: ...we talked about that paradox how could you have a real, allegory. or, you could make ⁽⁴⁾ you know supposedly people who buy art are real too. ...here you get a bunch of discarded clothing on the floor. ...usually when you have, ⁽⁷⁾ you know nymph and shepherd and- shepherd in a landscape or, naked figure of truth you don't get truth's underwear, there in front of you. ...if she really is an artist's model which is one way that you would see a nude in a contemporary nineteenth century Parisian context. ...you'll recall some of the comments that people made at the time...
3. directives	(Not found)
4. references to shared knowledge	⁴⁻¹ <u>you'll recall some of the comments that people made at the time about how Courbet applied his paint about his technique for painting.</u>
5. asides addressed to the listeners	(Not found)

The analyses of Excerpts 10.5 and 10.6 indicate that DMs are not the primary means of increasing interaction in lectures. Possible alternatives include the use of lexical items as boundary markers and the devices signalling listeners' engagement. In this analysis of engagement strategies, how speakers engage their listeners and construct their identities has been uncovered.

From the choices of engagement strategies, it is inferred that both lecturers in Excerpts 10.5 and 10.6 avoid conveying their academic authority and try to express their solidarity with the students. The lecturers rarely use interrogatives and directives to engage listeners. While explicit questions and directives tend to add an interactional dimension, these devices imply the speaker's authority based on knowledge and suggest that the speaker is in control of both the speech and audience. Moreover, the use of the inclusive pronoun, *we*, to replace of the first person singular pronoun *I* and the appeals to shared knowledge reduce the distance in status and power.

In the analyses of *you know* and *I mean* (Chapter 7), an example of a high user is found in the NSs' speech of highly monologic discourse mode. A closer look at the distribution of *you know* and *I mean* in the 14 texts in this sub-corpus reveals an unusual text. More than half the occurrences of *you know* (225 out of 388) and *I mean* (64 out of 130) are from the text, LES495JU063, which is a lecture, produced by a senior graduate with near-native proficiency. Referring to Excerpt (10.3.1) below (the same as Excerpt (7.3.1) in Chapter 7 and repeated here for ease of reading), with *you know* and *I mean* in bold face, it can be interpreted that the use of *you know* and *I mean* could be an idiosyncratic feature of the speaker's speech style and that frequently using *you know* and *I mean* seems not to help construct a credible academic identity. The reason is that the use of DMs *you know* and *I mean* on such occasions as lectures, where fluency is expected, is rightly criticised and the overuse of these two DMs in contexts which lack coherence may hinder the progress of understanding for the listeners. By contrast, the sentential boundary markers and engagement strategies used by the lecturers in Excerpts 10.5 and 10.6 are appropriate in the classroom context.

(10.3.1)

- (1) but see what we ha- what we have here, is **|you know|** ultimately **|you know|** long term animosity,
- (2) and, **|you know|** a lot of people **|you know|** if you if you ask them in Eastern Europe they'll, No main clause;
with illegitimate
end
- (3) these days **|i mean|** they're not very keen, on remembering either of course,
- (4) **|i mean|** those who were older of course they're not very keen on on remembering German domination, but they're ultimately the also not very keen on when it comes well,
- (5) **|you know|** um did you like the Soviet troops there?
- (6) mkay. so, even these days, and of course **|you know|**, if you think about it this long term in- animosity, and anxiety and fear installed, No main clause
- (7) **|i mean|**, if that wasn't there, if if the Soviets had established some **|you know|** common interest, more on legitimacy rather than based on force, f- **|you know|**, the the the the Poles, the Hungarians and the Czechs these days, or what is it a year ago or two years ago? it was last year when they were admitted to NATO? **|i mean|** that's, why, that's part of the reason why Eastern European countries these days are now very, uh **|you know|**, are very vociferous in asking for, membership in NATO and the European Union.
(MICASE: LES495JU063)

I have dwelt in some detail on Excerpts 10.5 and 10.6 because they are typical texts in the sub-corpus of the highly monologic discourse mode in MICASE. Looking at the broader context and analysing discoursal features help to explain why there are fewer DMs in the texts of this subset than those of the highly interactive discourse mode. This text-based analysis complements the corpus investigation of DMs, as the latter offers quantitative information and the former helps to look into the phenomena of the non-use and low/high use of DMs.

10.3.4 Texts of highly interactive discourse mode in the MICASE corpus

In the sub-corpus of the NSs' highly interactive discourse mode, two excerpts are selected on the basis of the frequencies of the DMs, *oh*, *like*, *now* and *you see*. Excerpts 10.7 and 10.8 have the highest incidence of these DMs.

Apparently, there are more instances of DMs in Excerpts 10.7 and 10.8 than in Excerpts 10.5 and 10.6 in the previous section. In reference to the contexts, four interpretations are put forward to explain why the speakers in Excerpts 10.7 and 10.8 use more DMs. First, the types of activity, for example, the office hours in Excerpt 10.7 and the study group discussion in Excerpt 10.8, are less likely to have been prepared beforehand. This may be the reason why the speakers employ DMs to search for lexis or content information and to indicate a restart and repair.

Another factor in the frequent use of DMs can be the number of speakers. Excerpts 10.7 and 10.8 involve more than one speaker and a mixture of students, faculty and staff members.

The involvement of several speakers and frequent turn-taking can lead to the use of *oh* as a response and *well* and *you know* with varied types of co-occurrence; therefore, this increases the total of DMs.

Third, it appears from the contrast between Excerpts 10.5 and 10.6 (texts of highly monologic discourse mode) and Excerpts 10.7 and 10.8 (texts of highly interactive discourse mode) that DMs are used in symmetrical conversations rather than asymmetrical ones. The primary identities of the interlocutors in Excerpts 10.5, 10.6 and 10.7 are either teachers or students. The teacher-student relationship between speakers is not always stable. The more symmetrical relationship built up through interaction is influential. For example, in Excerpt 10.7, both the instructor (S2) and the student (S1) use *you know* with varied types of co-occurrence and the functions of *you know* can be identified on the basis of the co-occurrences in its immediate contexts. However, in a sense, the instructor with her institutional identity probably uses *you know* (4 times) to demonstrate solidarity, to reduce the difference of status and to downplay her authority in this type of activity, i.e. discussion in office hours. As the student also uses 4 instances of *you know*, a less asymmetrical interaction is created. Other evidence that the instructor is downplaying her authority is that she constantly gives the floor to the student rather than holding the floor all the time, which does not often happen in such lectures as Excerpts 10.5 and 10.6.

Fourth, using DMs can be a way of expressing solidarity and establishing and maintaining rapport. In Excerpt 10.8, both of the two students (one is a peer group leader) use *like* as a DM. It is not possible to identify the functions of all the uses. It seems that the use of *like* has little connection with the proposition in the discourse, but is used to construct the speakers' personae. The two speakers use *like* to show solidarity with the peer students. *Like*, on the whole, is used at present as an in-group marker. At the same time, the speakers present the identity of a competent student in biochemistry. In this context, the two speakers make some changes in what Sacks (1992: 327-328) calls *operative identities*²². Take Speaker 1, for example. She makes some changes of identity as the conversation proceeds. Utterances (A) and (C), shown in Excerpt 10.8, are academic terms, suggesting that Speaker 1 is presenting the identity of a competent student or group leader. Utterances (B) and (D) display Speaker 1 in her role of a group member by employing *like*, as Speaker 2 does. The identities of Speaker 1 are dynamic rather than fixed.

²² Sacks (1991: 327-8) argues that speakers have *operative identities*, which are the identities they have in the world and are not employed at the beginning of the discourse.

DMs	Excerpt 10.7: NSs' highly interactive discourse mode (MICASE)	Remarks
(1) opening a new topic	<TITLE>Anthropology of American Cities Office Hours</TITLE> <WHO="S1" LANG="NS" ROLE="SU, Student" SEX="M" AGE="1" > <WHO="S2" LANG="NS" ROLE="SG, Instructor" SEX="F" AGE="2" >	
(2) indicating a cognitive process	S1: nice to meet you Mr Letterman S2: alright	
(3) unclassified	S2: ⁽¹⁾ so how's the paper? S1: it's good actually i um	
(4) unclassified	S2: excellent S1: i talked to, the day after i talked to you, outside of the building	
(5) co-occurring with reported speech	S2: uhuh i talked to the, got in contact with the mall manager S2: <EVENT DESC="LAUGH" /> ⁽²⁾ ooh ⁽³⁾ so cuz i asked um, when i was at Twelve Oaks, ⁽⁴⁾ you know i i went to the security department and i said ⁽⁵⁾ you know can i ask you a few questions and she's like no, i was like okay ⁽⁶⁾ well . um, she says you ⁽⁷⁾ you know , you have to get permission from the mall manager and she gave me	
(6) unclassified		
(7) co-occurring with reported speech		
(8) as a preface to the response	S2: <OVERLAP2> ⁽⁸⁾ oh</OVERLAP2> <OVERLAP1>her</OVERLAP1> number so interesting i was able to get in touch with her.	
(9) as a preface to the response to a question	S2: and uh, what did you guys talk about? Wow S1: ⁽⁹⁾ well i asked her ⁽¹⁰⁾ like about whoa, don't wanna do that, um, how the security ⁽¹¹⁾ you know ⁽¹²⁾ like what are th- ⁽¹³⁾ you know what what they have to deal with the most and	
(10), (14), (15), co-occurring with exemplifications	S2: uhuh S1: average age of employees ⁽¹⁴⁾ you know all kinds of general stuff like S2: good	
(11), (12), (13) prefacing a restart	S1: ⁽¹⁵⁾ you know what sh- how she feels about, surrounding malls how they've affected Twelve Oaks S2: great and she she talked my ear off, <OVERLAP1>so</OVERLAP1> <EVENT DESC="LAUGH" />	
(16) co-occurring with approximations	S2: <OVERLAP1>fantastic.</OVERLAP1> and is that already incorporated into this paper? <OVERLAP1>or, okay</OVERLAP1> S1: <OVERLAP1>yeah, yeah</OVERLAP1> that's in there. i um S2: great was it Tues- Tuesday night? yeah i just went to work. i spent ⁽¹⁶⁾ like just hours and hours, <OVERLAP1>so</OVERLAP1> <OVERLAP1>great</OVERLAP1> great ⁽¹⁷⁾ well it looks like you have a full draft um, why don't you tell me what you think the argument of the paper is now,	
(17) co-occurring with expressions of uncertainty, looks like	S1: okay and uh... first of all what questions you ha- like, things you want me to look for in a way	
(18) unclassified		
(19) co-occurring with expressions of uncertainty, about	S1: um okay ⁽¹⁸⁾ well it's about ⁽¹⁹⁾ like malls moving into wealthy areas S2: <OVERLAP2>okay</OVERLAP2> <OVERLAP1>and</OVERLAP1> um... and where they do like the_ changes the socialization like people, uh, ⁽²⁰⁾ you know in those areas will come to the mall to socialize instead of, ⁽²¹⁾ like i compared it to Howell,	
(20) prefacing elaborations	(MICASE: OFC115SU060)
(21) co-occurring with exemplifications		

DMs	Excerpt 10.8: NSs' highly interactive discourse mode (MICASE)	Remarks
	<TITLE>Biochemistry Study Group</TITLE> <WHO="S1" LANG="NS" ROLE="SU, Peer Group Leader" SEX="F" AGE="1" /> <WHO="S2" LANG="NS" ROLE="SU, Student" SEX="M" AGE="1" />	
(1) as a preface to the response to new information	S1: sorry we could only get the old book. S2: ⁽¹⁾ oh that's okay.	
(2) prefacing a new topic	S1: okay. <PAUSE DUR=":09" /> ⁽²⁾ so, we have questions on electron transport or what?	
(3) unclassified	S2: um, you kn- that oxyl acetate, you know that, there's ⁽³⁾ like a, a weird cycle. it's, it sorta seems <OVERLAP1>separate</OVERLAP1>	Utterance (A) identity of a competent student
(4) unclassified	S1: ^(A) <OVERLAP1> citric</OVERLAP1> acid cycle? S2: no, it's something, sort of related to it but it's not, it.	
(5), (6) expressing emotions	S1: ^(B) okay ⁽⁴⁾ like , i was reading in the book. it's like ⁽⁵⁾ oh ⁽⁶⁾ oh , S2: <OVERLAP2>the</OVERLAP2> <OVERLAP1>the</OVERLAP1> the t- okay i know S1: <OVERLAP1>what you're talking about</OVERLAP1> S2: <OVERLAP1>okay,</OVERLAP1> i just... S1: for the fatty acid breakdown right? S2: uh yeah, i <OVERLAP1>think so</OVERLAP1> S1: <OVERLAP1>okay</OVERLAP1> do you know what chapter that is? <OVERLAP1>it's (way) past this</OVERLAP1> S2: <OVERLAP1>um, in</OVERLAP1> alri- is it all in the same ch- it was like,	Utterance (B) identity of a group member
(7) co-occurring with an explanation	S1: ^(C) glyoxylate? S2: yeah, S1: ^(D) <OVERLAP2>okay</OVERLAP2> <OVERLAP1>i'm just</OVERLAP1> sort of confused about it, ⁽⁷⁾ like the purpose of it? or	Utterance (C) identity of a competent student
(8) co-occurring with a restart	S2: <OVERLAP1>what</OVERLAP1> <OVERLAP1> ⁽⁸⁾ like,</OVERLAP1> i know it's like, it breaks down fatty acids but ⁽⁹⁾ like , how is it related to ⁽¹⁰⁾ like the citric acid it ⁽¹¹⁾ like feeds to the citric a- acid cycle	Utterance (D) identity of a group member
(9) unclassified		
(10) unclassified		
(11) unclassified		
(12) prefacing a concluding remark	S1: right, right okay, ⁽¹²⁾ so , some of the same enzymes are involved, okay S2: okay ⁽¹³⁾ like , uh you go from acetyl CoA to citrate that's okay <OVERLAP2>yeah, yeah</OVERLAP2> <OVERLAP1>you know exactly</OVERLAP1> what you do in the citric acid cycle so you s- have a citrate synthase that does that. ⁽¹⁴⁾ okay , then you go to isocitrate, just like in the	
(13) co-occurring with exemplifications	S2: <OVERLAP2>okay</OVERLAP2> <OVERLAP1>citric acid</OVERLAP1> cycle, so <OVERLAP1>aconitase</OVERLAP1>	
(14) as a boundary marker		
(15), (16) co-occurring with an explanation	S2: <OVERLAP1>and</OVERLAP1> what i- what does this mean? ⁽¹⁵⁾ like , ⁽¹⁶⁾ you know that means that if there_ A-T-P would inhibit it? or, S1: <OVERLAP2>um, let's find out.</OVERLAP2> <OVERLAP1>is that j- like an inhibition?</OVERLAP1> and in calcium and A-D-P s- like, S1: that i- that's what it appears to be so, A-T-P would inhibit that S2: cuz i- is it cuz it bi- it binds to that, or or something?	
(17), (18), (19), (21) as a preface to the response	S1: it probably binds to the enzyme, S2: okay isocitrate dehydrogenase <OVERLAP1>which</OVERLAP1> <OVERLAP1> ⁽¹⁷⁾ oh okay</OVERLAP1> and it S1: <OVERLAP2>uh</OVERLAP2> <OVERLAP1>changes</OVERLAP1> the conformation yeah, so and it's it's an allosteric inhibitor.	

(20) unclassified S2: ⁽¹⁸⁾oh| okay and so these are allosteric activators.
 <OVERLAP1>okay?</OVERLAP1> <OVERLAP1> ⁽¹⁹⁾oh </OVERLAP1> so allosteric
 inhibitors bind to the enzyme and change the active site? or is that,
 <OVERLAP1> ⁽²⁰⁾so ⁽²¹⁾|oh| okay</OVERLAP1>

(MICASE: SGR175SU123)

10.4 Chapter summary and conclusions

The text-based analysis shows that it is possible to explain why DMs occur more in one text than another. More interpretations of the use of DMs are made with reference to the type of activity and relationship between speakers in the text-based analysis. The relationship between speakers can be built on the *discourse identities* and *situated identities* (Zimmerman 1998) which speakers employ in speech. The relationship can also be formed by the changes in *operative identities* (Sacks 1992: 327-328) as speech proceeds. In the text-based analyses of the above excerpts, it is reasonable to assume that the distinctions in using DMs correlate with the type of activity and the speakers' identity. These affect the speakers' decision to give priority to fluency, the engagement of the listeners and the creation of solidarity.

Regarding the connection between DMs and contexts, it is difficult to be precise, but it is interpreted either that there is a connection or occasionally that there is no connection and that DMs are being used for constructing speakers' personae.

It is found that the type of activity is a key factor in using DMs. Most of the texts in the NSs' highly monologic discourse mode in MICASE are lectures, in which speakers seldom use DMs but employ devices which signal listeners' engagement.

The literature has reported that the use of DMs is sensitive to type of activity (e.g. (Fuller 2003a)) and the relationship between speakers (e.g. (Jucker and Smith 1998)). In the above analysis, it is interesting to further identify that the use of DMs is relevant to the speakers' construction of dynamic identities in discourse. It also depends on the demands (e.g. fluency, the engagement of listeners and the creation of solidarity) of particular types of interaction.

In general, the corpus methodologies reveal general patterns of the use of DMs which the NNSs and NSs make. The text-based analysis helps to better understand the co-occurrence of all the DMs and the relationship between DMs and contexts. In this chapter, text-based analysis has explained how and why DMs occur more in one text than another. The corpus study and text-based analysis can be seen as complementary approaches which inform and enrich each other, thereby leading to a better understanding of the use of DMs.

CHAPTER 11: TEXT-BASED ANALYSIS OF DISCOURSE MARKERS IN THE SPEECH OF THE NON-NATIVE SPEAKERS

11.1 Introduction

As highlighted in the introduction of Chapter 10, the corpus approach to the data reveals the collocation phenomena surrounding DMs, which help to interpret the functions of the DMs under investigation. A more qualitative approach, text-based analysis, is then employed to examine the use of DMs from a complementary perspective. Following this analysis of the NNSs' speech, the present chapter examines the Chinese NNSs' speech and tests some hypotheses which cannot be proved or disproved by using corpus methodologies. This chapter seeks to answer whether or not the NNSs' speech can be evaluated according to the presence of DMs and how good the NNSs seem to be at using DMs. This makes it possible to discuss the success or otherwise of the NNSs' usage. The wider question of the NS speech as an appropriate target is also explored.

11.2 Texts for analysis

The six texts listed in Table 11.1 are particularly selected from the two subsets of the SECCL corpus to answer the following research questions:

1. How good do the NNSs seem to be at using DMs?
2. Can the NNSs' speech be evaluated according to the presence of DMs?
3. Are the NNSs successful in using DMs?
4. Can the use of DMs in the NS speech be taken as an appropriate target for NNSs?

Table 11.1: Texts of the non-native speaker speech for analysis

Excerpt number	Type of activity	Participant	Setting	Source
11.1	Talking on a given topic	1 student	Language lab	SECCL: B01-99-16 (Monologue)
11.2	Talking on a given topic	1 student	Language lab	SECCL: B96-08-03 (Monologue)
11.3	Talking on a given topic	1 student	Language lab	SECCL: B01-08-16 (Monologue)
11.4	Talking on a given topic	1 student	Language lab	SECCL: B02-150-32 (Monologue)
11.5	Role-playing	2 students	Language lab	SECCL: C97-01-09 (Dialogue)
11.6	Role-playing	2 students	Language lab	SECCL: C01-01-13 (Dialogue)

With reference to the frequency information of DMs, I searched for high users and low

users of DMs. I also wanted to compare speakers who intuitively seem to be proficient users of English with those who are less competent. I chose four excerpts (Excerpts 11.1 to 11.4) from the NNSs' monologues and two excerpts (Excerpts 11.5 and 11.6) from the dialogues.

11.3 Text-based analysis of discourse markers in the non-native speakers' speech

11.3.1 Monologues in the SECCL corpus

The NNS data are elicited in a rather restricted test-taking context. In the monologues, each speaker has three minutes of preparation time and is asked to talk for three minutes on a given topic (Wen *et al.* 2005: 12-13) (see Appendix 1 for the topics for the monologues). Therefore, the texts for analysis are similar in length. Each text is produced by one speaker only who has the legitimate right to the floor; thus, it can be argued that in this context DMs are very unlikely to be used for maintaining the floor.

The speakers in Excerpts 11.1 and 11.2 are high users of the DMs *oh*, *like*, *now*, *you see*, *you know* and *I mean*. There are 20.5 instances per 1,000 words of these DMs in Excerpt 11.1 and 23.4 instances in Excerpt 11.2. The speaker in Excerpt 11.3 is a non-user of DMs while the speaker in Excerpt 11.4 is a high user of DMs but not a proficient user of English grammar.

In Excerpt 11.1, of the 13 DMs, 8 instances are *oh* co-occurring with reported speech. This can be attributed to the speaker's narrative style and task questions²³. It is found in the NS speech also that NSs use *oh* and *you know* with reported speech. In Excerpt 11.2, 7 out of 17 instances of the DMs co-occur with the hesitation markers *eh* and *um* and pauses, probably suggesting the speakers' need to search for content information or lexical words. These two distinguishing uses of DMs in Excerpts 11.1 and 11.2 seem to be native-like, apart from their high frequencies.

Using DMs to indicate a repair can be something for NNSs to be aware of. 4 instances in Excerpt 11.1 and 6 instances in Excerpt 11.2 of self-repairs are circled. The speakers simply restart when a repair is needed, whereas NSs may use a DM to signal the coming self-repair. Take the second instance of self-repair in Excerpt 11.1, *I just... at that time I was very embarrassed*, for example. An NS may insert *well* between *I just* and *at that time* to give

²³ 2001 Test for English Majors Band 4 (TEM 4): Task B: Talking on a given topic – describe a teacher of yours whom you find unusual.

listener(s) a signal. In the NSs' speech under investigation, it is found that *well* in intra-clausal position prefaces a self-repair.

DMs	Excerpt 11.1: NNSs' monologue (SECCL)	Remarks
(1), (2), (5), (6), (9), (10), (11), (12) co-occurring with reported speech	The teacher I find most unusual is... my English teacher... Mr Tom. He has a very poor memory. Let's set an example. One day I had to set an examination. But I forgot... to take a pen with me. It's very <laugh> terrible. I just asked... eh... Mr Tom "excuse me, Mr tom?" "Yes?" he said. I said <say> "(1) oh ... could you lend me a pen? I forgot... take a pen with me... eh... so... Mr Tom just... eh... he was <is> a very kind and said <say>: "er... that's ok."	
(3), (4) co-occurring with hesitation marker <i>eh</i> and pauses	And gave me a pen... (2) oh ... it was <is> a fine pen and writes very smooth. I thought <think>... (3) now ... eh... (4) now I could <can>... take my examination. Then... after the examination I can not find Mr tom, so I... just kept the pen and intended to the return it to him next day in his class. Next morning I went to the classroom and... find Mr Tom was already... in... the classroom.	
(7) expressing emotions	So I just got <get> out the pen... and came up to him... said <say> "... (5) oh ... Mr tom, thank you very much... eh... ". Before I could say anything, he... he looks... he looked very surprised and said <say>: "(6) Oh , what's this? A nice pen? Is it a gift for me?.... " At that time I was confused and... I don't <laugh> know what to say.	
(8) co-occurring with a key point	And... he just took <take> the pen and thanked <thanks> me... said <say>: "thank you very much... (7) oh , (8) you know , you are the first student who gave <give>... who gave me a pu... a gift ... I like it very much." <u>I just <laugh>... at that time I was</u> very embarrassed and didn't <don't> know what to do and how... how to deal to deal with the situation, I... I... just told him, "(9) oh , Mr tom, I'm sorry, you must... ah... <u>mistook... misunderstood</u> me... eh... this is your pen actually. Do you remember?... eh... yesterday... eh... you <u>gave it... lent <lend> it</u> to me because I didn't take a pen with me at the examination... um... then... eh... ". He seemed <seems>... he looked very puzzled and said <say>: "(10) Oh ? Is that true? I lent <lend> it to you? I don't remember."	4 instances of self-repairs are circled.
(13) co-occurring with a repair	Ah <laugh>... at that time I just laughed and said <say>: "(11) oh ... (12) oh ... Mr tom, you... you are very... eh... absent-minded. (13) I mean you have... a poor memory." But he just laughed and said <say>, "maybe it's my pen... anyway, thank you very much." He... Mr Tom is the most unusual teacher find since... because... he has a very... extremely... poor memory I can not understand. And at the same time, he is very kind... very interesting... and very funny.	

(SECCL: B01-99-16)

DMs	Excerpt 11.2: NNSs' monologue (SECCL)	Remarks
(1), (2), (3), (4), (5), (10), (13) co-occurring with hesitation marker <i>eh</i> and pauses	(1) Well eh... I'll talk about a girl. Ye eh... (2) well eh... Sheer..... was my eh... classmate <classmates>, when I was eh... in the middle school. I will talk about the one thing that when I first eh... entered <enter> the middle school.	
(6) co-occurring with a repair	(3) You know eh... I eh... (4) you see eh... foreign language school, so... eh... it's a very hard it was very hard actually for new come student new coming student, since the test for a eh... eh... what for English eh... students was <is> very hard in the first year.	6 instances of self-repairs are circled.
(7) introducing a new topic	When I eh... first... eh... got <get> touched with eh... English, I found it's very it was very difficult to grasp.	
(8) co-occurring with a restart	(5) You know eh... I... the local dialect (6) I mean local accent... had great influence on my eh... English accent. So I cannot eh... speak English I could not speak Eng speak English eh... very eh... well.	
(9) co-occurring with elaborations	(7) You know this (8) I mean this girl this classmates eh... helped me a lot. First of all, she helped me to to practice some sounds.	
(11), (12) unclassified	(9) I mean this the sounds we we do not have these sounds in our do lo local dialect. (10) So eh... for example, the sound /she/, this sound we never have it in our local dialect. (11) So this girl helped me to practice this sound for eh... what for hundreds of hundreds of times until I eh... really eh... got know eh... I really eh... grasped <grasp> it.	
(14) co-occurring with a negative comment	(12) So it's just eh... it was a pretty good job she had done eh... for me. (13) And e r... (14) you know , since I did very poor eh... in my first two eh... English tests I eh... what I I lo lost my... confidence totally. Actually it's just to me it's the end it was the end of the world. I lost my confidence totally... eh... for English eh... grammar or something like that eh... eh... This time eh... that girl helped me eh... to eh... practice grammar eh... everyday, just every eh... one sentence and eh... another, just everyday to practice grammar.	
(15), (17) as a continuer	(15) And (16) you know , girls are always quite intelligent eh... with language but not boys. So... at that time, she is very eh... she was much advanced eh... than me, much more eh... advanced than me. So she taught me a lot eh... about grammar and some usage of English. So I learned quite a lot from her.	
(16) co-occurring with emphatic lexis, <i>always quite</i>	(17) And after eh... that period of time, my English had <has> improved greatly and I but I can never forget about eh... her behavior in the first eh... first period.	

(SECCL: B96-08-03)

Excerpts 11.3 and 11.4 are chosen for comparison. The speaker in Excerpt 11.3 is a non-user of DMs. This speaker is likely to be taken as a fluent speaker of English in the test-taking context, as she can show accurate usage of syntax, morphology and semantics, making the propositional content clear. In this utterance, the speaker's use of complete sentence structures and the presence of introductory and concluding remarks make this monologue sound like a piece of written English. Moreover, no DM is used and no hesitation markers and pauses occur. The lack of these features of spoken English, however, seems unnatural. In the spoken mode, an appropriate use of DMs can facilitate the understanding of

propositional meaning and interactional interpretations. For instance, in the circled instance in Excerpt 11.3, there is no signal for the repair. If DM *well* had been used, it might have been easier to understand. Otherwise, the listener(s) does not have a clue about why the phrase *but he was so impressive* is repeated.

DMs	Excerpt 11.3: NNSs' monologue (SECCL)	Remarks
No DMs are identified.	<p>He is such an unusual teacher that I have ever met before though he just teach us one year <u>but he was so impressive but he was impressive enough</u> for me to remember him. He's my teacher as I was a freshman the last year. He is not strict <srit> at all, not as strict <srit> as a teacher ought to be at least. I regard him as one of my friends more than one of my teachers. His humorous may be that's the most important reason why I admired him so much.</p> <p>Freshmen as we were at his first class, we all felt nervous. However after he made the funny joke all of us relax at once. His humorous way of teaching encourage us having the interest and right attitude towards our major course English. Besides he is a teacher full of responsibility. He had always been the last person who left the classroom if there if there were some question asked by his students, he won't run away as soon as he heard the ringing bell. He would be as patient as he can to answer all sorts of your questions, from study to life. That's why I said he had been a great teacher and a friend as well. Though he doesn't teach me any more, I will never forget what he once said to me, "don't be lazy or you will be <k?rzi>."</p>	(SECCL: B01-08-16)

In terms of the lack of hesitation and the correctness of grammar, the speaker in Excerpt 11.4 is not as proficient as the speaker in Excerpt 11.3. However, his utterance sounds more natural because it shows the use of DMs, hesitation markers and repetition. These features of spoken English are not found in Excerpt 11.3; however, it is almost impossible to conclude that the speaker in Excerpt 11.3 does not know how to use DMs. Her speech may be constrained by the test-taking context, since non-stop fluency in exams is usually seen as the priority.

DMs	Excerpt 11.4: NNSs' monologue (SECCL)	Remarks
	Task 2	
(1) followed by an evaluation	'This is enough time for college students in their spare time, and I will, (1) I think , I was will as good in the English, so I want to offend a part time job. When I contract <u>a, a</u> , family with eh..., um, with a primary school eh..., I I am the best to teach him, so, the next day, I want his, his room as a tutor, eh..., the, I, before that, I prepare the text very carefully, and consult a lot of books.	Constant use of the hesitation markers, <i>eh</i> and <i>um</i>
(2) unclassified	(2) I think , I was comfortable for that job, so I felt confidently when I, when I do there, the the, the twelfth home, um, then before the twelfth room, the family give me a warmly welcome when um the, I began to teach the the child on the English,	
(3), (4), (6) co-occurring with hesitation marker, <i>um</i>	(3) so , ' (4) so , um, in the um during the reaching, he asked me a word, and said that he didn't know the.. Exactly meaning' among the senior *** so, I, I, I look at that,	
(5) followed by an evaluation	(5) oh , it is easy, but unfortunately, I forgot the***, no, the exact meaning of that, um, eh...,	
(7) co-occurring with a negative point	(6) so , um, I was, (7) you know , upset ***. So ask him take the *** dictionary to me, and consult the dictionary, he lift form year to year, my face was brushed with that, after his mother was going home, he tell his mother he ask me word that he didn't know, his mother felt very surprised, and say that	
(8) omission of content	cannot you, you are the English Major, (8) so ...	

(SECCL: B02-150-32)

Answering the research questions stated earlier, the examination of the above four excerpts from the Chinese NNSs' monologues demonstrates that the NNSs, in general, use DMs appropriately and in a native-like way, although NNSs should become aware of certain ways of using DMs (e.g. signalling repairs). In the test-taking context from which the NNS data were collected, speakers with high marks do not necessarily use DMs in their speech. Speakers with low marks and speakers who are less fluent in oral tests are able to employ DMs for various functions. The analyses of Excerpt 11.3 and 11.4 show that it is difficult to evaluate the speech of the NNS according to the incidence of DMs.

11.3.2 Dialogues in the SECCL corpus

In the NNSs' dialogues, the two speakers have three minutes of preparation time and are asked to discuss for four minutes on the basis of the prompts (Wen *et al.* 2005: 12-13) (see Appendix 1 for the prompts for discussion). The relationship between the two speakers is symmetrical: both are undergraduates with equal power and control over the dialogue.

The two speakers in Excerpt 11.5 are high users of the DMs *oh*, *like*, *now*, *you see*, *you know* and *I mean*. There are 20.5 instances of the DMs per 1,000 words of text. In Excerpt 11.5, the floor is constantly negotiated and re-negotiated as the dialogue goes along. This can be attributed to the test-taking context, where the two speakers are required to express their

ideas and opinions and staying silent would not result in good marks. Furthermore, the nature of the discussion is likely to affect the frequency of DMs, because the turn-taking between the interlocutors is frequent and this leads to more uses of DMs for initiating topics and questions, prefacing responses, mitigating disagreement and criticism and the like.

About two-thirds (12 out of 35 turns) of the turns in this dialogue begin with a DM. They mainly co-occur with emphatic lexis (DMs (5), (11), (13), (19), (20) and (28)) and negative comments (DMs (8), (9) and (26)). It is difficult to be precise about why the 6 occurrences of *but* are used. They seem to preface a counter-claim. It is also possible that the DM collocation, *but you see*, appear to be an idiosyncratic feature of Speaker B's speech style, as Speaker A does not use it at all in this dialogue.

DMs	Excerpt 11.5: NNSs' dialogue (SECCL)	Remarks
	A: Hi. B: Hi, Nina.	
(1) prefacing a new topic	A: Hi, what's the matter? You look not so well? B: I come here for your for your advice. A: Advice for that?	
(2), (4) as a response	B: ⁽¹⁾ You see my department is going to have an English speech contest. A: ⁽²⁾ Oh .	
(3) as a continuer	B: ⁽³⁾ And I'm eager to enter it. But I have some worries. A: ⁽⁴⁾ Oh , that's a chance.	
(5) co-occurring with emphatic lexis, <i>of course</i>	B: ⁽⁵⁾ You see of course I'm eager to to attend it because ⁽⁶⁾ you see I... I want to change my characteristic and I... I want to improve my oral English. I think this is a good chance.	
(6) unclassified		
(7) prefacing a question	A: Um m. Yes, I think so. ⁽⁷⁾ So what do you worry about? B: ⁽⁸⁾ You see I'm a shy girl, and every time when I make speech in public, evrything goes out of my mind so I'm afraid that I can't make my speech fluently.	
(8), (9) co-occurring with negative comments	A: Um m. B: In big is ⁽⁹⁾ I think I will lose my face. A: I think all the things you worry about is not impor... important. Eh... B: Help me please. A: Yes. I... I almost every studentser..... make theirs first speech publicly will feel nervous. B: Yes, yeah. It's my case. A: So I think you should grasp this chance to to practice. B: Yeah I do want to practice to practice I do want to improve myself.	
(10), (12), (14), (16), (17) unclassified	A: Soer..... B: ⁽¹⁰⁾ But ⁽¹¹⁾ you see I'm so afraid of failure. Every time when I when felt <feeled> I my words only lose it's... A: Umm. Yes, I know. ⁽¹²⁾ But ⁽¹³⁾ I think you should pay attention to the anticipation not just for the for the rewards. You you you should thinking regard it as a chance for you to practice. You needn't to worry the result. B: ⁽¹⁴⁾ But ⁽¹⁵⁾ you see if once I I take part it. I want to win.	
(11) co-occurring with emphatic lexis, <i>so</i>		

(13) co-occurring with emphatic lexis, <i>should</i>	A: Um um... I know I know. I can feel the same as you. Last time I took part in this kind of contest also. At first I also felt nervous and I also worried the result. ⁽¹⁶⁾ But my teacher comfort me that eh..... all these things are not important. Just put yourself in it so that's enough.
(15) unclassified	
(18) unclassified	B: ⁽¹⁷⁾ But ⁽¹⁸⁾ you see make it instead every time when I speak in public.
(19), (20) co-occurring with emphatic lexis, <i>so</i>	Eh... I can not control control myself ⁽¹⁹⁾ you see I'm so nervous ⁽²⁰⁾ you see. A: Um m.
(21), (27) as a continuer	B: ⁽²¹⁾ And I'm so I'm afraid that people will laugh at me. A: Maybe maybe when you speak when you speak in on that moment, you can forget all the people. Eh... in the audience.
(22), (23) unclassified	B: I hope that but ⁽²²⁾ I think I... that is... that is not the case. A: You can try ⁽²³⁾ I think you can try.
(24), (25) unclassified	B: ⁽²⁴⁾ You see if if I failed I will loss I will lose all of my self-confidence. A: Everybody have have a failure and victory. So you... B: ⁽²⁵⁾ But ⁽²⁶⁾ you see I have failed so much. failed a lot so...
(26) co-occurring with a negative comment	A: Maybe this time you can success, succeed it. ⁽²⁷⁾ And ⁽²⁸⁾ I think several days before this contest, you have to, you had better to listen to some music or talk with your friends your friends freely.
(28) co-occurring with emphatic lexis, <i>have to,</i> <i>had better</i>	(SECCL: C97-01-09)

Excerpt 11.6 is selected as an example of speech with rather few DMs. The key feature of this excerpt is that the two speakers constantly exchange opinions and disagree with each other. For instance, Items (A) *No, I don't think so*, (B) *No, he is not experienced*, (C) *No, I don't think so* and (D) *No, I think* directly contradict the other speaker. As noted in the preceding chapter, these are called *dispreferred* responses. NSs will use such modal verb as *would* to hedge the disagreement (Carter and McCarthy 2006: 708). They would not place a *dispreferred* response in turn-initial position and they could well use hesitation markers, DMs, agreement prefaces and the like (Pomerantz 1984: 72) before a *dispreferred* response in order not to be too direct and imposing. The two speakers' direct speech style would be avoided by NSs, since direct speech tends to be associated with aggression, over-assertiveness and lack of consideration for people's feelings.

DMs	Excerpt 11.6: NNSs' dialogue (SECCL)	Remarks
(1) prefacing a new topic	A: Hello, Sdelor. B: Hello, John. A: ⁽¹⁾ Oh , one of my friends want to go abroad, um... what your opinion about that of that news.	
(2) as a response	B: ⁽²⁾ Oh , what he what he is study now? A: He has just finished his high school. B: High school? A: Yes. B: <i>I think</i> the time is not <raipl> is not <raipl> for him to go overseas. A: Not proper? B: Not <raipl>. A: Why? B: Because now he is only a... high school students. Yes? A: Yeah.	Type A <i>I think</i> in bold and italics
(3) as a continuer	B: ⁽³⁾ And ... one aspect he was he is young and not experience, do you think so? A: ^(A) No, I don't think so. Um... I <shu> <i>I think</i> most of the... foreign um... students just like his age can go... can um... deal with all the things by them themselves. <i>I think</i> I I my friend can do so also can do so.	Items (A), (B), (C) and (D) dispreferred responses
(4) co-occurring with emphatic lexis, <i>full of</i> and <i>very</i>	B: ^(B) No, he is not experienced. ⁽⁴⁾ You know now the foreign foreign countries full of competitions and the situation was is very serious. <i>I think</i> he he could he can't... confront all the situations. A: ^(C) No, I don't think so. <i>I think</i> a an experience um... everyone who goes abroad don't have experience they can practice abroad and <i>I think</i> they have studied um... in our country for so long and <i>I think</i> um... he had learned a lot of things um... about China, so he want he should go abroad to know some things about other countries.	
(5) co-occurring with emphatic lexis, <i>many</i>	B: <i>I think</i> he maybe he will <fis> ⁽⁵⁾ oh , many <i>many</i> ... difficulties, maybe he will not deal with all the situations. It it maybe make him very depressed. A: ^(D) No, <i>I think</i> it's a it's a good chance for him to broaden his his eyes, broaden his knowledge.	
(6) as a continuer	B: ⁽⁶⁾ And , now he is not rich and <knowledged>, so eh... especially he he haven't he hasn't a <kat> of... of major. ⁽⁷⁾ I mean the technic, ⁽⁸⁾ you know . And there are many examples <sav> um... many of our countries' student went to the abroad, but he couldn't...	
(7) prefacing a repair	A: ⁽⁹⁾ But ⁽¹⁰⁾ I think he can go abroad and choose a major. He can <sta> continue to study. ⁽¹¹⁾ I mean he can continue to study in the university also in the college and then he can um... choose a major.	
(8), (9), (10) unclassified	B: Do you think his foreign language's good? A: Yes, <i>I think</i> it it just ok. Maybe ha can't understand it at the beginning but his English is very good at at the middle school.	
(11) prefacing a restart	B: Do you do you think he can offered all the he can offer all the tuition tuition? A: Tuition? B: Yeah.	
(12), (13), (14), (15) unclassified	A: It doesn't matter. ⁽¹²⁾ I think his family is very rich and um... his mother his mother is a manager. His father is a computer computer um... in a computer company. B: ⁽¹³⁾ But ⁽¹⁴⁾ but still I ⁽¹⁵⁾ I think he is very young and couldn't go abroad by himself.	

(SECCL: C01-01-13)

The four speakers in Excerpts 11.5 and 11.6 use DMs in similar ways. In this particular context of exchanging opinions on a given topic, the exchange of claim and counter-claim is frequent and both Type A *I think* (shown in bold and italics) and Type B *I think* occur often.

In Excerpt 11.6, refusal items, marked with (A), (B), (C) and (D) in superscript seem, on the one hand, to be too direct but, on the other, they seem to be straightforward and explicitly express opinions as between fellow students. In academic discourse, NSs would use expressions such as *it seems to me*, *what puzzles me*, *I was wondering*, *what I'm saying*, *you're saying*, etc (Mauranen (2003a) cited in Mauranen (2006: 146)).

Considering the type of activity, the role of participants and the setting in Excerpts 11.5 and 11.6, I would argue that the NNSs are successful in using DMs. However, I have some reservations about the Chinese NNSs' use of DMs if they communicate with NSs and use DMs in the same ways as they talk with their peers.

This, however, raises the question of whether or not NNSs should aim to sound like NSs. In the above case, the two Chinese NNSs seem to mirror each other's linguistic usages. They probably do not perceive the other speaker to be too direct. However, if they were communicating with NSs, the NSs would probably interpret their speech negatively and have a bad impression of the NNSs.

Take the use of DM *like* in Excerpt (4.5.1) in Chapter 4 for example. It is repeated here for ease of reference as Excerpt 11.7. The three NS students in their study group discussion frequently use *like* as a DM. The co-occurrence of some instances of *like* can be identified but in some cases it is not possible to identify why they are used. My interpretation is that some instances of *like* in Excerpt 11.7 are used only as in-group markers in order to create solidarity. If this interpretation is correct, it will not be surprising that the NNSs under investigation do not pick up the use of *like* as a DM, as they have been learning to do with other NNSs in an environment where English is not spoken and contact with NSs is limited. It is also possible that creating solidarity is not an issue in the exam context. The NNSs do not have a personal need to form a relationship with their interlocutors.

To answer the question of whether or not NNSs should know how to use *like* as a DM, perhaps in the context of talking with professors or other NNSs who have different L1s, it does not matter if NNSs do not know the use of *like* as a DM. It is probably better not to use *like*, as shown in Excerpt 11.7, no matter what use might be given to it. In Speaker 3's first utterance, using too many instances of *like* seems not to help understanding. In particular, it cannot be objectively identified in any way why the three instances of *like* referred to as Items (11), (12) and (13) should have been used. Nevertheless, if NNSs would like to integrate with a particular group, it may be of help to use *like*. For instance, it will probably be beneficial for

international students in the English-speaking countries to be aware that using DMs as in-group markers can be a way of expressing solidarity and establishing membership.

DMs	Excerpt 11.7: NSs' highly interactive discourse mode (MICASE)	Remarks
(1), (2) unclassified S2: maybe you won't be able to yeah i might not be able to or i need to see if ⁽¹⁾ like someone else could ⁽²⁾ like kinda just take it on right now or, if Frank if i could be like ⁽³⁾ oh Frank will you just write it? or something like that ⁽⁴⁾ you know . uhuh mhm,	
(3) co-occurring with reported speech		
(4) co-occurring with approximations		
(5) unclassified	S3: yeah. ⁽⁵⁾ well so i got the extension from him and then i finally just ended up talking to my professor on North Campus about it. and i told her what i was interested in doing, and then i was like could i use the research? and she was like_ cuz there's several different parcels of research ⁽⁶⁾ like , there's ⁽⁷⁾ like five different communities ⁽⁸⁾ i think in Detroit that they did interviews with and ⁽⁹⁾ like s- ⁽¹⁰⁾ like (peer g-) the diff- there're different characteristics for all of 'em. some of 'em are like industrial areas some of 'em are retail areas, and so some of them a- um all of them are gonna be going into the research that she does at the end of the year and stuff but also ⁽¹¹⁾ like she went to ⁽¹²⁾ like different neighborhood organizations ⁽¹³⁾ like , maybe ⁽¹⁴⁾ like um, employment organizations and things in the area and was like do you guys want some of the information that we get from this stuff? ⁽¹⁵⁾ like cuz it would probably be	
(6) co-occurring with exemplifications		
(7) co-occurring with numbers and (8) <i>I think</i> , could be for hedging		
(9), (10) co-occurring with a restart	S3: <OVERLAP2> mhm </OVERLAP2>	
(11), (12), (13), (15) unclassified	S1: <OVERLAP2> yeah </OVERLAP2> S3: <OVERLAP1> helpful to </OVERLAP1> them. so that's what i've been working on a lot for her last year was doing that for a couple um ⁽¹⁶⁾ like uh, industrial organizations around the area and some of them haven't been done at all. so she wanted me to do some of those um for, my project. so then <EVENT WHO="SS" DESC="LAUGH" /> so then i would do them and i would put that in my project. ⁽¹⁷⁾ like i would have a part that was like this is empowerment zones in Detroit.	
(14) co-occurring with vague language <i>maybe</i>		
(16), (18) co-occurring with hesitation markers <i>um</i> and <i>uh</i>	S1: yeah this is um, ⁽¹⁸⁾ like what local labor market theory is and then ⁽¹⁹⁾ like this is one specific thing so i would on- also get i would paid for it. <EVENT DESC="LAUGH" /> <OVERLAP1> which is kind of cool.</OVERLAP1>	
(17) co-occurring with elaborations	S3: <OVERLAP1> that's nice. </OVERLAP1>	
(19), (20), (21) unclassified	S1: <OVERLAP1> i think that that, </OVERLAP1> that would be good because ⁽²⁰⁾ like , we're supposed to have ⁽²¹⁾ like a other aspect besides	

(MICASE: SGR999SU146)

11.4 Chapter summary and conclusions

In this chapter, I look at the high users and non-users of DMs among the NNSs to see if there is a correlation between the use of DMs and good speakers. It is concluded that the use of

DMs is not an indication of whether or not the speaker is a good speaker in the test-taking context. The speakers in Excerpts 11.1 and 11.2 are high users and their use of DMs seems to be appropriate. The speaker in Excerpt 11.3 is a good speaker of English but she does not use any DMs in the monologue, which makes the utterance sound like formal written English. In contrast, the speaker in Excerpt 11.4 is not competent to produce correctly grammatical English, but his use of DMs and other features of spoken English make the utterance sound more natural than Excerpt 11.3 does.

In the NNSs' dialogues, the speakers in Excerpt 11.5 are high users of DMs. Some of DMs used by Speaker B seem to be strong, such as *but* and *you see*. The two speakers in Excerpt 11.6 are low users. Their dialogue seems to be too direct, probably because they are not using DMs to soften or mitigate their speech. It would be ideal to look at non-users of DMs in the NNSs' dialogues, but I was unable to find any texts which had no DMs in this sub-corpus.

Both corpus study and text-based analysis of the speech of NSs and NNSs show distinctions in the use of DMs between them. However, the evidence does not tell us whether the distinctions are due to their being NNSs or due to the nature of test-taking discourse. This creates the dilemma of assessing spoken English through an exam. On the one hand, oral exams may not be able to assess learners' speaking ability in all aspects of spoken English; on the other, there seem to be no better ways to assess learners' performance.

As mentioned in the previous chapters, it is surprising that the frequency of DMs in the NNS speech in the test-taking context is notably high. In Hellermann and Vergun's study (2007) of 17 beginning-level adult learners, DMs *like*, *well* and *you know* are used by learners only once on average every 90 turns. By contrast, in the NNSs' dialogues in this study, these three DMs are used once every 23.7 turns²⁴. It remains a mystery why the NNSs under investigation use a noticeably high number of DMs, for these are not likely to have been taught in most language classrooms. Nor is it possible that the NNSs learnt the DMs through interactions in English with NSs in a country where English is not commonly spoken.

²⁴ Based on an extrapolation of the known incidence in random samples, the occurrences of *like*, *well* and *you know* as discourse markers in extra-clausal turn-initial position are possibly 7, 415 and 827 respectively. These numbers of incidence are obtained from (raw counts) x (proportion of Type B word/phrase) x (proportion of Type B in turn-initial position). In the case of *like*, 7 instances are from 984 x 3% (Table 4.1) x 22.2% (Table 4.14). In the case of *well*, 415 instances are from 1,384 x 37.3% (Table 6.1) x 80.4% (Table 6.14). For *you know*, 827 instances are from 3,263 x 82% (Table 7.1) x 30.9% (Table 7.30). There are possibly 29,542 turns in the sub-corpus of the NNSs' dialogues if we search for speaker referents *a*: (15,110 tokens) and *b*: (14,432 tokens). In the NNSs' dialogues, these three discourse markers are used once every 23.7 turns (29,542 turns/1,249 instances =23.7).

The non-native-like aspects of using DMs in the NNS speech have been pointed out in the literature (see the discussion in Chapter 2) and in this thesis. This leads to the question whether it is desirable for NNSs to sound like NSs. It is argued that deciding whether or not to take NS speech as an appropriate target should depend on NNSs' needs. When NNSs are talking to NSs or when they want to integrate into a particular group, it would be more desirable to sound like NSs. When they communicate with other NNSs, sounding like NSs is probably not a high priority for NNSs.

In the above text-based analysis, three findings which are not discussed in the corpus analysis can be reported. First, in the NNSs' texts for analysis, it is found that the NNSs tend to simply restart their utterance for a self-repair, rather than using a DM to signal it, as an NS would probably do.

Second, it seems to be impossible to point out what has gone wrong in the example of spoken English without any DMs. As exemplified in Excerpt 11.3, the speaker used no DMs in the given example. The utterance is similar to a piece of writing. The use of DMs would probably have facilitated the listener's process of understanding, even in a test-taking context.

Last, it is difficult to specify how often DMs should be used in spoken English, although, in the NSs' speech, good examples of the use of DMs can probably be identified. The speakers in Excerpts 10.1 and 10.2 in the previous chapter appear to manipulate DMs well in their lengthy utterances in the classroom.

The text-based analysis helps to identify some uses of DMs which are not found in the corpus investigation and to interpret why and how DMs are used. These two approaches to the investigation of DMs are found to be complementary, working well together.

Both corpus study and text-based analysis have implications and applications for pedagogy, which is explored in the next chapter. The collocation phenomena identified in the corpus study and the factors found in the text-based analysis will, it is hoped, contribute to the teaching of English to NNSs and to the investigation of varieties of global English.

CHAPTER 12: SUMMARY AND DISCUSSION

12.1 Introduction

This chapter provides a summary of the description and analyses of the Type B words/phrases under investigation, draws conclusions about the use of DMs with supporting evidence and discusses the implications of the research. The focus of the discussion is on two sets of comparisons: 1) DMs used in two types of genre – monologic and dialogic and 2) the use of DMs by two groups of speakers – Chinese NNSs and the NSs.

12.2 Summary of the analyses of discourse markers

12.2.1 The proportions of the words/phrases as discourse markers

The instances of the words and phrases have been manually classified into non-discourse use (Type A) and discourse use (Type B or DMs). Table 12.1 below presents the percentage of Type B words/phrases in the six sub-corpora. *You know* and *I mean* are central DMs. Over two-thirds of the instances across the two types of genre and in the two groups of speakers are used as DMs. *Well* is also a central DM, in particular in the NS dialogic genres, where over 85% of the instances of *well* as DMs, while *you see* is primarily used as a DM in the NNSs' speech. DM *like* apparently occurs more often in the highly interactive discourse mode in MICASE than in other five sub-corpora. Similarly, *now* as a DM appears relatively frequently in MICASE.

Table 12.1: Percentage of the words/phrases as discourse markers across corpora

Corpus/ Discourse marker	Monologic genres			Dialogic genres		
	SECCL: Monologues (Chinese NNSs)	MICASE: Highly monologic discourse mode (American NSs)	ICE-GB: Unscripted monologues (British NSs)	SECCL: Dialogues (Chinese NNSs)	MICASE: Highly interactive discourse mode (American NSs)	ICE-GB: Private direct conversations (British NSs)
<i>like</i>	1.7	5.3	2.6	3.0	57.3	15.7
<i>oh</i> *	n/a	n/a	n/a	n/a	n/a	n/a
<i>well</i>	16.0	52.0	44.8	37.3	90.0	85.7
<i>you know</i>	75.2	73.0	83.5	82.0	71.0	81.3
<i>I mean</i>	91.1	86.4	66.7	76.7	87.7	93.7
<i>you see</i>	74.4	29.3	37.1	89.2	17.1	67.4
<i>I think</i>	10.0	22.5	25.3	7.0	11.7	19.0
<i>now</i>	21.7	55.3	31.0	17.7	48.3	29.0

*There is no distinction between non-discourse use and discourse use in the case of *oh*.

12.2.2 The frequencies of the words/phrases as discourse markers

The normalised frequencies of Type B words/phrases are listed in Table 12.2 below. In most cases, there are more instances of DMs in the dialogic genres than in the monologic genres and this supports my hypothesis that the more interactive the genre or activity type is, the more DMs occur. However, *now* in MICASE and ICE-GB and *you see* in MICASE are rather different from other DMs. The frequencies of these in the monologic genres are higher than those in the dialogic genres. *Now*, in MICASE, occurs 15.1 times per 10,000 words in the highly monologic discourse mode, as opposed to 9.8 times in the highly interactive discourse mode and in ICE-GB 12.5 times in the unscripted monologues, as opposed to 5.8 times in the private direct conversations. *You see*, in MICASE, appears 1.3 times per 10,000 words in the highly monologic discourse mode, as opposed to 0.7 times in the highly interactive discourse mode.

Table 12.2: Frequency comparisons of the discourse markers in the monologic and dialogic genres under investigation (Normalised frequency per 10,000 words (times))

Corpus/ Discourse marker	Monologic genres			Dialogic genres		
	SECCL: Monologues (Chinese NNSs)	MICASE: Highly monologic discourse mode (American NSs)	ICE-GB: Unscripted monologues (British NSs)	SECCL: Dialogues (Chinese NNSs)	MICASE: Highly interactive discourse mode (American NSs)	ICE-GB: Private direct conversations (British NSs)
<i>like</i>	0.4	1.1	0.4	0.5	84.4	7.7
<i>oh</i>	11.8	1.8	2.9	48.3	48.4	60.7
<i>well</i>	2.4	7.8	10.3	8.7	32.9	70.5
<i>you know</i>	6.8	8.9	4.3	44.8	32.8	36.0
<i>I mean</i>	1.5	4.3	0.9	4.1	25.8	43.8
<i>you see</i>	0.9	1.3	0.8	6.2	0.7	5.2
<i>I think</i>	3.0	1.5	2.8	11.1	3.7	6.8
<i>now</i>	3.1	15.1	12.5	3.5	9.8	5.8

There are good grounds for comparing the frequencies of the DMs in the monologic genre with those in the dialogic genres. In the SECCL corpus, the speakers in the monologues are the same as those in the dialogues. In the two NS corpora, MICASE and ICE-GB, the speakers across the two types of genre are not the same, but their backgrounds are similar. Arguably, the frequencies in the sub-corpora of the monologic genres can be compared with those in the dialogic genres. However, the Chinese NNS corpus (SECCL) and the two NS corpora (MICASE and ICE-GB) are not designed for comparison (see Section 3.1 of Chapter 3, where the issue of comparability between corpora is discussed in some detail). As the nature of these three corpora varies, differences in frequency and the use of DMs between the

two groups of speakers cannot be made without controversy. In addition, taking the NS usages of DMs as the target norm for NNSs is another issue. With due consideration for these issues, in the discussion of differences between the NNSs and NSs, the neutral terms *over-* and *under-representation* are used (see Section 2.2.1 of Chapter 2 for a more detailed discussion). In Table 12.1 above, DM *like* is over-represented in the sub-corpus of the highly interactive discourse mode in MICASE, occurring 84.4 times per 10,000 words, as opposed to fewer than 7.7 times in the five other sub-corpora. *Like* in the two NNS sub-corpora and the sub-corpus of the unscripted monologues in ICE-GB is under-represented, occurring fewer than 0.5 times per 10,000 words. The normalised frequency is based on fewer than 10 instances in each sub-corpus. Due to the low number of occurrences, I argue that in the Chinese NNSs' speech and in the British NSs' unscripted monologues, *like* is almost never used as a DM.

12.2.3 Positional preference of discourse markers

Table 12.3 presents the predominant positions in an utterance/turn of the eight DMs in the monologic and dialogic genres. It can be clearly seen that the DMs have their preferred positions. The most common position of the DMs for analysis, except for *like*, *oh*, *well* and *I think*, is extra-clausal utterance/turn-medial position. *Like* shows a preference for the intra-clausal position, following an incomplete message (M-) element. *Oh* tends to occur in extra-clausal utterance/turn-initial position and *well* tends to have the same position in the dialogic genres. Most of the instances of *I think* in the unscripted monologues in ICE-GB are placed intra-clausally following an M- element, but in the NNSs' dialogues, the typical positions are turn-medial and turn-final.

Table 12.3: Summary of the positions in an utterance/turn of the discourse markers in the monologic and dialogic genres under investigation

Corpus/ Discourse marker	Monologic genres			Dialogic genres		
	SECCL: Monologues (Chinese NNSs)	MICASE: Highly monologic discourse mode (American NSs)	ICE-GB: Unscripted monologues (British NSs)	SECCL: Dialogues (Chinese NNSs)	MICASE: Highly interactive discourse mode (American NSs)	ICE-GB: Private direct conversations (British NSs)
<i>like</i>	n/a	After an M-	n/a	n/a	After an M-	After an M-
<i>oh</i>	Utterance-medial	Utterance-initial; after an M-	Utterance-medial	Turn-initial	Turn-initial	Turn-initial
<i>well</i>	Utterance-medial	Utterance-medial	Utterance-medial	Turn-initial	Turn-initial	Turn-initial
<i>you know</i>	Utterance-medial	Utterance-medial	Utterance-medial	Turn-medial	Turn-medial	Turn-medial
<i>I mean</i>	Utterance-medial	Utterance-medial	Utterance-medial	Turn-medial	Turn-medial	Turn-medial
<i>you see</i>	Utterance-medial	Utterance-medial	Utterance-medial	Turn-medial	Turn-medial	Turn-medial
<i>I think</i>	Utterance-medial	Utterance-medial	After an M-	Turn-medial/ Turn-final	Turn-medial	Turn-medial
<i>now</i>	Utterance-medial	Utterance-medial	Utterance-medial	Turn-medial	Turn-medial	Turn-medial

12.2.4 Co-occurrence and suggested functions of discourse markers

Table 12.4 below summarises the most frequent types of co-occurrence of the DMs, which present particular uses of each DM in relation to the two types of genre (see the tables in Sections 4.4.2.9 (*like*), 5.4.2.9 (*oh*), 6.4.2.12 (*well*), 7.4.2.12 (*you know*), 7.4.3 (*I mean*), 7.4.4 (*you see*), 8.4.2.6 (*I think*) and 9.4.2.10 (*now*) for the distribution of the identified types of co-occurrence of each DM). Most of the instances of *like* co-occur with exemplifications in the highly monologic discourse mode in MICASE, while they often co-occur with expressions of uncertainty in the highly interactive discourse mode. *Oh* and *well* in the monologic genres primarily mark reported speech, whereas, in the dialogic genres, they are often used as a (preface to a) response. *You know* is primarily used to emphasise a co-occurring statement. *I mean* predominantly co-occurs with clarifications and explanations in the two types of genre and across the two groups of speakers. The most frequent type of co-occurrence of *you see* is explanations, justifications and conclusions in both the monologic and dialogic genres. *I think* seems to be used as a delaying device by the NNSs, because it co-occurs most often with hesitation markers, pauses and restarts, while it co-occurs frequently with factual information either to express uncertainty or to appear less assertive in the NSs' speech. *Now* is primarily used as a boundary marker, indicating shifts of (sub)topic and viewpoint.

Table 12.4: Summary of the most frequent types of co-occurrence of the discourse markers in the monologic and dialogic genres under investigation

Corpus/ Discourse marker	Monologic genres			Dialogic genres		
	SECCL: Monologues (Chinese NNSs)	MICASE: Highly monologic discourse mode (American NSs)	ICE-GB: Unscripted monologues (British NSs)	SECCL: Dialogues (Chinese NNSs)	MICASE: Highly interactive discourse mode (American NSs)	ICE-GB: Private direct conversations (British NSs)
<i>like</i>	n/a	Exemplifications	n/a	n/a	Expressions of uncertainty	Expressions of uncertainty/ explanations
<i>oh</i>	Reported speech	Reported speech/ showing emotions	Showing emotions	As a (preface to) response	As a (preface to) response	As a (preface to) response
<i>well</i>	Reported speech	Transitions; shifts of topic	Transitions; shifts of topic	Disagreement; negative evaluation/ as a preface to response	Transitions; shifts of topic/ as a preface to response	As a preface to response/ disagreement; negative evaluation/ as a continuer
<i>you know</i>	<u>Emphatic lexis;</u> <u>key information</u>	Clarifications; explanations	<u>Emphatic lexis;</u> <u>key information/</u> Clarifications; explanations	Contrasting and negative points/ <u>Emphatic lexis;</u> <u>key information</u>	Shared knowledge presumed by speaker/ clarifications; explanations	<u>Emphatic lexis;</u> <u>key information/</u> Shared knowledge presumed by speaker
<i>I mean</i>	<i>Clarifications;</i> <i>explanations</i>	<i>Clarifications;</i> <i>explanations/</i> <i>Elaborations</i>	Hesitation markers; pauses; restarts/ <i>Clarifications;</i> <i>explanations</i>	<i>Clarifications;</i> <i>explanations/</i> Hesitation markers; pauses; restarts	<i>Clarifications;</i> <i>explanations/</i> Elaborations	<i>Clarifications;</i> <i>explanations/</i> Elaborations
<i>you see</i>	Explanations; justifications; conclusions/ Emphatic lexis	Explanations; justifications; conclusions	Indications of objects and places	Explanations; justifications; conclusions	Explanations; justifications; conclusions	Explanations; justifications; conclusions
<i>I think</i>	Hesitation markers; pauses; restarts	Factual information	Factual information/ Personal opinions & evaluation	Hesitation markers; pauses; restarts	Personal opinions & evaluation	Factual information
<i>now</i>	Opening & closing of topic; concluding remarks/ Shifts of (sub)topic and viewpoint	Shifts of (sub)topic and viewpoint	Shifts of (sub)topic and viewpoint	Emphatic lexical items and structure	Shifts of (sub)topic and viewpoint/ Questions	Shifts of (sub)topic and viewpoint

The collocation phenomena are used as categories for discussion and they lead to the interpretations of the functions of DMs. Table 12.5 below lists the types of co-occurrence of the DMs and the interpretations of their functions.

It can be seen that all the eight DMs for analysis are multi-functional. This characteristic of multi-functionality causes difficulty in interpreting the use of DMs, as highlighted at the beginning of this thesis. It is problematic to intuitively interpret the functions of DMs, as most previous studies have done, because a researcher cannot read a speaker's mind; in most cases the uses of DMs are not even easily available to introspection by the speaker. This thesis uses collocation phenomena to categorise the uses of DMs and clarify the logic of the

identification of their functions. However, occasionally more than one type of co-occurrence is found in the same instance. In cases of this kind, the classification has to be judged on the basis of intuition.

The functions listed in Table 12.5 can be split into two broad categories. The first twelve items are primarily for textual organisation. The use of DMs to perform these functions helps the process of comprehension. For example, the use of *oh*, *well*, *you know* and *I mean* signalling a repair and *well* and *now* marking a transition give the listeners a hint about the coming change. From the thirteenth item to the last, the items primarily contribute to the interpersonal aspect of interaction. For instance, *well*, *you know*, *I mean* and *I think* can be used as mitigators to avoid sounding too assertive and soften the impact of criticism. The function of *now* and *you see* is to attract the listeners' attention.

Table 12.5: Functions of discourse markers identified on the basis of co-occurrence

Co-occurrence	Function	Discourse markers (highlighted, relatively frequent in the NNSs' speech; underlined and in bold, relatively frequent in the NSs' speech)
1 Hesitation markers; pauses; restarts	<ul style="list-style-type: none"> To suggest a search for contents or lexis; to hold the floor To sound less direct To reformulate due to being interrupted 	<i>like, oh, well, you know, I mean, you see, I think</i>
2 Exemplifications	<ul style="list-style-type: none"> To introduce exemplifications 	<i>like, you know, I mean, you see, now</i>
3 Explanations; clarifications	<ul style="list-style-type: none"> To introduce explanations/clarifications 	<i>like, you know, I mean, you see, now</i>
4 Repaired/replaced items	<ul style="list-style-type: none"> To signal a repair 	<i>oh, well, you know, I mean</i>
5 Reported speech	<ul style="list-style-type: none"> To mark the boundary between the mode of the speaker and reported speech 	<i>like, oh, well, you know,</i>
6 Opening/changing of a topic	<ul style="list-style-type: none"> As a topic changer 	<i>oh, well, you know, you see, now</i>
7 Concluding remarks	<ul style="list-style-type: none"> To indicate a conclusion 	<i>You know, you see, I think, now</i>
8 Shifts of topic	<ul style="list-style-type: none"> To mark a transition 	<i>well, now</i>
9 Continuation of the earlier topic; elaborations	<ul style="list-style-type: none"> As a continuer 	<i>well, I mean</i>
10 Listing items and sequence of events	<ul style="list-style-type: none"> To separate units and draw attention to the following point 	<i>now</i>
11 Emphatic lexis and key information	<ul style="list-style-type: none"> To emphasise the statement 	<i>you know, you see, now</i>
12 Prefacing a question	<ul style="list-style-type: none"> To mark a boundary and shift of topic To sound less direct and imposing 	<i>now</i> <i>well, I think</i>
13 Prefacing responses	<ul style="list-style-type: none"> As a (preface to) response to a question and new information To mitigate indirect/insufficient answers 	<i>oh, well</i>
14 Disagreement and negative evaluation	<ul style="list-style-type: none"> As a mitigator 	<i>well</i>
15 Contrasting and negative points	<ul style="list-style-type: none"> As a mitigator 	<i>you know, I mean</i>
16 Personal opinions and evaluation	<ul style="list-style-type: none"> To avoid being too assertive with positive evaluation To mitigate negative evaluation 	<i>I think</i>
17 Factual information	<ul style="list-style-type: none"> To express uncertainty To appear less assertive To reduce commitment 	<i>I think</i>
18 Contrasting items	<ul style="list-style-type: none"> To draw attention to the following point 	<i>now</i>
19 Indications of objects and places	<ul style="list-style-type: none"> To gain attention 	<i>You see, now</i>
20 Numerical expressions	<ul style="list-style-type: none"> To make approximations To focus the coming information 	<i>like</i>
21 Expressions of certainty; key points	<ul style="list-style-type: none"> To focus a key point To draw attention 	<i>like, well</i>
22 Expressions of uncertainty; vague language	<ul style="list-style-type: none"> To express uncertainty 	<i>like</i>
23 Expressions of emotions	<ul style="list-style-type: none"> To show emotions 	<i>oh</i>
24 Cognition-related verbs	<ul style="list-style-type: none"> To indicate a cognitive process has been done 	<i>oh</i>
25 Shared knowledge presumed by the speaker	<ul style="list-style-type: none"> To build consensus 	<i>you know, you see</i>

In most contexts, the speaker has more than one DM to choose from. The choice may be affected by genre, activity type and identity construction (see Section 12.3.1 below for further

discussion). For instance, the speaker may use *now*, *well* and *I think* before raising a question. A faculty member in the classroom setting is likely to use *now* in order to mark a shift of topic and to sound confident, in particular when s/he knows the answer. The question may not be posed to elicit responses but for other purposes, such as marking the boundary in discourse and engaging the listener(s). A faculty member in an office hour session with a student may use *well* and *I think* prefacing a question to sound less imposing and to downplay her/his academic power and status.

In Table 12.5 above, some DMs are more frequent in one group of speakers than another. The highlighted DMs are more frequent in the NNSs' speech and those underlined and in boldface are relatively frequent in the NSs' speech. The different uses of DMs in the speech of the NNSs and NSs are summarised in Section 12.3.2 below.

12.2.5 Discourse marker collocations

The functions associated with the eight DMs are identified on the basis of the co-occurring linguistic items. It is revealed that the DMs have their own particular functions in discourse. When they co-occur with other DMs, they show a preference for the order of DM collocations. Table 12.6 below lists DM collocations in the speech of the NNSs and NSs under investigation. *Oh* and *well* in a DM collocation are usually followed by another DM, whereas *like*, *you know*, *I mean*, *you see*, *I think* and *now* are usually preceded by other DMs. Moreover, it is evident that DM collocations are much more varied in the NSs' speech than in the NNSs' speech. It would be intriguing to look at discourse markers collocations in more detail; however, the limited space does not allow it.

Table 12.6: Discourse marker collocations in the speech of the non-native speakers and native speakers under investigation

Discourse marker for analysis	Chinese NNSs	American and British NSs
<i>like</i>	--	<i>and like</i> <i>I mean like</i> <i>but like</i>
<i>oh</i>	<i>oh no</i>	<i>oh okay</i> <i>oh yeah</i> <i>oh yes</i> <i>oh no</i> <i>oh right</i> <i>oh well</i>
<i>well</i>	<i>oh well</i> <i>well I think</i> <i>well you know</i> <i>well but</i>	<i>well well</i> <i>well I think</i> <i>well you know</i> <i>ok (okay) well</i> <i>like well</i> <i>so well</i> <i>yeah well</i>
<i>you know</i>	<i>and you know</i> <i>but you know</i>	<i>well you know</i> <i>but you know</i> <i>you know and</i> <i>like you know</i> <i>I mean you know</i>
<i>I mean</i>	<i>I mean I mean</i> <i>but I mean</i>	<i>but I mean</i> <i>so I mean</i> <i>well I mean</i> <i>yeah (yes) I mean</i> <i>no I mean</i> <i>and I mean</i> <i>I mean like</i> <i>I mean and</i> <i>you know I mean</i> <i>I mean you know</i>
<i>you see</i>	<i>but you see</i> <i>and you see</i> <i>so you see</i> <i>yes (yeah) you see</i>	<i>and you see</i> <i>well you see</i> <i>but you see</i> <i>now you see</i>
<i>I think</i>	<i>Well I think</i> <i>I think I think</i>	<i>I mean I think</i>
<i>now</i>	<i>but now</i> <i>you know now</i> <i>well now</i>	<i>but now</i> <i>okay now</i> <i>so now</i> <i>right now</i> <i>well now</i> <i>yes (yeah) now</i>

The above subsections have summarised the analyses of the eight DMs for analysis from their frequency comparisons, typical positions in an utterance/turn, the most frequent types of co-occurrence and functions and the co-occurring DMs. The next subsection reviews the the use of *Linear Unit Grammar* (Sinclair and Mauranen 2006) in this thesis.

12.2.6 The usefulness and limitations of *Linear Unit Grammar* in the analysis

The use of *LUG* (Sinclair and Mauranen 2006) is one of the innovative aspects of the thesis and this innovation deserves extended discussion in this section. One of the two uses of *LUG* in this thesis is to assign units in speech making it possible to indicate the point where DMs occur in an utterance (see Section 2.5 of Chapter 2 for some detailed discussion). This has met with success. Since one of the main characteristics of DMs is flexibility of position (see Section 2.3.2 for a discussion of the characteristics and definitions of DMs), DMs can appear anywhere in an utterance. This makes it less easy to describe their positions by means of traditional hierarchical grammars. The step before labelling units is chunking. The ability to chunk an utterance/text is assumed by Sinclair and Mauranen to be possessed by speakers/analysts. The *LUG* analysis presupposes that an utterance will always be divided into chunks in similar ways. However, some doubts have to be raised whether NNSs (without prior linguistic research training) can process English and perceive the divisibility of text as NSs can. This is without doubt an interesting area to investigate in the future.

The other use of *LUG* in this thesis is to help distinguish non-discourse use (Type A) and discourse use (Type B) of the words/phrases under investigation in a different context. In the *LUG* analysis, Type A words/phrases are categorised into an M- element and Type B into an OI element. As demonstrated in Section 2.5 of Chapter 2 and the analysis chapters (Chapters 4 to 9), the distinction between M- and OI in the *LUG* analysis is not always clear-cut. The analyses of one-word DMs, *like*, *well* and *now*, appear to be less complicated and less problematic than those of two-word DMs, *you know*, *I mean*, *you see* and *I think*. Type A *like*, *well* and *now* in the *LUG* analysis are part of a message-oriented element, for example, *I like to* (M-) in Excerpt (4.2.9), *he was well* (M) in Excerpt (6.2.5) and *right now* (MS) in Excerpt (9.2.2), while Type B *like*, *well* and *now* are a single OI element. In the case of two-word phrases, both those of Type A *you know*, *I mean*, *you see* and *I think* and of Type B are a single element, in which an M- element needs to be distinguished from an OI element. This judgement has to be based on the analyst's interpretation of the discourse and context where the element is used. Therefore, the distinction between M- and OI is sometimes unavoidably subjective.

As noted above, it is problematic to use *LUG* to make the distinction between Type A and Type B in the case of two-word phrases. Among the four two-word phrases under investigation, the classification of *I think* in *LUG* is most difficult. *I think* as an M- element

often follows a personal opinion in a reported clause, classified as a +M element. *I think* as an OI element also co-occurs with personal opinions. The identification of the second case, as illustrated in Example (2.5.1) in Chapter 2, depends on the judgement whether the co-occurring elements give strong grounds for classifying *I think* as M-. Sinclair and Mauranen (2006: 75) point out that prosodic information would probably facilitate the classification; however, they decide not to make use of it in the initial stage of analysis, but claim that their classification of *I think* as OI is supported by some researchers' findings (e.g. Stenström (1994) and Aijmer (2002)) that *I think* is typically used as a DM rather than a reporting clause. In contradiction to this argument, this thesis finds that *I think* is not primarily used as a DM. This is a result of the types of activity in which *I think* occurs. For instance, in the sub-corpus of Chinese NNSs' dialogues, the speakers are asked to exchange opinions and this leads to the frequent use of *I think* as M- and not as OI.

Overall, *LUG* is a useful device for assigning units in speech and placing the decision of non-discourse use and discourse use of the words/phrases within a different kind of context, which makes a clear binary distinction between message-oriented and organisation-oriented. For making distinctions between non-discourse examples and discourse examples, in cases (i.e. *like*, *well* and *now*) where the distinction is easy, *LUG* gives another terminology but does not make this distinction easier. In the borderline cases (i.e. *you know*, *I mean*, *you see* and *I think*) the distinction is often controversial, because the identification requires larger sections of discourse and more subjective interpretations. Applying the *LUG* analysis does not make this process objective, but provides another way to clarify the process of making the distinction.

12.3 The use of discourse markers

There are two major difficulties to face in the investigation of the use of DMs. One is in differentiating between the non-discourse use and the discourse use of the words/phrases and the other is in the identification of functions of DMs.

It is discussed in Chapter 2 that although there are some clear-cut examples of DMs (e.g. *well* and *you know*), the distinguishing characteristics of DMs are still not clear enough to give a precise definition of them. Due to the lack of clarity in their definition, it is sometimes difficult to distinguish non-discourse uses from discourse uses of some words/phrases. The *LUG* analysis is of some use in the classification, since it accommodates DMs in one of the

two major categories. (See Section 2.5 for more details of *LUG* and Section 12.2.6 for its usefulness and limitations.)

It is argued in this thesis that it is difficult to demonstrate unequivocally the function of any instance of a DM, because DMs are produced without conscious knowledge and their functions cannot be easily identified by either speakers or researchers. In the light of this, the collocation phenomena surrounding DMs are used to determine the co-occurrence categories with the suggested functions being secondary interpretations. It is acknowledged that this process, which sometimes calls for unavoidably subjective judgements, does not always reliably identify the functions of DMs. However, this research approach begins with linguistic evidence for logically progressing to the identification of functions.

In this thesis, the investigation of DMs benefits from a combination of quantitative corpus methodologies and qualitative text-based analyses. The similarities and differences in the use of DMs between the two types of genre and between the two groups of speakers are identified and the factors, such as genres, contexts, types of activity and identities of the speaker in the use of DMs are discussed. All such factors affect the speakers' choice of a DM to use when giving priority to discourse organisation, fluency, the engagement of listeners, the construction of the speaker's persona and the creation of solidarity.

The two general hypotheses proposed in the introductory chapter are: 1) across the monologic and dialogic genres under investigation, the more interactive the genre or activity type is, the more DMs occur and 2) the uses of DMs in the Chinese NNSs' speech are different from and are possibly not as varied as those in the NSs' speech. The sections below first discuss the findings which confirm my hypothesis that the use of DMs is genre- and context-dependent and then report the different uses of DMs in the Chinese NNSs' and NSs' speech.

12.3.1 Factors in the use of discourse markers: genre dependency, context and activity type sensitivity and identity construction

The corpus methodologies and text-based analyses demonstrate that the use of DMs correlates with genres, contexts, types of activity and identities of the speaker. In the analyses of the DMs, it is found that all the DMs under scrutiny (except *now*) occur more often in the dialogic genres than in the monologic genres. For a frequency comparison between the two types of genre, the log-likelihood test (see Appendix 6) and *z* test for two proportions (see Appendix 7)

are calculated to assess the significance of differences. Except the case of *like* in SECCL and *now* in SECCL and MICASE, the values indicate significance between the monologic genres and dialogic genres as well as under-representation in the monologic genres. This supports my hypothesis that the use of DMs is affected by genre. The more interactive the genre is, the more often DMs occur.

In addition to genre, context and type of activity are also factors in the use of DMs. For example, DM *like* occurs much more frequently in the American NSs' highly interactive discourse mode than in the highly monologic discourse mode (84.4 vs. 1.1 times per 10,000 words). It is used mostly by fellow students in informal contexts, such as study group discussion, rather than by faculty members in lectures. Another example is *oh*. *Oh* is used more than four times as frequently in the NNSs' dialogues as in the monologues and it is more than twenty times as frequent in the NSs' dialogic genres as in the monologic genres. The use of *oh* is found to be context sensitive. About two thirds of the instances of *oh* in the NSs' unscripted monologues in ICE-GB are used to show the speaker's emotions. A further look reveals that almost all (23 out of 26) instances are from sports commentaries, in which commentators use *oh* to show their emotions.

In the case of DM *well*, there are marked differences in the distribution of the types of co-occurrence of *well* across the six sub-corpora. This can be attributed to the variations in the type of activity. For example, the NNSs' monologues are mainly accounts of personal experience, in which there are more opportunities for using *well* to mark reported speech. Their dialogues are for exchanging opinions and therefore more chances are for the use of *well* prefacing disagreement to soften the speech (see Appendix 1 for the topics of the NNSs' monologues and dialogues).

Nevertheless, *now* is a rather different DM. In the NSs' speech, *now* occurs more frequently in the monologic genres than in the dialogic genres. The identification of co-occurrence reveals that *now* is mostly used a boundary marker, indicating shifts of (sub)topic and viewpoint. This function is of particular use in a lengthy monologue.

It is evident that the use of DMs is genre-dependent and context-sensitive. In addition to these two factors, the use of DMs relates to the identities and personae which the speakers intend to construct in speech. Take Excerpt 10.8 in Chapter 10 for example. Speaker 1's use of disciplinary terminology constructs her identity as a competent student and also as a group leader and her use of DM *like* constructs the persona of American young people to show

solidarity with her peers.

12.3.2 Different uses of discourse markers in the Chinese NNSs' and NSs' speech

In general, there are more similarities than differences in the use of DMs between the Chinese NNSs' speech and NSs' speech. Among the eight words and phrases for analysis, *like*, *oh*, *well*, *you know*, *I mean*, *you see*, *I think* and *now*, the Chinese NNSs seem to achieve the non-discourse use (Type A) as well as the discourse use (Type B or DM) of *oh*, *well*, *you know*, *I mean* and *you see* in a similar way to those of the NSs, but with different frequencies across the six sub-corpora under investigation. This does not fully support my hypothesis that the uses of DMs in the Chinese NNSs' speech are different from and are possibly not as varied as those in the NSs' speech.

However, some uses of DMs in the Chinese NNSs' speech are different from those in the NSs' speech. Some of them may be attributed to generic constraints. An example of this is that *I mean* is under-represented in the NNSs' speech. One contributing factor in the under-representation of *I mean* in the NNSs' speech may be the nature of test language. Similar topics to the NNSs' speech under investigation are likely to be practised before the recording, which leads to less use of *I mean* as a DM co-occurring with clarifications, explanations and elaborations (the major types of co-occurrence with *I mean*). This could result in the under-representation of *I mean* in the NNSs' speech.

Another example of the generic constraints is that *now* prefacing a question is relatively frequent in the NSs' highly monologic discourse mode in MICASE. It is likely that in this sub-corpus, the main speakers in each text are mostly lecturers, who have a position of power in the discipline. In this particular context, *now* prefacing a question is primarily used as a boundary marker to shift (sub)topics. The speakers do not usually expect an answer or any response from the students. In contrast, in conversations, speakers could use *well* and *I think* to preface a question in order to sound less direct and imposing.

Some differences in the use of DMs between the NNSs and NSs resist reasonable explanation. The investigation of *like* (Chapter 4) reveals that the NNSs tend to employ Type A *like* in their speech and Type B *like* represents only 1.7 and 3 per cent of the instances of *like* in the monologues and dialogues respectively. In contrast, Type B *like* is highly represented in the NSs' dialogic genres in MICASE and ICE-GB. It can be argued that the Chinese NNSs perceive *like* as an inappropriate DM in the test-taking setting and therefore

provide almost no instance of it. However, it is more likely that the Chinese NNSs do not know how to use *like* as a DM, while using the other DMs under investigation.

Another distinction can be made relates to *oh* as a (preface to a) response to a question and to new information. Both the NNSs and NSs use *oh* in turn-initial position as a (preface to a) response. Further examination reveals that the Chinese NNSs tend to use *oh* as a (preface to a) response to a question and the implications conveyed by *oh* are probably different from NSs' understanding of *oh*. In the NS speech, *oh* is used as a marker of change-of-state, indicating "a problem about a question's relevance, appropriateness, or presuppositions" (Heritage 1998: 294-295). However, it is found that the Chinese NNSs use *oh* in a neutral tone as a token of acknowledgement.

In the case of *you know*, the NNSs and NSs use *you know* in similar ways, but there are differences in its position in an utterance/turn. Most of the instances (86.4% in the monologues and 91.9% in the dialogues) in the NNSs' speech are placed in an extra-clausal position, whereas about half the instances in the NSs' speech occur in intra-clausal positions. The NNSs' awareness of *you know* in an intra-clausal position could be raised to facilitate the process of comprehension.

The way to signal a repair varies in the NNSs' and NSs' speech. In the speech of the NSs, *you know* and *well* are used to mark a repair. In Müller's study (2005) of the German NNSs' and American NSs' speech, she finds that almost one third of the NSs use *you know* between one and three times to mark a repair, whereas only 3 (3.8%) out of the 77 NNSs use *you know* once to mark a self-repair. Likewise, in the present study, the Chinese NNSs seldom use *you know* to signal a repair and yield only one instance of *well* marking a repair. They are found to use *I mean* to signal a repair. The category of repairs accounts for 19.6% of the instances of Type B *I mean* in the NNSs' monologues and 11% in the dialogues, as opposed to 3.3% on average in the NSs' speech. In addition to using *I mean* to mark a repair, the NNSs tend for a self-repair to simply restart their utterance rather than using a DM to signal it. This way is non-native-like, as the NSs would use such DMs as *well* and *you know* to orient listeners to the coming repair.

The speech of the NNSs and NSs shows a marked difference in the use of *I think* as a DM. In terms of the positions in an utterance/turn of Type B *I think*, in the NNSs' speech, *I think* seldom occurs in an intra-clausal position, whereas in the NSs' monologues, more than half the instances of *I think* occur in an intra-clausal position. In terms of the types of

co-occurrence, in the NNSs' speech, most of the instances of *I think* (63.5% on average) co-occur with hesitation markers, pauses and restarts. This could suggest that the NNSs use *I think* as a filler in their speech. Some of the NSs also use *I think* in this way, but the percentage is much lower (18% on average). In the NSs' speech, factual information is the most frequent type of co-occurrence, representing 45% and 28.6% respectively of the instances of Type B *I think* in the two sub-corpora in MICASE and 41.9% and 50.9% respectively in the two subsets in ICE-GB. However, *I think* co-occurring with factual information is seldom used by the Chinese NNSs: only two instances in their monologues.

You see is another non-native-like DM. The phrase *you see* is not primarily used as a DM in the NSs' speech, but 74.4% and 89.2% of the instances of *you see* are used as DMs in the NNSs' monologue and dialogues (see Table 12.1 above). As with the use of *I think*, *you see* is found to co-occur frequently with hesitation markers, pauses and restarts in the NNSs' speech, accounting for 10% of the 403 instances of Type B *you see*. Only two instances of it are found in the NSs' speech.

The identification of co-occurrence of DMs with corpus methodologies is used as the basis for interpreting their functions. The text-based analysis, as a complementary view, looks at the speech of high users and low users of DMs. More interpretations are made of the ways in which DMs are used, with reference to type of activity and the relationship between speakers. For example, the NS high users of DMs use *oh* to show emotions when the type of activity (e.g. sports commentary) requires it. The NS low users of DMs (e.g. a lecturer in the classroom setting) seldom use DMs but employ devices signalling listeners' engagement. In some contexts, it is arguably better to use explicit devices to replace the function of a simple DM. The Chinese NNS high users of DMs, in a test-taking context, are not necessarily fluent speakers, but their use of DMs makes the utterance sound more natural, whereas the speech of non-users of DMs sounds like formal written English.

12.4 Implications for the investigation of discourse markers in a non-native variety of English

The different uses of DMs in the Chinese NNSs' and NSs' speech are discussed above. This raises the question of how we see the differences in the use made by Chinese NNSs and NSs of DMs. There are two possible, though contradictory, arguments as to whether it is necessary for NNSs to speak in a native-like way. Both of these have some appeal. First, there is an

argument that it is practically unnecessary for NNSs to take the NS usages as a target norm, considering that NNSs outnumber NSs (Graddol 1998) and the population of the former is still rapidly growing and may reach two billion people by the year 2020 (Graddol 2006: 62). For NNSs, there are many more opportunities to communicate in English with other NNSs with different L1 backgrounds than with the equivalent NSs. Furthermore, with regard to grammar, there is nothing right and wrong in using DMs. It is inevitably to some degree a matter of ideology. The NNS usages of DMs should be acknowledged as an acceptable feature in its own right of an emerging variety of English. However, it can equally be argued that NNSs should use DMs in a native-like way and this argument takes NS usages as the target norm for NNSs who want to improve their speech. This point is discussed in the next section.

As noted above, as the number of Chinese users of English is growing, it seems that NSs can acknowledge the use that Chinese speakers make of DMs, which may facilitate their ability to use interpretation when they communicate with speakers whose L1 is Chinese. Some uses of DMs are distinctive in the Chinese NNSs' speech. For example, as demonstrated in the analysis of *oh*, DMs are probably culturally specific and contextually dependent. Other examples in the Chinese NNSs' speech are not using *like* as a DM; using the phrase *you see* as primarily a DM and co-occurring with hesitation markers, pauses and restarts; over-representing the phrase *I think*. It is arguable that these distinctive uses of DMs can be acknowledged as a feature of a non-native variety of English.

This research offers insights into one variety of global English. As the number of Chinese expert users of English is growing, it is worth considering possible ways to address the role of NSs in acknowledging this English variety and in some cases to improve the communication between NSs and Chinese speakers of English.

In addition, the issue of whether NNSs should speak English like NSs is still under debate. Learners should not be punished for not speaking like NSs and, if they desire, they can keep their cultural identity, as long as the NNS language does not cause any misunderstanding or impede the development of interpersonal relations. Meanwhile, it is of importance for NNSs to be aware of how NSs use DMs in certain contexts.

The argument thus far is that, on the one hand, the global community of English speakers should be educated to accept the use of DMs in a local variety of English, for example, the subtle implications of *oh* and the frequent use of *I think* in the Chinese NNSs' speech. NSs are encouraged to have an attitude of tolerance towards some areas in the NNSs' use of DMs.

On the other hand, the Chinese NNSs (as well as other NNSs of different L1s) should be instructed to become aware of NSs' usages of DMs and to use DMs differently, taking NSs' usages as a target norm in order to improve interaction and prevent misunderstanding in communication with NSs. On this aspect of reality, this research certainly sheds some light on pedagogy in ELT, an area which is further discussed in the next section.

12.5 Implications for pedagogy in English language teaching

We now turn to the second argument mentioned in Section 12.4 above. This seeks to address the practical needs of the learners of English in a more practical way. As argued in the previous section, it is probably not necessary for NNSs to sound like NSs in their use of DMs when they communicate with other NNSs of different L1s. Admittedly, it is an ideal situation that every group of English speakers should use English as the way they do and other groups will be able to adjust to it. In practice, however, some measures have to be taken in the teaching and learning of English in Chinese-speaking countries. It makes little sense to tell learners that they can simply speak English in any way they like so long as they keep their cultural identity. For one thing, many learners of English wish to speak in a native-like way (Timmis 2002). In addition, there are certainly those (e.g. Svartvik (1980: 171), Erman (1987: 1) and Fung and Carter (2007)) who believe that an inappropriate use of DMs may cause misunderstanding and lead to negative effects in communication. Therefore, it is worth raising the Chinese learners'²⁵ awareness of the uses of DMs among NSs and enabling them to use DMs appropriately, as well as helping them to speak like NSs.

In this thesis, the investigation of the eight DMs reveals surprising similarities between the speech of the Chinese learners and NSs. Some differences are also identified (see Section 12.3.2) and, based on these findings, some areas are suggested for pedagogical interventions. Some approaches to the teaching of DMs are also recommended in this section.

12.5.1 Pedagogical aspects of discourse markers

With limited exposure to naturally-occurring spoken English among the Chinese learners and few opportunities to use DMs in their classroom discourse, I would argue that creating space in the classroom for the teaching and learning of DMs is necessary. Teachers should evaluate, depending on learners' needs, to what extent learners have to understand NSs' use of DMs and

²⁵ This section talks about pedagogy, in which it makes more sense to refer to the Chinese NNSs under investigation as learners. See more discussion in Section 1.4 of Chapter 1.

to speak in a native-like way. The following sections examine some pedagogical aspects of the use of DMs and approaches to the instruction of the use of DMs in the classroom setting. The suggested approaches and activities below aim to raise learners' awareness and to enhance their understanding in order to improve their receptive competence. Their productive competence of using DMs, as McCarthy (1998: 60) maintains, should be allowed room to be displayed in a more natural context in the future, rather than in immediate production in the classroom.

As noted in the previous chapter, how the Chinese NNSs learn the use of DMs and why they frequently use DMs in the exam context remain a mystery. It is probable that DMs can be acquired without consciously learning them in the classroom. Nevertheless, the corpus studies and text-based analyses of the NSs' and NNSs' speech reveal some NNS usages for NSs to acknowledge and some areas for NNSs to become aware of.

The selection of the types of discourse features to be taught is based on local contexts, such as learners' age, proficiency level and the needs and objectives of the language programmes. As argued earlier, not all learners need to be native-like in using DMs, but some learners would certainly benefit from understanding DMs better and using them appropriately. In this section, the relevant research outcomes are highlighted for pedagogical use.

12.5.1.1 Making learners aware of the use of discourse markers in speech

In most contexts where English is used as a foreign language, the mode of written English has been the norm, probably firmly rooted in learners since the outset of their English learning. Learners have a bias towards the grammar of written English. Because of this, I would anticipate a certain degree of difficulty in addressing the issue of learners' expectations and prejudices. It may be a good choice to provide authentic data as strong evidence and to begin with the features of spoken English and then introduce DMs as a prominent feature in speech.

To make learners aware of the use of DMs in speech, a quick starting point can be a consciousness-raising activity (see Section 12.5.2.1 below for more details) of comparing an academic word list (e.g. Coxhead's *academic word list* (1998, 2000)) and frequent words in spoken English (e.g. the most frequent words in spoken English in O'Keeffe *et al.* (2007: 35)). It is clearly evident that words, such as *yeah*, *so*, *like*, *well* and *right*, listed in the first 50 most frequent words in the 5-million-word spoken section of the CANCODE corpus (O'Keeffe *et al.* 2007: 33-36), are absent from an academic word list. These words are likely to be used as

DMs by NSs. The Chinese learners under investigation in this thesis, however, are more likely to employ the non-discourse uses of these words (referred to as Type A in this study), in particular *like* and *well*. The learners seem to be competent in using *like* as a verb and a preposition and *well* as an adverb in their speech. They seldom use *like* and *well* as DMs. It is also probable that they do not know how to use them as DMs. When NNSs encounter the use of DMs in real life, they are likely to experience, to some extent, linguistic shocks, because these words are frequently used as DMs in the NS speech.

In addition to consciousness-raising activities, *Linear Unit Grammar (LUG)* (Sinclair and Mauranen 2006) may be used as a new tool for learners (see Section 2.5 in Chapter 2 for a detailed description of *LUG* as an approach to the description of spoken English). The use of *LUG* in pedagogical grammar has so far, to my knowledge, remained un-explored. The researchers maintain that some of the procedures are of use for all learners. They claim that naturally-occurring language in real situations should be used to train learners' ability to chunk and argue that the distinction between an organisational (O) element and a message-oriented (M) element is crucial. Distinguishing between O and M elements can help learners to notice the contributions to organisation and interaction of O elements. In the situations where learners try to understand propositions, interactive organisational (OI) elements can be ignored. In situations where speakers' attitudes and feelings are stressed, OI elements should be attended to (Sinclair and Mauranen 2006: 163-164).

Excerpt (4.5.1) in the analysis of *like* (Chapter 4), repeated here as Excerpt (12.5.1), would provide good learning material for some students (e.g. international students in the United Kingdom and the United States). In this excerpt, *like* as a DM is frequently used by Speakers 2 and 3, who are American undergraduate students. If there are NNS students in the discussion group, they may have comprehension problems in figuring out what these instances of *like* mean, in the meanwhile missing out the main message. One possible activity for Excerpt (12.5.1) is asking students to circle words which do not carry literal meanings (i.e. OI elements), as shown in Excerpt (12.5.1). The clean version, Excerpt (12.5.2), delivers the proposition better than the original one. Some questions and discussion can be followed up, such as why the OI elements are used and if learners of English would like to speak in this way. Some learners may consider imitating the use of *like* in the NS speech in order to express solidarity and integrate with fellow students.

(12.5.1)

.....

- S2: maybe you won't be able to
[yeah] i might not be able to
or i need to see if [like] someone else could [like] kinda just take it on right now
or, if Frank if i could be like oh Frank will you just write it?
or something like that [you know]. uhuh mhm,
- S3: [yeah]. [well] [so] i got the extension from him
and then i finally just ended up talking to my professor on North Campus about it.
[and] i told her what i was interested in doing,
and then i was like could i use the research?
and she was like cuz there's several different parcels of research [like],
there's [like] five different communities i think in Detroit that they did interviews with
and [like] s- [like] (peer g-) the diff- there're different characteristics for all of 'em.
some of 'em are [like] industrial areas
some of 'em are retail areas,
and so some of them a- um all of them are gonna be going into the research that she does at the end of
the year and stuff
but also [like] she went to [like] different neighborhood organizations [like], maybe [like] um, employment
organizations and things in the area
and was like do you guys want some of the information that we get from this stuff?
[like] cuz it would probably be
- S3: <OVERLAP2> mhm </OVERLAP2>
- S1: <OVERLAP2> yeah </OVERLAP2>
- S3: <OVERLAP1> helpful to </OVERLAP1> them.
so that's what i've been working on a lot for her last year was doing that for a couple um [like] uh,
industrial organizations around the area
and some of them haven't been done at all.
so she wanted me to do some of those um for, my project.
[so] then <EVENT WHO="SS" DESC="LAUGH" />
[so] then i would do them and i would put that in my project.
[like] i would have a part that was like this is empowerment zones in Detroit.

.....

(MICASE: SGR999SU146)

(12.5.2)

.....

- S2: maybe you won't be able to
i might not be able to
or i need to see if someone else could kinda just take it on right now
or, if Frank if i could be like oh Frank will you just write it?
or something like that. uhuh mhm,
- S3: i got the extension from him
and then i finally just ended up talking to my professor on North Campus about it.
i told her what i was interested in doing,
and then i was like could i use the research?
and she was like cuz there's several different parcels of research,
there's five different communities i think in Detroit that they did interviews with
and s- (peer g-) the diff- there're different characteristics for all of 'em.
some of 'em are industrial areas
some of 'em are retail areas,
and so some of them a- um all of them are gonna be going into the research that she does at the end of
the year and stuff
but also she went to different neighborhood organizations, maybe um, employment organizations and
things in the area

and was like do you guys want some of the information that we get from this stuff?
 cuz it would probably be
 S3: <OVERLAP2> mhm </OVERLAP2>
 S1: <OVERLAP2> yeah </OVERLAP2>
 S3: <OVERLAP1> helpful to </OVERLAP1> them.
 so that's what i've been working on a lot for her last year was doing that for a couple um uh, industrial
 organizations around the area
 and some of them haven't been done at all.
 so she wanted me to do some of those um for, my project.
 then <EVENT WHO="SS" DESC="LAUGH" />
 then i would do them and i would put that in my project.
 i would have a part that was like this is empowerment zones in Detroit.

.....

(MICASE: SGR999SU146, revised, taken out discourse markers)

12.5.1.2 Areas which may require pedagogical interventions

The corpus-based investigation of DMs across the monologic and dialogic genres shows that there are more occurrences of DMs in the dialogic genres and reveals differing usages across the two types of genre. The qualitative text-based analyses indicate that the use of DMs are dependent on genre and activity type and have a connection with the speakers' construction of relationship and creation of solidarity. These findings imply that context-appropriateness in using DMs, rather than frequency, should be the focus in the pedagogical interventions, if needed.

The frequency information derived from corpus studies tell us that certain types of co-occurrence of DMs are most frequent in an NS corpus, but this does not necessarily raise the use to prominence in a beginners' class or an introductory session to DMs. Similarly, when corpus studies tell us that a certain use of DMs is not frequent in itself compared with other uses, it is not a reason for ignoring this use. It is suggested that the pedagogical interventions of the use of DMs should aim to enable learners to use DMs appropriately on the basis of their priorities, competing with fluency, the construction of relationships, creation of solidarity, etc.

Based on the analyses of the DMs, the following six areas of the uses of DMs are suggested for pedagogical interventions to raise learners' awareness of NSs' usages of DMs and to help them sound native-like if they wish to.

One of the areas that the Chinese learners may improve is using DMs to signal a repair. It is found that the Chinese learners seldom use DMs to signal a repair but merely pause for thought and restart, whereas NSs would use a DM (e.g. *well* and *you know*) to give listeners a signal that a correction is coming.

DMs prefacing *dispreferred* responses is another use that the Chinese learners can give

attention to. It is common for the Chinese learners to give *dispreferred* responses without any preface. This kind of direct speech tends to be interpreted by NSs as aggressiveness, over-assertiveness and a lack of consideration for people's feelings. Nevertheless, in Chinese, the learners' L1, *dispreferred* responses are usually prefaced with hesitation markers and DMs just as those in English are by NSs. Instruction may begin by referring to learners' communication strategies in their L1 and then demonstrate how NSs use such DM as *well* and hesitation markers to introduce *dispreferred* responses in order to make speech less direct and face-threatening.

Although it is nothing to do with right or wrong grammars if the Chinese learners frequently use *I think* in their speech, they might use it at the expense of other alternatives and this probably requires some pedagogical interventions for them. Chinese learners tend to use *I think* with personal opinions, whereas NSs use *I think* for hedging (O'Keeffe *et al.* 2007: 174-176). The Chinese learners' awareness can be raised for the various alternatives to *I think* in the NS speech.

The positions in an utterance/turn of DMs may also be an area for pedagogical intervention. In the case of *you know*, most of the instances (86.4% in the monologues and 91.9% in the dialogues) in the Chinese learners' speech are placed in an extra-clausal position, whereas about half the instances in the NSs' speech occur in an intra-clausal position. Because the difference in proportion is huge, it is likely that Chinese learners are unfamiliar with *you know* in intra-clausal positions; therefore, their awareness can be raised to facilitate the process of comprehension when they talk with NSs.

Similarly, the NS use of *like* is another area to introduce to learners to aid their process of comprehension. Since *like* as a DM is a fairly recent use, it is suspected that the Chinese learners do not know how and when to use *like* in this way and they lack familiarity with the NSs' usage of *like*. It is probable that the NSs' constant use of *like* distracts learners' attention from the proposition that the speaker aims to deliver. In the NS data, it is found that DM *like* is used as an in-group marker to express solidarity in certain groups. In some contexts, learners may need to communicate in English with NSs and assert in-group membership, for example, overseas students in programmes of English for Academic Purposes in English-speaking countries. For this group of learners, instruction can focus on raising learners' awareness of NSs' use of *like* and further encourage learners to observe the use of *like* in the group they wish to integrate with. The native-like use of *like* probably enables

learners to express their solidarity. As Cutting's study (2006: 177) on vague language suggests, learners' precise speech may create a barrier to communicating with native fellow students. This can also be true of the use of DMs.

The last area that probably requires pedagogical intervention is not directly related to the use of DMs but the language surrounding them. In the analysis of *oh*, it is found that *oh* co-occurs with indicators of *misplacement* in both Chinese learners' and NSs' speech. It is found that the language of *misplacement* in the Chinese learners' speech may be interpreted as too direct (e.g. *I have another point*), while the NSs' language is more hedging (e.g. *the other thing I wanted to just mention*). Other less direct uses of language for softening speech can be instructed together with the use of DMs.

12.5.2 Pedagogical approaches to the instruction of discourse markers

It is generally believed that language teachers adapt an eclectic approach catering to learners in context in order to maximise learning outcomes. This underlying principle is also applied to the instruction of DMs. As pointed out in the previous section, not all learners need to use DMs in a native-like way and it is very likely that DMs cannot be taught but are acquired unconsciously by learners. Because of these arguments, some teaching methodologies for grammatical items may not be appropriate for DMs. The following sections suggest two approaches, consciousness-raising activities (Willis and Willis 1996) and data-driven learning (DDL) (Johns 1991) and controlled exercises.

12.5.2.1 Consciousness-raising activities and data-driven learning

Willis (2003) argues strongly that distinctions between spoken and written English should be taught and that consciousness-raising activities are the best approach, using naturally-occurring data. Consciousness-raising activities (Willis and Willis 1996: 64) encourage learners "to notice particular features of the language, to draw conclusions from what they notice and to organize the view of language in the light of the conclusions they have drawn". This exercise on transcripts helps learners to notice features which may be ignored in the process of real-time communication (Willis and Willis 1996: 75-76).

Due to the lack of teaching materials on the grammar of spoken English, Hobbs (2005) takes transcripts and recordings of NSs performing tasks from a coursebook and uses them in the cycle of a Task-based Teaching and Learning (TBTL) framework so as to raise students'

consciousness of the differing use of DMs and other features in conversations between students and NSs. Jones (2001) reports a similar approach to raising learners' awareness about spoken narrative structures. Two versions of a transcribed narrative are available to learners. One of these is a simple narrative and the other contains elaboration and evaluation from a main story-teller and an active listener. A series of consciousness-raising activities on these two versions of narratives helps learners to develop storytelling ability.

A similar procedure can be followed to raise learners' awareness about using DMs. To demonstrate how NSs use *like* as a DM, two versions of one excerpt from MICASE are provided. One (see Version 2 in Table 12.6 below) is the original transcript and the other (see Version 1) is adapted to take out all instances of *like* as a DM.

Learners are asked to read Version 1 first. They are then given Version 2 and asked to highlight the differences, which are the instances of *like*, highlighted here for convenience's sake. Contextual information, such as type of activity (i.e. senior thesis study group discussion) and the roles of the speakers (i.e. senior undergraduates, female NSs, age ranging between 17 and 23 years old), should be provided for learners to make interpretations of the use that NSs make of *like*.

Table 12.7: Example material for consciousness-raising activities

Version 1	Version 2
<p>S2: maybe you won't be able to yeah i might not be able to or i need to see if someone else could kinda just take it on right now or, if Frank if i could be like oh Frank will you just write it? or something like that you know. uhuh mhm,</p> <p>S3: yeah. well so i got the extension from him and then i finally just ended up talking to my professor on North Campus about it. and i told her what i was interested in doing, and then i was like could i use the research? and she was like cuz there's several different parcels of research, there's five different communities i think in Detroit that they did interviews with and s- (peer g-) the diff- there're different characteristics for all of 'em. some of 'em are like industrial areas some of 'em are retail areas, and so some of them a- um all of them are gonna be going into the research that she does at the end of the year and stuff but also she went to different neighborhood organizations, maybe um, employment organizations and things in the area and was like do you guys want some of the information that we get from this stuff? cuz it would probably be </p>	<p>S2: maybe you won't be able to yeah i might not be able to or i need to see if like someone else could like kinda just take it on right now or, if Frank if i could be like oh Frank will you just write it? or something like that you know. uhuh mhm,</p> <p>S3: yeah. well so i got the extension from him and then i finally just ended up talking to my professor on North Campus about it. and i told her what i was interested in doing, and then i was like could i use the research? and she was like_ cuz there's several different parcels of research like, there's like five different communities i think in Detroit that they did interviews with and like s- like (peer g-) the diff- there're different characteristics for all of 'em. some of 'em are like industrial areas some of 'em are retail areas, and so some of them a- um all of them are gonna be going into the research that she does at the end of the year and stuff but also like she went to like different neighborhood organizations like, maybe like um, employment organizations and things in the area and was like do you guys want some of the information that we get from this stuff? like cuz it would probably be </p>

(MICASE: SGR999SU146)

Such consciousness-raising activities as above provide authentic English to learners to spot the use of DMs in the NS speech. When given the activity, learners should be informed of the objective of the instruction, which can be either raising their awareness or making them sound like NSs.

The consciousness-raising activities arguably involve one interpretation of DDL. In a sense, learners are presented with a certain amount of data and the data drive learning. Traditionally, DDL has been used in conjunction with concordance lines, which are directly linked with corpora research.

Johns's DDL approach (1991) involves three stages: the observation of evidence arranged in concordance lines, classification of salient features and generalisation of rules. Some may argue that reading concordance lines is likely to be unfamiliar to learners and indeed to some language teachers. In spite of this, the practice is backed up by the notion of

noticing (Skehan 2001, Tognini-Bonelli 2001) and previous studies (e.g. Lynch (2001)) favour form-focused learning. The DDL approach is inductive and learner-centred, thereby training students to become better language learners (Johns 1994). This enables them to interpret the use of DMs when they encounter them outside the classroom.

If a multi-modal corpus is available, concordance lines of DMs as well as their aligned audio and video data can be offered to learners, acting as researchers, to observe the discourse environment where DMs are used by NSs. The video clips would provide such contexts as the roles of speakers, types of activity and settings, to facilitate learners' investigation and for further discussion. Otherwise, the concordance lines can be provided on screen or on paper.

The use of DMs is often context-dependent and therefore to make interpretations of their use requires larger co-texts, as when the use of *like* is shown in a text in Table 12.6 above. For some uses of DMs, a display of DMs in concordance lines makes the uses obvious. For example, Figure 12.1 presents the DMs *oh*, *well* and *you know* signalling reported speech in concordance lines. The use of DMs marking reported speech can be identified in their immediate contexts.

Figure 12.1: Concordance for *oh*, *well* and *you know* marking reported speech in MICASE (WordSmith 4, Scott 2004)

N	Concordance
1	she wasn't as like, forward about things but she'd be like oh guys i have an idea for the bathroom to make it neater
2	see i_ students who say oh well i might as well just take, Honors Math because i
3	he's like yes and i'm like, oh and it just hit
4	we're gonna make ourselves look (bad.) (and they're) like oh yeah? (xx)
5	(xx) (xx) like somehow goes oh he thinks he's God's gift to women, well what if you know
6	that he can like (pick) up all the little pieces and be like oh that's the (da-da-da)
7	ying to remember it but, i think, i got out of it and he's like oh oh that's not that's not how you do it or something like
8	but, but then critics will sort of slam him and he'll say well let's play. let's leave it up to the music to decide. and
9	to this node? so you first would follow this path and say well , is this a connection or is there something that
10	need to meet with me about thi- but i go in and she's like well , Margie said that o- or, i guess that this woman had tried
11	point to disprove my theory. then it's just like saying, well nuh-uh and uh-huh a
12	because this is all (xx) so i'm i'm going around saying that you know , you're a big white bunny rabbit that, likes to lick
13	of the sign, so it didn't say, New York Daily News it just said you know the number, twelve. uh and the A-P protested
14	it go? chka chka chka chk chk right here it just says you know , no prerequisites just you know, it's a humanities.
15	tone but you know counselors often hear students say well you know i kind of understood the basic concepts but i had a

Corpus-related resources can be an effective tool for second language teaching and learning (Hunston 2002, Sinclair 2004a: 2). The above two approaches, consciousness-raising activities and DDL approach, will train students to be active and independent learners as well as prompting learners' autonomy. In particular the use of DMs is dependent on genre and

context. Actively involving learners in the classroom setting would enable them to acquire the use of DMs when they encounter DMs in other sources, such as films and conversations with NSs.

12.5.2.2 Controlled exercises

One disadvantage of the above two approaches is the source of material. It is very likely that teachers have to prepare their own materials. A quick way to introduce DMs to learners is using ready-made exercises. A number of corpus-informed coursebooks are available on the market. In these coursebooks, the actual language examples, although they may be adapted, reflect authentic language. The exercise below, taken from *Touchstone Student's Book 1* (McCarthy *et al.* 2006: 39), can be used to explicitly teach DMs or help learners recall the use of DMs.

The exercise below begins with some questions, such as *what does B say 'well' in the answer?* and *Is 'well' in the top 25, 50, 75, or 100 words?*, to direct students' attention to the use of *well*. Then, students practise the conversations with a partner and do it again with true answers (McCarthy, McCarten and Sandiford 2005: T-39).

Figure 12.2: Example material for the use of *well* (McCarthy, McCarten and Sandiford 2006:39)

Unit 4 *Everyday life*

2 Strategy plus *Well*

Start your answer with **Well** if you need time to think, or if your answer is not a simple yes or no.

Are you from California?

Well, I'm from Chicago originally, . . .

In conversation . . .

Well is one of the top 50 words.

About you → **Pair work** Practice the conversations. Then ask the questions again. Give your own answers.

- A What are your neighbors like?
B Well, they're very noisy.
They like loud music.
- A Do you see your family a lot?
B Well, not really. They don't live around here.
- A Do you study every day?
B Well, not every day. I go out with friends on the weekends.



The above sections have presented possible ways for the instruction of DMs. Whether such instruction is needed or not depends on the local context. Moreover, as noted by McCarthy (1998: 60), learners may not be asked for immediate production but may delay it until appropriate contexts appear.

CHAPTER 13: CONCLUSIONS

13.1 Introduction

This thesis examines a Chinese NNS corpus (SECCL) and two NS corpora (MICASE and ICE-GB) in order to shed light on the use of DMs across the monologic and dialogic genres and between the Chinese NNSs and NSs. Both quantitative corpus methodologies and qualitative text-based analyses are used. This chapter assesses the thesis, pointing to its achievements, strengths and limitations; it goes on to provide suggestions for future research and makes some concluding remarks.

13.2 Achievements and strengths of the thesis

The objective of this thesis, stated in Chapter 1, is using collocation phenomena and co-text analyses to empirically derive the functions of DMs rather than interpreting them intuitively. This thesis has demonstrated this by using quantitative corpus methodologies, combined with the *LUG* analysis (Sinclair and Mauranen 2006) and qualitative text-based analysis to identify the use made by Chinese NNSs and NSs of DMs. The functions of the eight DMs under investigation, *like, oh, well, you know, I mean, you see, I think* and *now*, have been empirically established, making a contribution to the investigation of DMs in the NNS speech and the modelling of NNS English.

The six sub-corpora extracted from SECCL, MICASE and ICE-GB are by no means comparable. However, it is less problematic to have comparability across the two types of genre in each of the three corpora and to test my hypothesis that the more interactive the genre or type of activity is, the more DMs occur. One great advantage, beyond my expectations, of using three different corpora is the result of identifying the factors in using DMs. If a small comparable NS corpus had been compiled, factors such as genre, type of activity and identity construction might not have been uncovered.

An innovative aspect of this thesis is the use of *LUG* (Sinclair and Mauranen 2006), which is a newly-established device for the description of language. At the time of writing, there has not been any study of NNS English using *LUG*. In this thesis, *LUG* has been used successfully to assign units in spoken English and describe the intra-clausal positions of DMs in both the speech of the NNSs and NSs.

A strength which makes this thesis distinct from most corpus studies is the use of

text-based analysis. Since the quantitative corpus-based approach to language has been criticised for its de-contextualisation, the qualitative text-based analysis, as a supplement, looks at wider discourse and contextual issues. It reveals some findings for and against some hypotheses which cannot be tested with corpus methodologies. The corpus study and text-based analysis can be seen as complementary approaches which can inform and enrich each other, thereby leading to a better understanding of the use of DMs.

The implications of this research for the investigation of non-native varieties of English and for pedagogy should benefit the English teaching for Chinese speakers of English, help raise learners' awareness, prevent misunderstanding between speakers and facilitate inter-cultural communication in English.

13.3 Limitations and weaknesses of the research

The argument thus far in this final chapter is that the broad aim of empirically identifying the use of DMs on the basis of their co-occurring evidence. Additionally, some factors influencing the (non-)use of DMs have been identified in both corpus-based investigation and text-based analyses. Nevertheless, no approach is without its weaknesses and neither kind of research is free of limitations. In this section, I would like to acknowledge six limitations and weaknesses of this thesis. The first two limitations relate to the corpora in use. The remaining four weaknesses are about the methodology used in this thesis.

The first limitation relating to the corpora under investigation is that using three corpora with mark-ups and annotations in different ways makes it difficult to conduct an investigation across corpora with regard to certain questions. The Chinese NNS corpus (SECCL) under investigation clearly identifies speaker change in the texts, but similar information in the two NS corpora (MICASE and ICE-GB) is included in the mark-up and cannot be searched for. For instance, if a group of speaker use a particular DM (e.g. *well* co-occurring with hesitation markers, pauses and restarts) in turn-initial position more often than another group of speakers does, it is difficult to find the answer but the topic remains for future research.

The second limitation is inherent in the Chinese NNS data, which are contrived and collected in a restricted test-taking setting. The speakers' use of DMs is likely to be controlled and affected by the unnaturalness of the means of getting information and role-play activity; for example, *oh* prefacing a simple question at the beginning stage of the dialogue discussed in Section 5.5. This limitation has been imposed on studies of Chinese NNSs' speech, as the

publicly available corpora of this group of speakers consist of elicited data. The selection of corpora for analysis is justified in the methodology chapter (Chapter 3).

The third limitation, which relates to methodology, has been widely acknowledged. Since the corpus-based approach is based on the data of production, it is inevitably restricted to the evidence which is present in linguistic forms. In other words, this approach misses out absent features. It is not impossible but would be extremely difficult to investigate any item which was not present in the corpus. (For example, such feature as ellipsis in a corpus can be manually tagged before the corpus is processed by software tools.) In this thesis, the corpus-based approach to the investigation of DMs looks at areas where DMs are used, but the areas where DMs might have been used have been left out. The qualitative text-based analysis helps find some places where DMs might have been used. For instance, in some cases in the Chinese NNSs' speech, there is no signal for a coming correction, but the use of *well* and *you know* signalling repairs would have been of help for the listener. It is likely that there are some areas which lack DMs and which cannot be found with corpus methodologies.

Another weakness in methodology in this thesis is that the process involved in identifying and describing the collocation phenomena surrounding DMs is sometimes unavoidably dependent on the knowledge and intuitions of the researcher. In addition, some may raise doubts whether the *functions* of DMs can be fully supported by the identification of co-occurrence. In the discussion of the contexts where DMs tend to occur, it has been pointed out that, in a few instances, no linguistic evidence can be found and I have had to resort to interpreting based on intuition. In the cases where more than one type of co-occurrence is identified, intuition-based judgements have to be made about which co-occurrence is stronger. The reliance on intuition and the making of subjective judgements may result in a slightly different frequency in duplicated studies in the future. However, since such examples account for a very small proportion, the overall distribution of the co-occurrence of DMs is expected to be reliable.

The fifth problem to be acknowledged here is the limitation of the *LUG* analysis in the cases of two-word phrases. This is discussed in more detail in Section 12.2.6 above. For the time being, there seems to be no other better way to distinguish two-word DMs from their non-discourse uses.

The last weakness of this thesis is the possibility of generalisation. Further investigation shows that the distribution of DMs in each text varies. In other words, high users and low

users of DMs can be identified. Moreover, the frequencies of DMs are affected by genre, type of activity, identity construction and other contextual factors. These all make it difficult to generalise the results.

13.4 Implications for future research

The research and research methodologies reported in this thesis point towards some promising lines of inquiry for further research. First of all, more work can be done to investigate DMs in the NS speech, pointing towards the definition, functions and subtle implications that they carry. The literature survey in Chapter 2 shows that there is no agreement on the definition of DMs, what items are DMs and what their functions are. Although some central DMs, such as *well* and *you know*, have been studied intensively and extensively, the subtle implications of DMs still need to be uncovered. To facilitate interpretation of the use of DMs, the recent availability of multi-modal corpora (e.g. the Nottingham Multi-modal Corpus (Knight, Adolphs, Tennent and Carter 2008)) will be of great help, as they provide audio and visual resources in addition to co-texts.

Future research can be extended to DMs in other types of activity in the Chinese NNS speech. Studies of this kind will help us better understand the uses of DMs among Chinese NNSs. It will also be interesting to apply the analysis modelled here on the speech of other NNSs with different mother tongues, so that non-native varieties of English may be compared and discussed together. A possible starting point could be the investigation of *oh* in another non-native variety of English to find what implications it carries.

The pedagogical implications discussed in the previous chapter open up considerable scope for further work in the classroom setting. Even though I have concluded that it is probably not necessary to formally teach DMs, a certain amount of pedagogical interventions may be of use for some learners. In addition, there has been little empirical research on the efficacy of inductive approaches, such as DDL (Johns 1991), in the classroom.

13.5 Concluding remarks

Corpus-based investigation of spoken English is not recent, but a corpus-linguistic approach to NNS/learner language study is fairly new in Applied Linguistics. There is no doubt that this is a very promising approach to the investigation of NNS/learner language, because it can throw new light on language acquisition, language learning and teaching and varieties of

English as well as other neighbouring branches. A corpus-based approach to the investigation of DMs in Chinese NNS and NS English, such as I have demonstrated in this thesis, is relatively under-explored, compared with investigations into features in written English, and this approach may still be unfamiliar to most researchers, language material developers, practitioners, learners and relevant parties. I hope that this study of the use of DMs, a prominent feature in spoken English, may make a contribution to the investigation of NNS/learner language and that the implications and applications of corpus research will be carefully considered and examined in the future.

APPENDIX 1: Details of the two sub-corpora extracted from the Spoken English Corpus of Chinese Learners (SECCL)

The Spoken English Corpus of Chinese Learners (SECCL) is available via the Foreign Language Teaching and Research Press, Beijing, China (Wen *et al.* 2005). The data in this corpus were collected from the Test for English Majors Band 4 (TEM 4) in China between 1996 and 2002. This test of spoken English was taken by second-year English majors. It consisted of three parts: Task A – retelling a story, Task B – talking on a given topic and Task C – role-playing.

Task A was to retell a roughly 300-word story. The story was played twice. While the story was playing, the test-taker was allowed to take notes. At the end of the second playing, the test-taker had three minutes in which to retell the story.

Task B was to talk on a given topic. Each test-taker had three minutes in which to prepare and was asked to talk for three minutes on a given topic, which was related to Task A.

Task C was role-playing. The two speakers were given three minutes for preparation and were asked to discuss a topic for four minutes on the basis of prompts.

The sub-corpora of the Chinese NNSs' monologues and dialogues consist of transcripts of recordings of Task B and Task C respectively (1,148 texts of each). The topics for discussion, below, are listed in chronological order:

Task B: Talking on a given topic

- 1996 Tell your university friend one unforgettable event in which one of your best friends in your middle school gave you great help when you were in difficulty.
- 1997 Suppose a boy who is living next to your house is a mischief-maker. He often plays a trick on you, which makes you very annoyed. One day you decided to take some action, which turned out to be very successful.
- 1998 Tell us your successful or unsuccessful experiences in your part-time job. If you haven't got any experience yourself, you may tell us what you have heard about others'.
- 1999 Describe one of your experiences in which you had a burning desire to learn something.
- 2000 Describe the unforgettable birthday party you've ever had.
- 2001 Describe a teacher of yours whom you find unusual.
- 2002 Describe an embarrassing situation in which you got very angry.

Task C: Role-playing

- 1996 Student A: You discuss with your partner whether second-year university students should take part-time jobs. The answer to the question itself is not important. What you and your partner should do is to try to think of all the advantages and disadvantages of taking part-time jobs. Remember you should initiate the conversation.
- Student B: You discuss with your partner whether second-year university students should take part-time jobs. The answer to the question itself is not important. What you and your partner should do is to try to think of all the advantages and disadvantages of taking part-time jobs. Remember it is your partner who initiates the conversation.
- 1997 Student A: Your department is going to have an English speech contest. You are eager to enter it but you have a lot of worries. So you come to your friend for advice. First of all, you and your friend are talking about the reasons why you are eager to take part in the contest and what is worrying you. Your friend is trying to help you. However, you don't think all his/her suggestions are helpful. You are trying to give him/her

more explanations about your own situation.

Student B: Your department is going to have an English speech contest. Your friend wants to enter it and at the same time he/she has a lot of worries. Now he/she comes to you for advice. After hearing his/her problems attentively, you give him/her some suggestions by sharing your own experience with him/her. However, your friend doesn't think all your suggestions are helpful and he/she tries to give you more explanations about his/her situation. By having a better understanding of his/her situation, you are able to give him/her better suggestions.

1998 Student A: You and your friend are discussing what to do in future. You and your friend have different plans for future, which, however, are unrealistic. Through the discussion, you and your friend have come to realize what you want to do is different from what you can do in future, and when people make decisions on what job to take, they have to take several factors into consideration. Remember that you should initiate the conversation.

Student B: You and your friend are discussing what to do in future. You and your friend have different plans for future, which, however, are unrealistic. Through the discussion, you and your friend have come to realize what you want to do is different from what you can do in future, and when people make decisions on what job to take, they have to take several factors into consideration. Remember that your partner will initiate the conversation.

1999 Student A: You have been offered two jobs: One is working in the government and the other in a joint venture. You feel difficult to decide which one to take. Now you go to your friend and ask for his/her advice. First of all, you and your friend are discussing advantages and disadvantages of each option. Eventually both you and your friend have found it is not easy at all to select one job out of the two.

Student B: Your friend has been offered two jobs: One is working in the government and the other in a joint venture. She/he feels difficult to make a final decision on which one to take. Now she/he comes to you to seek your advice. First of all, you and your friend are discussing advantages and disadvantages of each option. Eventually, both you and your friend have found it is not easy at all to select one job out of the two.

2000 Student A: You are a freshman who has just enrolled in your university. You are wondering about what the university life would be though you have heard a lot of it. To be honest, you are a little bit lost and afraid of the coming university life. Now you meet a sophomore of your department and ask him/her for advice on how to make good use of time, how to study well and how to make friends in the university. Eventually you become confident about your future again.

Student B: You are a sophomore at the university. A freshman of your department comes to you to ask for advice on how to become successful in the university. You share your experience of the university life with him/her. You try to tell him/her how to make good use of time, how to study well and how to make friends in the university. Eventually you make him/her confident about his/her future again.

2001 Many high school graduates in China are going overseas for their college education. A friend of yours is graduating this year and would like to ask for your advice on whether it is a good idea for a high school graduate to go abroad to study.

Student A: You think this friend should go by all means, and you should try to convince your

partner. Remember you should start the conversation.

Student B: You think this friend should finish college in China before thinking about going abroad, and you should try to convince your partner. Remember your partner will start the conversation.

2002 The geology department of a major university is planning to admit 30 male and 5 female students. However, the results of the college entrance examination show that by average, of all the applicants, females have scored higher than most males. Should the department stick to its original plan?

Student A: You think the department should still stick to the original plan, and you should try to convince your partner. Remember you should start the conversation.

Student B: You DON'T think the department should stick to the original plan, and you should try to convince your partner. Remember your partner will start the conversation.

APPENDIX 2: Details of the two sub-corpora extracted from the Michigan Corpus of Academic Spoken English (MICASE)

The sub-corpus of the NSs' highly monologic discourse mode

(retrieved on 18 February 2009)

The sub-corpus of the American NSs' highly monologic discourse mode consists of the following 13 texts:

	Transcript ID *	File Name	Recording Length	Transcript Word Count
1.	COL999MX036	Provost Public Lecture	61 min.	9,116
2.	COL605MX039	Women's Studies Guest Lecture	65 min.	10,370
3.	COL385MU054	Public Math Colloquium	51 min.	7,664
4.	LEL300SU020	Literature and Social Change Lecture	84 min.	10,207
5.	LEL500JU034	Intro Psychology Lecture	47 min.	7,845
6.	LEL500SU088	Drugs of Abuse Lecture	68 min.	11,115
7.	LEL485JU097	Intro to Physics Lecture	49 min.	7,880
8.	LEL200JU105	Inorganic Chemistry Lecture	50 min.	6,918
9.	LEL175SU106	Biology of Cancer Lecture	70 min.	11,647
10.	LEL320JU143	Renaissance to Modern Art History Lecture	50 min.	8,332
11.	LEL215SU150	Sports and Daily Life in Ancient Rome Lecture	71 min.	12,958
12.	LEL175JU154	Intro to Evolution Lecture	98 min.	12,427
13.	LES405JG078	Graduate Cellular Biotechnology Lecture	83 min.	13,409

The sub-corpus of the NSs' highly interactive discourse mode

(retrieved on 18 February 2009)

The sub-corpus of the American NSs' highly interactive discourse mode consists of the following 48 texts:

	Transcript ID *	File Name	Recording Length	Transcript Word Count
1.	ADV700JU023	Honors Advising	52 min.	9,519
2.	ADV700JU047	Academic Advising	124 min.	28,160
3.	DIS175JU081	Intro Biology Discussion Section	59 min.	7,791
4.	DIS495JU119	Intro to American Politics Discussion Section	55 min.	7,751
5.	INT425JG001	Graduate Student Research Interview 1	34 min.	5,168
6.	INT425JG002	Graduate Student Research Interview 2	20 min.	2,963
7.	INT175SF003	Interview with Botanist	31 min.	5,159
8.	LAB200JU018	Chemistry Lab	47 min.	8,169
9.	LAB175SU026	Biology of Birds Field Lab	92 min.	11,769
10.	LAB175SU032	Biology of Fishes Field Lab	89 min.	11,370
11.	LAB175SU033	Biology of Fishes Lab	95 min.	8,153
12.	LAB500SU044	Biopsychology Lab	52 min.	9,455
13.	LES385SU007	Number Theory Math Lecture	36 min.	4,144
14.	LES175SU031	Biology of Fishes Group Activity	19 min.	2,866
15.	LES215MU056	Intro Latin Lecture	50 min.	5,883
16.	LES320SU085	Visual Sources Lecture	69 min.	12,526
17.	LES565SU137	Sex, Gender, and the Body Lecture	73 min.	14,629
18.	LES220SU140	Ethics Issues in Journalism Lecture	83 min.	16,291

19.	MTG425JG004	Natural Resources Research Group Meeting	83 min.	9,382
20.	MTG400MX008	Immunology Lab Meeting	60 min.	9,523
21.	MTG999ST015	Forum for International Educators Meeting	102 min.	17,323
22.	MTG485SG142	Physics Research Group Meeting	41 min.	9,076
23.	OFC301MU021	English Composition Tutorial	45 min.	3,586
24.	OFC578SG037	Technical Communications Tutorial	25 min.	4,178
25.	OFC150MU042	Astronomy Peer Tutorial	102 min.	21,798
26.	OFC575MU046	Statistics Office Hours	52 min.	11,265
27.	OFC270MG048	Computer Science Office Hours	116 min.	19,977
28.	OFC115SU060	Anthropology of American Cities Office Hours	178 min.	31,268
29.	OFC105SU068	American Culture Advising	42 min.	8,511
30.	OFC355SU094	Linguistics Independent Study Advising	52 min.	6,943
31.	OFC280SU109	Economics Office Hours	92 min.	14,050
32.	OFC195SU116	Heat and Mass Transfer Office Hours	137 min.	20,603
33.	OFC175JU145	Intro Biology Exam Review	55 min.	9,014
34.	OFC320SU153	Art History Office Hours	66 min.	9,233
35.	SEM475JU084	First Year Philosophy Seminar	72 min.	13,906
36.	SEM300MU100	English Composition Seminar	125 min.	21,442
37.	SGR385SU057	Math Study Group	132 min.	17,753
38.	SGR999MX115	Objectivism Student Group	125 min.	22,416
39.	SGR175SU123	Biochemistry Study Group	109 min.	17,530
40.	SGR200JU125	Organic Chemistry Study Group	101 min.	18,124
41.	SGR175MU126	Intro Biology Study Group	103 min.	24,514
42.	SGR195SU127	Chemical Engineering Group Project Meeting	77 min.	11,289
43.	SGR565SU144	American Family Group Project Meeting	85 min.	14,116
44.	SGR999SU146	Senior Thesis Study Group	64 min.	15,483
45.	STP200JU019	Chemistry Discussion Section Student Presentations	51 min.	7,303
46.	STP125JG050	Architecture Critiques	123 min.	24,228
47.	SVC999MX104	Media Union Service Encounters	187 min.	19,072
48.	SVC999MX148	Science Learning Center Service Encounters	121 min.	8,613

* The first three letters in the transcription ID stand for the type of speech event. The two letters in the middle indicate the level of participants (MICASE Manual 2007: 5-7).

ADV	advising sessions	interactions between students and academic advisors
COL	colloquia	departmental or university-wide lectures, panel discussions, workshops, brown bag lunch talks, etc.
DIS	discussion sections	additional section of a lecture class designed for maximum student participation; may also be called recitation
INT	interviews	interviews for research purposes
LAB	lab sections	lab sections of science and engineering classes; may include problem solving sessions
LEL	large lectures	lecture class; class size= more than 40 students
LES	small lectures	lecture class; class size= 40 or fewer students
MTG	meetings	faculty, staff, student government, research group meetings, not including study group meetings
OFC	office hours	held by faculty or graduate student instructors in connection with a specific class or project
SEM	seminars	any class defined as a seminar (primarily graduate level)
SGR	study groups	informal student-led study groups, one time or on-going
STP	student presentations	class other than a seminar in which one or more students speak in

		front of the class or lead discussion
SVC	service encounters	library, computer center and financial aid office services
MX	mixed	mixed faculty, staff and students
MU	mixed undergrad	mixed undergraduates
SU	senior undergrad	third year and above undergraduates
JU	junior undergrad	first and second year undergraduates
JG	junior graduate	first and second year or master's level graduate students
SF	senior faculty	associate professors and above
ST	staff	non-teaching university employees
SG	senior graduate	third year and above Ph.D. students

APPENDIX 3: Details of the two sub-corpora extracted from the International Corpus of English – Great Britain (ICE-GB)

The sub-corpus of the NSs' unscripted monologues

The sub-corpus of the NSs' unscripted monologues consists of 70 texts of four types of monologue: 1) spontaneous commentaries (File: S2A-001 – S2A-020), 2) unscripted speeches (File: S2A-021 – S2A-050), 3) demonstrations (File: S2A-051 – S2A-060) and 4) legal presentations (File: S2A-061 – S2A-070).

	Transcript ID	Title	Channel/Place	Date	Transcript Word Count
Spontaneous commentaries					
1.	S2A-001	Soccer	Radio, BBC 5	1991	2,135
2.	S2A-002	Sport on Five	Radio, BBC 5	1991	2,132
3.	S2A-003	Football Extra	Radio, BBC 5	1991	2,260
4.	S2A-004	Rugby League	TV, BBC 1	1990	2,092
5.	S2A-005	The Grand National	Radio, BBC 5	1991	2,119
6.	S2A-006	The Epsom Derby	Radio, BBC 5	1991	2,204
7.	S2A-007	Athletics	TV, ITV	1991	1,956
8.	S2A-008	Snooker	TV, BBC 1	1991	681
		Andrew Neil on Sunday	Radio, LBC	1991	925
		Athletics	TV, ITV	1991	253
9.	S2A-009	Champion Sport	Radio, BBC 5	1991	2,072
10.	S2A-010	International Soccer Extra	Radio, BBC 5	1991	1,988
11.	S2A-011	Trooping the Colour	Radio, BBC 4	1991	2,326
12.	S2A-012	Sunday Sport	Radio, BBC 5	1991	2,135
13.	S2A-013	Sunday Sport	Radio, BBC 5	1991	2,003
14.	S2A-014	International Soccer Extra	Radio, BBC 5	1991	1,925
15.	S2A-015	LBC Sport	Radio, LBC	1991	2,340
16.	S2A-016	Tour de France	TV, Channel 4	1991	1,893
17.	S2A-017	Capital FM Soccer	Radio, Capital Radio	1991	2,141
18.	S2A-018	LBC Sport	Radio, LBC	1991	2,229
19.	S2A-019	The Gulf Ceremony	Radio, BBC 4	1991	2,096
20.	S2A-020	The Maundy Thursday Service at Westminster Abbey;	Radio, BBC 4	1991	1,422
		Gulf Service: National Service of Remembrance and Thanksgiving			1,139
Unscripted speeches					
21.	S2A-021	Teaching the Teachers	Royal Society of Arts, The Strand, London	1991	2,246
22.	S2A-022	The Ancient Celts Through Caesar's Eyes	British Museum	1991	2,139
23.	S2A-023	Getting Britain Moving	Royal Society of Arts, The Strand, London	1991	2,197
24.	S2A-024	Greek Temples	British Museum	1991	2,247
25.	S2A-025	Earthquakes and Buildings	Darwin Theatre, UCL	1991	2,154
26.	S2A-026	Joseph Hekekyan	Darwin Theatre, UCL	1991	2,244
27.	S2A-027	An Academic's Path through the Media	Darwin Theatre, UCL	1991	2,554
28.	S2A-028	n/a	UCL	1991	2,167
29.	S2A-029	n/a	UCL	1991	2,160
30.	S2A-030	Prosodic Phonology	London	1991	2,024

31.	S2A-031	Towards a Quality Workforce	Royal Society of Arts, The Strand, London	1991	2,268
32.	S2A-032	Eurotra and some other machine translation research systems	King's College, London	1991	2,249
33.	S2A-033	n/a	UCL	1991	2,050
34.	S2A-034	n/a	UCL	1991	2,196
35.	S2A-035	n/a	UCL	1991	2,322
36.	S2A-036	n/a	UCL Department of Jewish & Hebrew Studies	1991	2,085
37.	S2A-037	The relationship between industrial innovation and academic research	Darwin Theatre, UCL	1991	2,251
38.	S2A-038	The Third Age	Royal Society of Arts, The Strand, London	1991	2,186
39.	S2A-039	Citizen Who, Citizen How?	Royal Society of Arts, The Strand, London	1991	2,094
40.	S2A-040	The Ark	UCL Chemistry Theatre	1991	2,150
41.	S2A-041	n/a	Darwin Theatre, UCL	n/a	2,046
42.	S2A-042	The Immunological Compact Disc	Darwin Theatre, UCL	1991	2,371
43.	S2A-043	Studying Climate Change from Outer Space	Darwin Theatre, UCL	1991	2,498
44.	S2A-044	Lawyers' Stories	Darwin Theatre, UCL	1992	2,067
45.	S2A-045	Managing a Great Estate	Royal Society of Arts, The Strand, London	1991	2,285
46.	S2A-046	n/a	UCL	1991	2,170
47.	S2A-047	n/a	UCL	1991	2,142
48.	S2A-048	Write with you hand, read with your mouth	Darwin Theatre, UCL	1991	1,938
49.	S2A-049	Creating a learning organisation	Royal Society of Arts, The Strand, London	1992	2,167
50.	S2A-050	n/a	London (private home)	1991	2,398
Demonstrations					
51.	S2A-051	Movement in the Microscopical World	Darwin Theatre, UCL	1991	2,190
52.	S2A-052	New Kingdom Paintings and Reliefs	British Museum	1991	1,987
53.	S2A-053	Looking into the Brain with Light	Darwin Theatre, UCL	1991	2,005
54.	S2A-054	Pass Your Motorbike Test	n/a	1991	2,417
55.	S2A-055	Top Gear	TV, BBC 2	1991	2,302
56.	S2A-056	n/a	UCL, London	1991	2,173
57.	S2A-057	n/a	UCL, London	1991	2,446
58.	S2A-058	n/a	UCL, London	1991	1,089
				1992	408
				1992	579
59.	S2A-059	Persian Manuscripts	British library	1990	2,226
60.	S2A-060	The Mosaics of Torcello	British Museum	1990	2,741
Legal presentations					
61.	S2A-061	n/a	Queen's Bench	1990	2,214
62.	S2A-062	n/a	Queen's Bench	1990	2,136
63.	S2A-063	n/a	Queen's Bench	1990	2,097
64.	S2A-064	n/a	Queen's Bench	1990	2,114
65.	S2A-065	n/a	Queen's Bench	1991	2,133
66.	S2A-066	n/a	Queen's Bench	1991	2,124
67.	S2A-067	n/a	Queen's Bench	1991	2,327

68.	S2A-068	n/a	Queen's Bench	1990	2,157
69.	S2A-069	n/a	County Court	1990	1,870
70.	S2A-070	n/a	Queen's Bench	1990	2,563

The sub-corpus of the NSs' private direct conversations

This sub-corpus consists of 90 private direct conversations by British NSs.

	Transcript ID	Title	Channel/Place	Date	Transcript Word Count
1.	S1A-001	n/a	Middlesex Polytechnic	1991	2,050
2.	S1A-002	n/a	Middlesex Polytechnic	1991	2,055
3.	S1A-003	n/a	Middlesex Polytechnic	1991	2,146
4.	S1A-004	n/a	Middlesex Polytechnic	1991	2,090
5.	S1A-005	n/a	London	1991	2,156
6.	S1A-006	n/a	London	1991	2,105
7.	S1A-007	n/a	London (private home)	1991	2,044
8.	S1A-008	n/a	London (private home)	1991	2,249
9.	S1A-009	n/a	London (private home)	1991	2,006
10.	S1A-010	n/a	London (private home)	1991	2,077
11.	S1A-011	n/a	London/ Cambridge	1991	852 1,124
12.	S1A-012	n/a	London (private home)	1991	1,893
13.	S1A-013	n/a	Times Books Ltd (office)	1991	2,217
14.	S1A-014	n/a	Survey of English Usage	1991	2,093
15.	S1A-015	n/a	London (public house)	1991	2,013
16.	S1A-016	n/a	Times Books Ltd (office)	1991	2,072
17.	S1A-017	n/a	London (private home)	1991	1,832
18.	S1A-018	n/a	London (private home)	1991	2,049
19.	S1A-019	n/a	London (private home)	1991	2,280
20.	S1A-020	n/a	London (private home)	1991	2,026
21.	S1A-021	n/a	Tunbridge Wells	1991	1,896
22.	S1A-022	n/a	Cambridge (private home)	1991	1,901
23.	S1A-023	n/a	London (private home)	1991	2,184
24.	S1A-024	n/a	Survey of English Usage	1991	2,183
25.	S1A-025	n/a	London (private home)	1991	2,179
26.	S1A-026	n/a	London (private home)	1991	1,998
27.	S1A-027	n/a	London (private home)	1991	1,856
28.	S1A-028	n/a	Cambridge (private home)	1991	2,183
29.	S1A-029	n/a	Survey of English Usage	1991	2,189
30.	S1A-030	n/a	London (private flat)	1991	1,950
31.	S1A-031	n/a	London (private home)	1991	2,045
32.	S1A-032	n/a	Cambridge (private home)	1991	1,703
33.	S1A-033	n/a	UCL Careers Office	1992	2,023
34.	S1A-034	n/a	UCL Careers Office	1992	2,066
35.	S1A-035	n/a	UCL Careers Office	1992	2,027
36.	S1A-036	n/a	London (private home)	1991	2,089
37.	S1A-037	n/a	London (private home)	1991	2,228
38.	S1A-038	n/a	Cambridge (private home)	1991	1,917
39.	S1A-039	n/a	Cambridge (private flat)	1991	2,067
40.	S1A-040	n/a	Cambridge (private flat)	1991	2,446
41.	S1A-041	n/a	Cambridge (private flat)	1991	2,223
42.	S1A-042	n/a	Cambridge (private flat)	1991	2,024
43.	S1A-043	n/a	London (private home)	1990	1,837

44.	S1A-044	n/a	London (private home)	1990	1,825
45.	S1A-045	n/a	London (private home)	1990	2,111
46.	S1A-046	n/a	Cambridge (private home)	1991	1,921
47.	S1A-047	n/a	Cambridge (private home)	1991	2,107
48.	S1A-048	n/a	London (private home)	1992	2,012
49.	S1A-049	n/a	London (private home)	1991	2,025
50.	S1A-050	n/a	UCL Counselling Service	1991	2,240
51.	S1A-051	n/a	UCL Health Centre	1991	1,995
52.	S1A-052	n/a	London (private home)	1991	2,036
53.	S1A-053	n/a	London (private home)	1992	2,211
54.	S1A-054	n/a	London (private home)	1991	2,214
55.	S1A-055	n/a	Cambridge (CUP Staff Canteen)	1992	1,898
56.	S1A-056	n/a	London (private home)	1992	1,964
57.	S1A-057	n/a	London (private home)	1992	2,157
58.	S1A-058	n/a	London (private home)	1992	2,090
59.	S1A-059	n/a	UCL Student Counselling Office	1991	1,996
60.	S1A-060	n/a	UCL Student Counselling Office	1990	2,181
61.	S1A-061	n/a	London (restaurant)	1992	2,141
62.	S1A-062	n/a	UCL Student Counselling Office	1991	2,033
63.	S1A-063	n/a	London (private home)	1991	1,942
64.	S1A-064	n/a	London, Central School of Speech & Drama	1991	2,039
65.	S1A-065	n/a	Duns, Berwickshire	1991	1,937
66.	S1A-066	n/a	UCL Careers Office	1992	2,078
67.	S1A-067	n/a	London (private home)	1991	2,094
68.	S1A-068	n/a	UCL Students' Union office	1992	1,956
69.	S1A-069	n/a	UCL Students' Union office	1992	2,121
70.	S1A-070	n/a	UCL Students' Union office	1992	1,844
71.	S1A-071	n/a	London (private home)	1992	2,101
72.	S1A-072	n/a	Middlesex Polytechnic	1991	1,994
73.	S1A-073	n/a	London (private home)	1991	1,889
74.	S1A-074	n/a	UCLU Travel Office/ University of London Union, General Office	1991 1992	1,475 335
75.	S1A-075	n/a	Middlesex Polytechnic	1991	2,187
76.	S1A-076	n/a	Middlesex Polytechnic	1991	2,173
77.	S1A-077	n/a	UCL Careers Office	1992	2,098
78.	S1A-078	n/a	UCLU Rights & Advice Office	1992	1,940
79.	S1A-079	n/a	UCLU Rights & Advice Office	1992	2,131
80.	S1A-080	n/a	n/a	1992	2,198
81.	S1A-081	n/a	London (private home)	1992	2,106
82.	S1A-082	n/a	London, Central School of Speech & Drama	1992	2,036
83.	S1A-083	n/a	London (private home)	1991	1,991
84.	S1A-084	n/a	UCL English Department Student Common Room	1992	2,196
85.	S1A-085	n/a	London (private home)	1992	2,090
86.	S1A-086	n/a	London (private home)	1991	1,978
87.	S1A-087	n/a	London (Dental Surgery)	1992	2,288
88.	S1A-088	n/a	London (Dental Surgery)	1992	2,097
89.	S1A-089	n/a	London (Dental Surgery)/ UCL Health Centre	1992 1991	978 998
90.	S1A-090	n/a	UCL English Department Student Common Room	1992	1,968

APPENDIX 4: Labels in *Linear Unit Grammar* (Sinclair and Mauranen 2006)

Label	Element
O	organisational element
OI	interactive organisational element
OT	text-oriented organisational element
M	message-oriented element
MF	message fragment
M-	incomplete message unit
+M	completion of message unit
+M-	partial completion of message unit
MS	supplement to message unit
MA	adjustment to message unit
MR	revision to message unit

APPENDIX 5: Frequency comparison of discourse markers in the manual classification and in the tagged ICE-GB

Discourse marker	Method	ICE-GB: Unscripted monologues	ICE-GB: Private direct conversations
<i>like</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	6 (Type B) out of 235	143* (Type B) out of 913
	<i>Like</i> tagged with discourse marker, searched by ICECUP	4 (Type B) out of 234	140 (Type B) out of 882
<i>well</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	158 (Type B) out of 353	1,303* (Type B) out of 1,521
	<i>Well</i> tagged with discourse marker, searched by ICECUP	153 (Type B) out of 342	1,287 (Type B) out of 1,493
<i>you know</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	66 (Type B) out of 79	666* (Type B) out of 819
	<i>You know</i> tagged with discourse marker, searched by ICECUP	60 (Type B) out of 79	704 (Type B) out of 823
<i>I mean</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	14 (Type B) out of 21	811* (Type B) out of 865
	<i>I mean</i> tagged with discourse marker, searched by ICECUP	13 (Type B) out of 21	811 (Type B) out of 862
<i>you see</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	13 (Type B) out of 35	97 (Type B) out of 144
	<i>You see</i> tagged with discourse marker, searched by ICECUP	7 (Type B) out of 34	93 (Type B) out of 144
<i>now</i>	Manual classification of Types A and B, processed in <i>WordSmith 4</i>	192* (Type B) out of 620	108 (Type B) out of 372
	<i>Now</i> tagged with discourse marker, searched by ICECUP	206 (Type B) out of 617	105 (Type B) out of 367

* The frequency of Type B is extrapolated from its proportion in the three sets of 100-line concordance samples. For example, the frequency count of Type B *like* is extrapolated from its proportion in the three sets of 100-line concordance samples. The number of Type B *like* (143) = total occurrences (913) multiplied by 15.7% (proportion of Type B *like*) (see Table 4.1).

APPENDIX 6: The log-likelihood test of discourse markers between two types of genre and between Chinese non-native speakers and native speakers

Log-likelihood (LL) tests of discourse markers between the sub-corpora of the monologic genres and dialogic genres

Corpus ¹ \ Discourse marker	Chinese NNSs				American NSs				British NSs			
	A1	A2	+/- ²	LL score ³	B1	B2	+/- ²	LL score ³	C1	C2	+/- ²	LL score ³
<i>like</i>	13	30	-	0.65	15	4878	-	1882.00	6	143	-	132.10
<i>oh</i>	397	2882	-	966.54	24	2795	-	969.87	45	1123	-	1047.74
<i>well</i>	82	516	-	150.62	104	1904	-	323.49	158	1303	-	824.19
<i>you know</i>	228	2676	-	1259.88	119	1896	-	284.35	66	666	-	466.18
<i>I mean</i>	51	243	-	50.06	57	1493	-	324.92	14	811	-	860.89
<i>you see</i>	29	403	-	206.81	17	39	+	4.29	13	97	-	57.92
<i>I think</i>	102	663	-	200.12	20	215	-	19.70	43	126	-	28.63
<i>now</i>	103	206	-	1.00	203	568	+	25.97	192	108	+	42.02
Total tokens	336303	596639			134096	577996			153646	185000		

Corpus ¹ \ Discourse marker	Monologic genres								Dialogic genres							
	A1	B1	+/- ²	LL score ³	A1	C1	+/- ²	LL score ³	A2	B2	+/- ²	LL score ³	A2	C2	+/- ²	LL score ³
<i>like</i>	13	15	-	7.70	13	6	-	0.00	30	4878	-	6593.41	30	143	-	268.75
<i>oh</i>	397	24	+	142.58	397	45	+	112.27	2882	2795	-	0.00	2882	1123	-	40.74
<i>well</i>	82	104	-	60.84	82	158	-	119.94	516	1904	-	891.44	516	1303	-	1864.41
<i>you know</i>	228	119	-	5.50	228	66	+	11.53	2676	1896	+	110.11	2676	666	+	26.91
<i>I mean</i>	51	57	-	27.91	51	14	+	3.12	243	1493	-	1040.84	243	811	-	1330.43
<i>you see</i>	29	17	-	1.53	29	13	+	0.00	403	39	+	337.48	403	97	+	5.28
<i>I think</i>	102	20	+	9.80	102	43	+	0.20	663	215	+	225.72	663	126	+	28.27
<i>now</i>	103	203	-	187.75	103	192	-	141.14	206	568	-	187.78	206	108	-	18.34
Total tokens	336303	134096			336303	153646			596639	577996			596639	185000		

1: Corpus A1- SECCL, monologues
 Corpus A2- SECCL, dialogues
 Corpus B1- MICASE, highly monologic discourse mode
 Corpus B2- MICASE, highly interactive discourse mode
 Corpus C1- ICE-GB, unscripted monologues
 Corpus C2- ICE-GB, private direct conversations

2: + indicates over-representation in the 1st corpus relative to the 2nd corpus
 - indicates under-representation in the 1st corpus relative to the 2nd corpus

3: Scores of 3.84 or above are regarded as significant
 Critical value for corpus studies is 15.13 (Rayson et al. 2004)
 Critical value = 3.84; p < 0.05
 Critical value = 6.63; p < 0.01
 Critical value = 10.83; p < 0.001
 Critical value = 15.13; p < 0.0001

APPENDIX 7: The z test for two proportions of discourse markers between two types of genre and between Chinese non-native speakers and native speakers

Z tests for two proportions of discourse markers between the sub-corpora of the monologic genres and dialogic genres

Corpus ¹ \ Discourse marker	Chinese NNSs				American NSs				British NSs			
	A1	A2	z score ²	p-value	B1	B2	z score ²	p-value	C1	C2	z score ²	p-value
<i>like</i>	13	30	-0.82	0.410	15	4878	-67.3	0.001	6	143	-11.03	0.001
<i>oh</i>	397	2882	-33.94	0.001	24	2795	-47.38	0.001	45	1123	-31.1	0.001
<i>well</i>	82	516	-32.32	0.001	104	1904	-23.53	0.001	158	1303	-28.52	0.001
<i>you know</i>	228	2676	-39.06	0.001	119	1896	-21.6	0.001	66	666	-21.29	0.001
<i>I mean</i>	51	243	-7.59	0.001	57	1493	-24.71	0.001	14	811	-27.6	0.001
<i>you see</i>	29	403	-15.82	0.001	17	39	1.82	0.069	13	97	-7.56	0.001
<i>I think</i>	102	663	-15.37	0.001	20	215	-5.32	0.001	43	126	-5.41	0.001
<i>now</i>	103	206	-1.01	0.312	203	568	4.66	0.001	192	108	6.27	0.001
Total tokens	336303	596639			134096	577996			153646	185000		

Z tests for two proportions of discourse markers between the sub-corpora of the Chinese non-native speakers and native speakers

Corpus ¹ \ Discourse marker	Monologic genres								Dialogic genres							
	A1	B1	z score ²	p-value	A1	C1	z score ²	p-value	A2	B2	z score ²	p-value	A2	C2	z score ²	p-value
<i>like</i>	13	15	-2.38	0.017	13	6	-0.02	0.984	30	4878	-69.52	0.001	30	143	-11.07	0.001
<i>oh</i>	397	24	14.39	0.001	397	45	12.07	0.001	2882	2795	-0.04	0.967	2882	1123	-6.16	0.001
<i>well</i>	82	104	-6.59	0.001	82	158	-9.11	0.001	516	1904	-28.77	0.001	516	1303	-31.19	0.001
<i>you know</i>	228	119	-2.26	0.024	228	66	3.58	0.001	2676	1896	10.51	0.001	2676	666	5.4	0.001
<i>I mean</i>	51	57	-4.54	0.001	51	14	1.87	0.061	243	1493	-30.35	0.001	243	811	-25.52	0.001
<i>you see</i>	29	17	-1.17	0.242	29	13	0.06	0.954	403	39	17.21	0.001	403	97	2.4	0.016
<i>I think</i>	102	20	3.44	0.001	102	43	0.45	0.653	663	215	14.37	0.001	663	126	5.78	0.001
<i>now</i>	103	203	-10.94	0.001	103	192	-9.93	0.001	206	568	-13.36	0.001	206	108	-3.9	0.001
Total tokens	336303	134096			336303	153646			596639	577996			596639	185000		

1: Corpus A1- SECCL, monologues
 Corpus A2- SECCL, dialogues
 Corpus B1- MICASE, highly monologic discourse mode
 Corpus B2- MICASE, highly interactive discourse mode
 Corpus C1- ICE-GB, unscripted monologues
 Corpus C2- ICE-GB, private direct conversations

2: + z scores indicate over-representation in the 1st corpus relative to the 2nd corpus
 - z scores indicate under-representation in the 1st corpus relative to the 2nd corpus

REFERENCES

- Ackerley, K. 2008. 'Using comparable expert-writer and learner corpora for developing report-writing skills' in Torsello, C. T., K. Ackerley and E. Castello (eds.). *Corpora for University Language Teachers*. Bern: Peter Lang.
- Aertselaer, J. N. v. 2008. 'Contrasting English-Spanish interpersonal discourse phrases: A corpus study' in Meunier, F. and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Aijmer, K. 1987. 'Oh and ah in English conversation' in Meijs, W. (ed.) *Corpus Linguistics and Beyond*. Amsterdam: Rodopi.
- Aijmer, K. 1997. 'I think - an English modal particle' in Swan, T. and O. J. Westvik (eds.). *Modality in Germanic Languages. Historical and Comparative Perspectives*. Berlin: Mouton de Gruyter.
- Aijmer, K. 2002. *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins.
- Altenberg, B. 1993. 'Recurrent verb - complement constructions in the London-Lund corpus' in Aarts, J., P. d. Haan and N. Oostdijk (eds.). *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi.
- Altenberg, B. 2002. 'Using bilingual corpus evidence in learner corpus research' in Granger, S., J. Hungand and S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Altenberg, B. and S. Granger. 2001. 'The grammatical and lexical patterning of *make* in native and non-native student writing'. *Applied Linguistics* 22: 173-194.
- Andersen, G. 1998. 'The pragmatic marker *like* from a relevance-theoretic perspective' in Jucker, A. H. and Y. Ziv (eds.). *Discourse Markers: Description and Theory*. Amsterdam: John Benjamins.
- Andersen, G. 2001. *Pragmatic Markers and Sociolinguistic Variation: A Relevance-theoretic Approach to the Language of Adolescents [electronic resource]*. Amsterdam: John Benjamins.
- Anthony, L. 2007. *AntConc 3.2.1*.
- Aston, G. 1997. 'Enriching the learning environment: Corpora in ELT' in Wichmann, A., S. Fligelstone, T. McEnery and G. Knowles (eds.). *Teaching and Language Corpora*. Essex: Longman.
- Barbieri, F. 2009. 'Quotative 'be *Like*' in American English: Ephemeral or here to stay?'. *English World-Wide* 30/1: 68-90.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, D. and E. Finegan. 1991. 'On the exploitation of computerized corpora in variation studies' in Aijmer, K. and B. Altenberg (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- Biber, D., E. Finegan, S. Johansson, S. Conrad and G. Leech. 1999. *Longman Grammar of Spoken and Written English*. Essex: Pearson Education Limited.
- Blakemore, D. 1987. *Semantic Constraints on Relevance*. Oxford: Blackwell.
- Blakemore, D. 1992. *Understanding Utterances: An Introduction to Pragmatics*. Oxford: Blackwell.
- Blakemore, D. 2002. *Relevance and Linguistic Meaning: The Semantics and Pragmatics of*

- Discourse Markers*. Cambridge: Cambridge University Press.
- Bolinger, D. 1989. *Intonation and Its Uses: Melody and Grammar in Discourse*. Stanford: Stanford University Press.
- Brazil, D. 1995. *A Grammar of Speech*. Oxford: Oxford University Press.
- Brown, G. and G. Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Brown, P. and S. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Cameron, D. 2001. *Working with Spoken Discourse*. London: Sage Publications.
- Carlson, L. 1984. *Well in Dialogue Games: A Discourse Analysis of the Interjection well in Idealized Conversation*. Amsterdam: John Benjamins.
- Carter, R. and M. McCarthy. 1995. 'Grammar and the spoken language'. *Applied Linguistics* 16/2: 141-158.
- Carter, R. and M. McCarthy. 1997. *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Carter, R. and M. McCarthy. 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Cheng, W. and M. Warren. 1999. 'Facilitating a description of intercultural conversations: The Hong Kong Corpus of Conversational English'. *ICAME Journal* 23: 5-20.
- Clark, H. H. and A. Bangerter. 2004. 'Changing ideas about reference' in Noveck, I. and D. Sperber (eds.). *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan.
- Coates, J. 2003. 'The role of epistemic modality in women's talk' in Facchinetti, R., M. Krug and F. Palmer (eds.). *Modality in Contemporary English*. New York: Mouton de Gruyter.
- Collins COBUILD Advanced Learner's English Dictionary* (5th edn.). 2006. Glasgow: HarperCollins Publishers.
- Collins COBUILD English Language Dictionary*. 1987. London: Collins.
- Coxhead, A. 1998. *An Academic Word List*. Wellington, New Zealand: Victoria University of Wellington.
- Coxhead, A. 2000. 'A new academic word list'. *TESOL Quarterly* 34/2: 213-238.
- Crystal, D. 1988. 'Another look at, well, you know'. *English Today* 13: 47-49.
- Crystal, D. and D. Davy. 1975. *Advanced Conversational English*. London: Longman.
- Cuenca, M.-J. 2008. 'Pragmatic markers in contrast: The case of well'. *Journal of Pragmatics* 40: 1371-1391.
- Cukor-Avila, P. 2002. '*She say, she go, she be like*: Verbs of quotation over time in African American Vernacular English'. *American Speech* 77/1: 3-31.
- Cutting, J. 2006. 'Spoken grammar: Vague language and EAP' in Hughes, R. (ed.) *Spoken English, TESOL and Applied Linguistics*. London: Palgrave Macmillan.
- D'Arcy, A. 2005. *Like: Syntax and Development*. Unpublished PhD thesis. Department of Linguistics. University of Toronto. Toronto, Canada.
- Dagneaux, E., S. Denness and S. Granger. 1998. 'Computer-aided error analysis'. *System* 26/2: 163-174.
- Dailey-O'Cain, J. 2000. 'The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*'. *Journal of Sociolinguistics* 4/1: 60-80.
- Dalziel, F. and F. Helm. 2008. 'Exploring modality in a learner corpus on online writing' in Torsello, C. T., K. Ackerley and E. Castello (eds.). *Corpora for University Language Teachers*. Bern: Peter Lang.
- De Cock, S. 2007. 'Routinized building blocks in native speaker and learner speech: Clausal sequences in the spotlight' in Campoy, M. C. and M. J. Luzón (eds.). *Spoken Corpora*

- in *Applied Linguistics*. Bern: Peter Lang.
- Dunning, T. 1993. 'Accurate methods for the statistics of surprise and coincidence'. *Computational Linguistics* 19/1: 61-74.
- Erman, B. 1987. *Pragmatic Expressions in English: A Study of You Know, You See and I Mean in Face-to-face Conversation*. Stockholm: English Department, Stockholm University.
- Erman, B. 1992. 'Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction'. *Language Variation and Change* 4: 217-234.
- Erman, B. 2001. 'Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk'. *Journal of Pragmatics* 33: 1337-1359.
- Flowerdew, J. 2006. 'Use of signalling nouns in a learner corpus'. *International Journal of Corpus Linguistics* 11/2: 345-362.
- Fortanet, I. 2004. '*I think*: Opinion, uncertainty or politeness in academic spoken English?'. *RAEL: Revista Electronica de Linguica Aplicada* 3: 63-84.
- Fox Tree, J. E. 2006. 'Placing *like* in telling stories'. *Discourse Studies* 8/6: 723-743.
- Fox Tree, J. E. 2007. 'Folk notions of *um* and *uh, you know, and like*'. *Text and Talk* 27/3: 297-314.
- Fox Tree, J. E. and J. C. Schrock. 1999. 'Discourse markers in spontaneous speech: *Oh* what a difference an *oh* makes'. *Journal of Memory and Language* 40/2: 280-295.
- Fox Tree, J. E. and J. C. Schrock. 2002. 'Basic meanings of *you know* and *I mean*'. *Journal of Pragmatics* 34/6: 727-747.
- Fox Tree, J. E. and J. M. Tomlinson Jr. 2008. 'The rise of *like* in spontaneous quotations'. *Discourse Processes* 45: 85-102.
- Fraser, B. 1990. 'An approach to discourse markers'. *Journal of Pragmatics* 14: 383-395.
- Fraser, B. 1996. 'Pragmatic markers'. *Pragmatics* 6/2: 167-190.
- Fraser, B. 1999. 'What are discourse markers?'. *Journal of Linguistics* 31/7: 931-952.
- Fraser, B. 2009. 'Topic orientation markers'. *Journal of Pragmatics* 41: 892-898.
- Fromkin, V. and R. Rodman. 1974. *An Introduction to Language*. New York: Holt, Rinehart and Winston, Inc.
- Fromkin, V., R. Rodman and N. Hyams. 2007. *An Introduction to Language*, (8th edn.). Boston: Thomson.
- Fuller, J. M. 2003a. 'The influence of speaker roles on discourse marker use'. *Journal of Pragmatics* 35/1: 23-45.
- Fuller, J. M. 2003b. 'Use of the discourse marker *like* in interviews'. *Journal of Sociolinguistics* 7/3: 365-377.
- Fung, L. and R. Carter. 2007. 'Discourse markers and spoken English: Native and learner use in pedagogic settings'. *Applied Linguistics* 28/3: 410-439.
- Gilquin, G., S. Papp and M. B. Díez-Bedmar. (eds.) 2009. *Linking up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi.
- Goffman, E. 1981. *Forms of Talk*. Oxford: Basil Blackwell.
- Graddol, D. 1998. *The Future of English?* London: The British Council.
- Graddol, D. 2006. *English Next*. London: British Council.
- Granger, S. 1998a. 'The computer learner corpus: A versatile new source of data for SLA research' in Granger, S. (ed.) *Learner English on Computer*. Harlow: Longman.
- Granger, S. 1998b. 'Introduction' in Granger, S. (ed.) *Learner English on Computer*. London: Longman.
- Granger, S. (ed.) 1998c. *Learner English on Computer*. London: Longman.
- Granger, S. 2002. 'A bird's-eye view of learner corpus research' in Granger, S., J. Hung and S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and*

- Foreign Language Teaching*. Amsterdam: John Benjamins.
- Granger, S., J. Hung and S. Petch-Tyson. (eds.) 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Granger, S. and F. Meunier. 2008. 'Phraseology in language learning and teaching: Where to from here?' in Meunier, F. and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Granger, S. and M. Paquot. 2009. 'Lexical verbs in academic discourse: A corpus-driven study of learner use' in Charles, M., D. Pecorari and S. Hunston (eds.). *Academic Writing*. London: Continuum.
- Greenbaum, S. (ed.) 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Oxford University Press.
- Groom, N. W. 2007. *Phraseology and Epistemology in Humanities Writing: A Corpus-driven Study*. Unpublished PhD thesis. Department of English. University of Birmingham. Birmingham, UK.
- Guo, X. 2006. *Verbs in the Written English of Chinese Learners: A Corpus-based Comparison Between Non-native Speakers and Native Speakers*. Unpublished PhD thesis. Department of English. University of Birmingham. Birmingham, UK.
- Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, M. A. K. and R. Hasan. 1985. *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Oxford: Oxford University Press.
- Hasselgren, A. 2002. 'Learner corpora and language testing: Smallwords as markers of learner fluency' in Granger, S., J. Hung and S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- He, A. W. and B. Lindsey. 1998. "'You know" as an information status enhancing device: Arguments from grammar and interaction'. *Functions of Language* 5: 133-155.
- Hellermann, J. and A. Vergun. 2007. 'Language which is not taught: The discourse marker use of beginning adult learners of English'. *Journal of Pragmatics* 39: 157-179.
- Heritage, J. 1984. 'A change-of-state token and aspects of its sequential placement' in Atkinson, J. M. and J. Heritage (eds.). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Heritage, J. 1998. 'Oh-prefaced responses to inquiry'. *Language in Society* 27/3: 291-334.
- Hirschberg, J. and D. Litman. 1993. 'Empirical studies on disambiguation of cue phrases'. *Computational Linguistics* 19/3: 501-530.
- Hobbs, J. 2005. 'Interactive lexical phrases in pair interview tasks' in Edwards, C. and J. Willis (eds.). *Teachers Exploring Tasks in English Language Teaching*. Basingstoke: Palgrave Macmillan.
- Holmes, J. 1986. 'Functions of *you know* in women's and men's speech'. *Language in Society* 15/1: 1-22.
- Holmes, J., B. Vine and G. Johnson. 1998. *Guide to the Wellington Corpus of Spoken New Zealand English*. Wellington: Victoria University of Wellington.
- Housen, A. 2002. 'A corpus-based study of the L2-acquisition of the English verb system' in Granger, S., J. Hung and S. Petch-Tyson (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.

- Hunston, S. and G. Francis. 1999. *Pattern Grammar*. Amsterdam: John Benjamins.
- Hyland, K. 2009. 'Corpus informed discourse analysis: The case of academic engagement' in Charles, M., D. Pecorari and S. Hunston (eds.). *Academic Writing*. London: Continuum.
- Hyland, K. and M. Bondi. (eds.) 2006. *Academic Discourse Across Disciplines*. Bern: Peter Lang.
- Johns, T. 1991. '“Should you be persuaded”: Two samples of data-driven learning materials' in Johns, T. and P. King (eds.). *Classroom Concordancing ELR Journal 4*. Birmingham: Birmingham University Press.
- Johns, T. 1994. 'From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning' in Odlin, T. (ed.) *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge University Press.
- Johns, T. and M. Scott. 1993. *MicroConcord Corpus (Collection B)*. Oxford: Oxford University Press.
- Jones, G. M. and B. B. Schieffelin. 2009. 'Enquoting voices, accomplishing talk: Uses of *be + like* in instant messaging'. *Language and Communication 29*: 77-113.
- Jones, R. 2001. 'A consciousness-raising approach to the teaching of conversational storytelling skills'. *ELT Journal 55/2*: 155-163.
- Jucker, A. H. 1993. 'The discourse marker *well*: A relevance-theoretical account'. *Journal of Pragmatics 19*: 435-452.
- Jucker, A. H. and S. W. Smith. 1998. 'And people just you know like 'wow': Discourse markers as negotiating strategies' in Jucker, A. H. and Y. Ziv (eds.). *Discourse Markers: Description and Theory*. Amsterdam: John Benjamins.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. Essex: Pearson Education.
- Knight, D., S. Adolphs, P. Tennent and R. Carter. (2008, 28 May). *The Nottingham Multi-Modal Corpus: A demonstration*. Paper presented at The 6th International Conference for Language Resources and Evaluation (LREC), Marrakech, Morocco.
- Knight, D., D. Evans, R. Carter and S. Adolphs. 2009. 'Redrafting corpus development methodologies: Blueprints for 3rd generation multimodal, multimedia corpora'. *Corpora 4/1*: 1-32.
- Knowles, G., L. Taylor and B. Williams. 1992. *A Corpus of Formal British English Speech*. London: Longman.
- Krishnamurthy, R. 1987. 'The process of compilation' in Sinclair, J. (ed.) *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Harper Collins.
- Krishnamurthy, R. 2004a. 'Editor's preface' in Krishnamurthy, R. (ed.) *English Collocation Studies: The OSTI Report John M. Sinclair, Susan Jones and Robert Daley*. London: Continuum.
- Krishnamurthy, R. 2004b. 'Interview with John Sinclair conducted by Wolfgang Teubert' in Krishnamurthy, R. (ed.) *English Collocation Studies: The OSTI Report John M. Sinclair, Susan Jones and Robert Daley*. London: Continuum.
- Lakoff, R. 1973. 'Questionable answers and answerable questions' in Kachru, B. B., R. B. Lees, Y. Malkiel, A. Pietrangeli and S. Saporta (eds.). *Issues in Linguistics. Papers in Honor of Henry and Renée Kahane*. Urbana: University of Illinois Press.
- Lenk, U. 1995. 'Discourse markers and conversational coherence'. *Anglicana Turkuensia 14* /Special Issue Organization in Discourse: 341-352.
- Lenk, U. 1998. 'Discourse markers and global coherence in conversation'. *Journal of Pragmatics 30/2*: 245-257.
- Longman Dictionary of Contemporary English* (5th edn.). 2009. Essex: Pearson Education

- Limited.
- Luke, K. K. (1987). *The establishment of common ground in natural conversation: The case of the utterance particle LA in Cantonese*. Paper presented at The 1987 International Pragmatics Conference, Antwerp.
- Luke, K. K. 1990. *Utterance Particles in Cantonese Conversation*. Amsterdam: John Benjamins.
- Lynch, T. 2001. 'Seeing what they meant: Transcribing as a route to noticing'. *ELT Journal* 55/2: 294-304.
- Müller, S. 2005. *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins.
- Macaulay, R. 2002. 'You know, it depends'. *Journal of Pragmatics* 34/6: 749-767.
- Macmillan English Dictionary for Advanced Learners* (2nd edn.) 2007. Oxford: Macmillan.
- Mark, K. L. 1998. 'The significance of learner corpus data in relation to the problems of language teaching'. *Bulletin of General Education* 312: 77-90.
- Martin, J. R. and P. R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
- Mauranen, A. 2003a. '“But here's a flawed argument”. Socialisation into and through metadiscourse' in Leistyna, P. and C. F. Meyer (eds.). *Corpus Analysis, Language Structure and Language Use*. Amsterdam: Rodopi.
- Mauranen, A. 2003b. 'The corpus of English as lingua franca in academic settings'. *TESOL Quarterly* 37/3: 513-527.
- Mauranen, A. 2006. 'Spoken discourse, academics and global English: A corpus perspective' in Hughes, R. (ed.) *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice*. Basingstoke: Palgrave Macmillan.
- Mauranen, A. 2007. 'Investigating English as a lingua franca with a spoken corpus' in Campoy, M. C. and M. J. Luzón (eds.). *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang AG.
- McCarthy, M. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M., J. McCarten and H. Sandiford. 2005. *Touchstone Teacher's Edition 1*. Cambridge: Cambridge University Press.
- McCarthy, M., J. McCarten and H. Sandiford. 2006. *Touchstone Student's Book 1*. Cambridge: Cambridge University Press.
- McEnery, T. and A. Wilson. 2001. *Corpus Linguistics*, (2 edn.). Edinburgh: Edinburgh University Press.
- McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based Language Studies*. Oxon: Routledge.
- Meunier, F. and S. Granger. (eds.) 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- MICASE Manual*. 2007. Michigan: University of Michigan.
- Nelson, G. 1996. 'The design of the corpus' in Greenbaum, S. (ed.) *Comparing English Worldwide: The International Corpus of English*. Oxford: Oxford University Press.
- Nesi, H. and P. Thompson. 2006. *The British Academic Spoken English Corpus Manual*.
- Nesselhauf, N. 2004. *Collocation in a Learner Corpus*. Amsterdam: John Benjamins.
- O'Donnell, W. R. and L. Todd. 1991. *Variety in Contemporary English*, (2nd edn.). London: Harper Collins.
- O'Keefe, A., M. McCarthy and R. Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Osborne, J. 2008. 'Phraseology effects as a trigger for errors in L2 English: The case of more

- advanced learners' in Meunier, F. and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Östman, J.-O. 1981. *You Know: A Discourse Functional Approach*. Amsterdam: John Benjamins.
- Paquot, M. 2008. 'Exemplification in learner writing: A cross-linguistic perspective' in Meunier, F. and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Pomerantz, A. 1984. 'Agreeing and disagreeing with assessents: Some features of preferred/dispreferred turn shapes' in Atkinson, J. M. and J. Heritage (eds.). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Prodromou, L. 1996. 'Correspondence'. *ELT Journal* 50/4: 371-373.
- Prodromou, L. 2008. *English as a Lingua Franca: A Corpus-based Analysis*. London: Continuum.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rayson, P. 2011. *Log-likelihood calculator*. Accessed on 18 June 2011. Retrieved from <http://ucrel.lancs.ac.uk/llwizard.html>
- Rayson, P., B. Damon and F. Brian. 2004. 'Extending the Cochran rule for the comparison of word frequencies between corpora' in G., P., F. C. and D. A. (eds.). *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004) Louvain-la-Neuve, Belgium, March 10-12, 2004*. Louvain: Presses universitaires de Louvain.
- Redeker, G. 1990. 'Ideational and pragmatic markers of discourse structure'. *Journal of Pragmatics* 14/3: 367-381.
- Redeker, G. 1991. 'Linguistic markers of discourse structure'. *Linguistics* 29/6: 1139-1172.
- Romero Trillo, J. 1997. 'Your attention, please: Pragmatic mechanisms to obtain the addressee's attention in English and Spanish conversations'. *Journal of Pragmatics* 34: 769-784.
- Sacks, H. 1992. *Lectures on Conversation 2*. Oxford: Blackwell.
- Saslow, J. and A. Ascher. 2006. *Top Notch 1*. White Plains, NY: Pearson Education.
- Saslow, J., A. Ascher and S. C. Tiberio. 2006. *Top Notch 3 Teacher's Edition and Lesson Planner*. White Plains, NY: Longman.
- Schegloff, E. A. and H. Sacks. 1973. 'Opening up closings'. *Semiotica* 8/4: 289-327.
- Schiffrin, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Schourup, L. 1985. *Common Discourse Particles in English Conversation: Like, Well, Y'know*. New York: Garland.
- Schourup, L. 1999. 'Discourse markers'. *Lingua* 107: 227-265.
- Schourup, L. 2001. 'Rethinking well'. *Journal of Pragmatics* 33: 1025-1060.
- Scott, M. 1998. *WordSmith Tools 3*. Oxford: Oxford University Press.
- Scott, M. 2004. *WordSmith Tools 4*. Oxford: Oxford University Press.
- Scott, M. 2008. *WordSmith Tools 5*. Liverpool: Lexical Analysis Software.
- Seidlhofer, B. 2001. 'Closing a conceptual gap: The case for a description of English as a lingua franca'. *International Journal of Applied Linguistics* 11/2: 133-158.
- Seidlhofer, B. 2004. 'Research perspectives on teaching English as a lingua franca'. *Annual Review of Applied Linguistics* 24: 209-239.
- Seidlhofer, B., A. Breiteneder, T. Klimpfinger, S. Majewski and M.-L. Pitzl. 2009. *The Vienna-Oxford International Corpus of English (version 1.0 online)*. Accessed on 28

- February 2010. Retrieved from <http://voice.univie.ac.at>
- Semino, E. and M. Short. 2004. *Corpus Stylistics*. London: Routledge.
- Shirato, J. and P. Stapleton. 2007. 'Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan'. *Language Teaching Research* 11/4: 393-412.
- Siegel, M. E. A. 2002. 'Like: The discourse particle and semantics'. *Journal of Semantics* 19/1: 35-71.
- Simon-Vandenberg, A.-M. 2000. 'The functions of *I think* in political discourse'. *International Journal of Applied Linguistics* 10/1: 41-63.
- Simpson, R. C., S. L. Briggs, J. Ovens and J. M. Swales. 2002. *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Accessed on 18 February 2009. Retrieved from <http://quod.lib.umich.edu/m/micase/>
- Sinclair, J. (ed.) 1990. *Collins COBUILD English Grammar*. London: Collins.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (ed.) 2004a. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. 2004b. 'Introduction' in Sinclair, J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. 2004c. *Trust the Text. Language, Corpus and Discourse*. London: Routledge.
- Sinclair, J. (ed.) 2006. *Collins COBUILD Advanced Learner's English Dictionary*. Glasgow: HarperCollins Publishers.
- Sinclair, J. and M. Coulthard. 1975. *Towards an Analysis of Discourse. The English Used by Teachers and Pupils*. Oxford: Oxford University Press.
- Sinclair, J., R. Daley and S. Jones. 1970. *English Lexical Studies*. Report No. 5060. London: Office of Scientific and Technical Information.
- Sinclair, J. and A. Mauranen. 2006. *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam: John Benjamins.
- Sinclair, J. and A. Renouf. 1988. 'A lexical syllabus for language learning' in Carter, R. and M. McCarthy (eds.). *Vocabulary and Language Teaching*. London: Longman.
- Skehan, P. 2001. 'The role of a focus on form during task-based instruction' in Muñoz, C. C., M. L. Celaya, M. Fernández-Villanueva, T. Navés, O. Strunk and E. Tragant (eds.). *Trabajos en Lingüística Aplicada*. Barcelona: Univerbook SL.
- Sperber, D. and D. Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Stenström, A.-B. 1984. *Questions and Responses*. Lund: Gleerup.
- Stenström, A.-B. 1990. 'Lexical items peculiar to spoken discourse' in Svartvik, J. (ed.) *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.
- Stenström, A.-B. 1994. *An Introduction to Spoken Interaction*. London: Longman.
- Stubbe, M. and J. Holmes. 1995. "'You know', 'eh' and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in the sample of New Zealand English'. *Language and Communication* 15: 63-88.
- Svartvik, J. 1980. 'Well in conversation' in Greenbaum, S., G. Leech and J. Svartvik (eds.). *Studies in English Linguistics for Randolph Quirk*. London: Longman.
- Svartvik, J. and R. Quirk. 1980. *A Corpus of English Conversation*. Lund: CWK Gleerup.
- Tagliamonte, S. 2005. 'So who? Like how? Just what? Discourse markers in the conversations of young Canadians'. *Journal of Pragmatics* 37/11: 1896-1915.
- Tankó, G. 2004. 'The use of adverbial connectors in Hungarian university students'

- argumentative essays' in Sinclair, J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Tannen, D. 1989. *Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.
- Tannen, D. 2007. *Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse*, (2nd edn.). Cambridge: Cambridge University Press.
- Timmis, I. 2002. 'Native-speaker norms and international English: A classroom view'. *ELT Journal* 56/3: 240-249.
- The International Corpus of English The British Component*. 2006. London: University College London.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Trester, A. M. 2009. 'Discourse marker 'oh' as a means for realizing the identity potential of constructed dialogue in interaction'. *Journal of Sociolinguistics* 13/2: 147-168.
- Tribble, C. 1997. 'Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching' in Lewandowska-Tomaszczyk, B. and J. Melia (eds.). *Proceedings of the First International conference on Practical Applications in Language Corpora*. 106-118.
- Underhill, R. 1988. 'Like is, like, focus'. *American Speech* 63/3: 234-246.
- Wade, E. and H. H. Clark. 1993. 'Reproduction and demonstration in quotations'. *Journal of Memory and Language* 32: 805-819.
- Wang, 王. L.-F. and 祝. W.-H. Zhu. 2005. '中國學生英語口語中話語標記語的使用研究 [A corpus-based study of the use of discourse markers in Chinese EFL learners' spoken English]'. *外語研究 [Foreign Language Research]* 3: 40-44, 48.
- Watts, R. J. 1989. 'Taking the pitcher to the 'well': Native speakers' perception of their use of discourse markers in conversation'. *Journal of Pragmatics* 13/2: 203-237.
- Wen, 文. Q.-F., 梁. M.-C. Liang and 晏. X.-Q. Yen. 2008. *中國學生英語口筆語語料庫 2.0 版 [Spoken and Written English Corpus of Chinese Learners, Version 2.0]*. 北京 [Beijing]: 外語教學與研究出版社 [Foreign Language Teaching and Research Press].
- Wen, 文. Q.-f., 王. L.-f. Wang and 梁. M.-c. Liang. 2005. *中國學生英語口筆語語料庫 1.0 版 [Spoken and Written English Corpus of Chinese Learners, Version 1.0]*. 北京 [Beijing]: 外語教學與研究出版社 [Foreign Language Teaching and Research Press].
- Willis, D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching*. London: Harper Collins.
- Willis, D. 2003. *Rules, Patterns and Words*. Cambridge: Cambridge University Press.
- Willis, J. and D. Willis. 1988. *Collins COBUILD English Course: Student's Book 1*. London: Collins.
- Willis, J. and D. Willis. 1996. 'Consciousness-raising activities in the language classroom' in Willis, J. and D. Willis (eds.). *Challenge and Change in Language Teaching*. Oxford: Heinemann.
- Xu, 許. J.-J. and 許. Z.-R. Xu. 2007. '中國大學生英語口語中的互動化與詞塊研究 [Discourse management chunks in Chinese college learners' English speech: A spoken corpus-based study]'. *外語教學與研究 [Foreign Language Teaching and Research]* 39/6: 437-443.
- Yang, 楊. H.-C. and 衛. N.-H. Wei. 2005. *中國學習者英語口語語料庫建設與研究 [Construction and Data Analysis of a Chinese Learner Spoken English Corpus]*. 上海

[Shanghai]: 上海外語教育出版社 [Shanghai Foreign Language Education Press].
Zimmerman, D. H. 1998. 'Identity, context and interaction' in Antaki, C. and S. Widdicombe
(eds.). *Identities in Talk*. London: Sage.