The Hong Kong Polytechnic University
Department of English

Discourse Particles in an Intercultural Corpus of Spoken English

Phoenix Wai-ying Lam

A thesis submitted in partial fulfilment of the requirements
for the Degree of Doctor of Philosophy
December 2007

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Phoenix Wai-ying Lam

Abstract

Abstract of thesis entitled "**Discourse particles in an intercultural corpus of spoken English**" submitted by Phoenix Wai-ying Lam in December 2007 for the Degree of Doctor of Philosophy at The Hong Kong Polytechnic University.

Based on a large number of authentic examples from an intercultural corpus of spoken English, the present study investigates the relationship between the use of discourse particles and various linguistic, sociolinguistic and contextual factors. Variables such as the collocational pattern of particles, their positional distribution in an utterance, as well as the linguistic background and gender of speakers are examined. The systematic analysis of the intonational pattern of discourse particles, in particular, is made possible by the prosodic transcription of the corpus. The wide range of text types in the data also presents an opportunity to look at the distribution of frequency rates and discourse functions of particles across different contextual settings. In order to study the influence of linguistic background of speakers in greater detail, and to verify the core findings generated from the intercultural corpus, a customised subset of texts from a reference corpus of British English is consulted. Conclusions drawn from these two corpora are then contrasted with the descriptions and presentations of discourse particles in English textbooks to determine to what extent the teaching of discourse particles in textbooks reflects real-world usage.

Two of the most frequently occurring English discourse particles in the spoken language, namely *well* and *so*, are analysed in the present study. Results from the intercultural corpus show that the two discourse particles under investigation serve a variety of discourse functions in the textual, interpersonal and interactional domains. In terms of first language and gender influence, the two particles display varying degrees of sociolinguistic variation. The research also demonstrates the importance of taking into account a range of linguistic and contextual features in the interpretation of discourse particles. These core findings as regards the pragmatic functions of particles and the effect of first language of speakers are largely substantiated by the customised British corpus. In reference to the pedagogical aspect of the present study, the fact that a wide discrepancy is observed between textbooks and naturally-occurring data in the use of discourse particles has significant implications for the improvement of teaching materials for spoken English.

Publications arising from the thesis

Forsyth, R., Clarke, D. & Lam, P. (2008). Timelines, talk and transcription: A chronometric approach to simultaneous speech. *International Journal of Corpus Linguistics*, 13(2), 225-250.

Lam, P. (2006). *Well but that's the effect of it*: The use of *well* as a discourse particle in talk shows. *Sprache und Datenverarbeitung (International Journal for Language Data Processing)*, 30(1), 99-108.

Papers submitted to journals

Lam, P. A needle in a haystack: The making of a BNC customised corpus. (manuscript submitted to the journal *Corpora* under review).

Lam, P. Discourse particles in corpus data and textbooks: The case of *well*. (manuscript submitted to *Applied Linguistics* under review).

Acknowledgements

Education is an admirable thing, but it is well to remember from time to time that nothing that is worth knowing can be taught. (Oscar Wilde)

I am greatly indebted to Professor Winnie Cheng and Professor Martin Warren for their inspiring and stimulating teaching since my undergraduate years. They are instrumental in making my study possible in the first place and have always been unfailingly supportive to me throughout my time as a student. Their critical, insightful and constructive comments have been invaluable to my study as much as to me as a person. To both of them I am forever grateful. My thanks are also due to everyone at the Department of English, The Hong Kong Polytechnic University, for helping me and for giving me academic advice, administrative assistance and technical support whenever I needed them.

Over the course of my doctoral study, I was privileged to be substantially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. B-Q714) as well as awarded a generous scholarship from The Hong Kong Polytechnic University, for both of which I am grateful.

I would also like to express my heartfelt gratitude to all the staff members and doctoral students at the Department of English, Göteborg University, Sweden, for giving me a wonderful academic and social experience during my three-month academic exchange there and for making me feel at home in Sweden. Special thanks are due to Professor Karin Aijmer, who is an inspirational figure with exceptional kindness and whom I greatly admire both academically and personally.

It must be said that any errors that remain are my own.

On a personal note, I would like to thank my partner Richard for making my life meaningful and transforming it from gloom to glory, and above all, for your precious love which I could never fully repay.

Table of contents

Abstract **iv**

Publications arising from the thesis **v**

Acknowledgements **vi**

Table of contents **vii**

List of figures **xiii**

List of tables **xiv**

List of abbreviations **xix**

Transcription notation **xx**

Chapter 1

Introduction **1**

1.1 Background to the study 1

1.2 Rationale for the study 2

1.2.1 Why study discourse particles? 2

1.2.2 Why study the prosody of discourse particles? 3

1.2.3 Why study discourse particles in intercultural communication? 4

1.2.4 Why study discourse particles in relation to pedagogy? 5

1.3 Purpose, research questions and scope of the present study 6

1.4 Outline of the study 7

Chapter 2

Literature Review **8**

2.1 Overview 8

2.2 The notion of discourse particles and related items: an ongoing debate 9

2.2.1 Terminology 9

2.2.2 Definition 11

2.2.3 Members of the class 12

2.3 Properties of discourse particles 13

2.3.1 No or little propositional meaning 14

2.3.2 Syntactic optionality 14

2.3.3 Indexicality 15

2.3.4 Multi-functionality 16
2.3.5 Non-truth-conditionality 18
2.3.6 Positioning 19
2.3.7 Syntactic diversity 21
2.3.8 Orality 22
2.4 Major approaches to discourse particles 23
2.4.1 Relevance theory 23
2.4.2 Coherence-based theory 26
2.4.3 Speech act theory 28
2.4.4 Theory of politeness 30
2.4.5 Corpus-linguistic approach 31
2.5 Functions of discourse particles 33
2.6 Prosodic patterns of discourse particles 36
2.7 Discourse particles and language learning 38
2.8 Discourse particles in grammar books and dictionaries 40
2.8.1 Traditional grammars and dictionaries 41
2.8.2 Corpus-based grammars and dictionaries 42
2.8.3 Linear Unit Grammar (LUG) 44
2.9 The present study 45

Chapter 3

Data description and research methodology 47

3.1 Data description 47
3.1.1 The Hong Kong Corpus of Spoken English (prosodic) 47
3.1.2 The customised corpus from the British National Corpus (BNC) 55
3.1.3 The textbook database 60
3.2 Research methodology 65
3.2.1 Data selection and pre-processing 66
3.2.2 Data retrieval / extraction 66
3.2.3 Codification 69
3.2.4 Functional analysis 75
3.2.5 Data evaluation 77
3.2.6 Statistical analysis 81

Chapter 4

Study of *well* **87**

4.1 Previous studies of *well* 87

4.2 Distinguishing the discourse use and propositional use of *well* 92

4.3 *Well* in the HKCSE 94

4.3.1 Frequency of occurrence 94

4.3.2 Position (utterance and tone unit) 97

4.3.2.1 Position in utterance 97

4.3.2.2 Position in tone unit 100

4.3.3 The prosody of *well* 102

4.3.4 Discourse functions of *well* identified in the HKCSE 106

4.3.4.1 Frame (framing) 107

4.3.4.2 Link (linking) 113

4.3.4.3 Responsive signal (responsive) 114

4.3.4.4 Expressing feelings (emotive) 119

4.3.4.5 Processing device (processing) 121

4.3.4.6 Turn management (turn managing) 123

4.3.4.7 Unclassified instances 125

4.3.5 Linguistic factors associated with the functions of *well* 126

4.3.5.1 Collocations 126

4.3.5.2 Positions 128

4.3.5.3 Prosodic features 131

4.3.6 Sociolinguistic and contextual factors associated with the functions of *well* 134

4.3.6.1 Language background 134

4.3.6.2 Gender 137

4.3.6.3 Domain and text type 140

4.3.7 Summary of findings in the HKCSE 142

4.4 Comparison of findings in the HKCSE and the BNC customised corpus 143

4.4.1 Frequency of occurrence 143

4.4.2 Functional analysis 145

4.4.3 Collocations 148
4.4.4 Utterance positions 151
4.4.5 Domain and text type 153
4.4.6 Summary of findings and comparisons between the HKCSE and the BNC (customised) 156
4.5 Comparison of findings in authentic speech and textbooks 157
4.5.1 Frequency of occurrence 157
4.5.2 Positions 161
4.5.3 Functional analysis 162
4.6 Pedagogic implications 166

Chapter 5
Study of *so* **169**
5.1 Previous studies of *so* 169
5.2 Distinguishing the discourse use and propositional use of *so* 174
5.3 *So* in the HKCSE 176
5.3.1 Frequency of occurrence 177
5.3.2 Position (utterance and tone unit) 179
5.3.2.1 Position in utterance 179
5.3.2.2 Position in tone unit 183
5.3.3 The prosody of *so* 187
5.3.4 Discourse functions of *so* identified in the HKCSE 192
5.3.4.1 Frame (framing) 193
5.3.4.2 Link (linking) 197
5.3.4.3 Marker of result or consequence (consequential) 198
5.3.4.4 Responsive signal (responsive) 200
5.3.4.5 Processing device (processing) 204
5.3.4.6 Turn management (turn managing) 205
5.3.4.7 Unclassified instances 207
5.3.5 Linguistic factors associated with the functions of *so* 208
5.3.5.1 Collocations 208
5.3.5.2 Positions 210
5.3.5.3 Prosodic features 213

5.3.6 Sociolinguistic and contextual factors associated with the functions of *so* 216
5.3.6.1 Language background 216
5.3.6.2 Gender 219
5.3.6.3 Domain and text type 221
5.3.7 Summary of findings in the HKCSE 225
5.4 Comparison of findings in the HKCSE and the BNC customised corpus 226
5.4.1 Frequency of occurrence 226
5.4.2 Functional analysis 229
5.4.3 Collocations 231
5.4.4 Utterance positions 233
5.4.5 Domain and text type 237
5.4.6 Summary of findings and comparisons between the HKCSE and the BNC (customised) 239
5.5 Comparison of findings in authentic speech and textbooks 240
5.5.1 Frequency of occurrence 240
5.5.2 Positions 244
5.5.3 Functional analysis 245
5.6 Pedagogic implications 249

Chapter 6
Towards a functional framework of discourse particles **252**
6.1 Are *well* and *so* interchangeable? 252
6.2 Qualitative contrastive analysis of *well* and *so* 253
6.2.1 Frame (framing) 253
6.2.2 Link (linking) 256
6.2.3 Responsive signal (responsive) 257
6.2.4 Processing device (processing) 259
6.2.5 Turn management (turn managing) 260
6.3 Quantitative contrastive analysis of *well* and *so* 261
6.4 The general applicability of the functional framework arising from the present study 264

Chapter 7
Conclusions **266**
7.1 Summing up 266
7.2 Research questions revisited: Implications of the present study 269
7.2.1 Are English discourse particles largely without meaning? 269
7.2.2 Do discourse particles have distinctive prosodic profiles and is there a correlation between their prosodic features and discourse functions? 270
7.2.3 How do the linguistic features of discourse particles, the sociolinguistic background of speakers and the context of communicative events influence the use of discourse particles in authentic speech? 271
7.2.4 Do textbook descriptions and presentations of discourse particles conform to how they are used as observed in corpus data? 272
7.3 Teaching discourse particles: Pedagogical suggestions 274
7.4 Limitations of the present study 276
7.5 Areas for further research 276

Appendix 1
Index of the HKCSE **278**
Appendix 2
Index of the BNC (customised) **295**
Appendix 3.1
Two pages from the teaching section of the textbook database **318**
Appendix 3.2
Two pages from the sample section of the textbook database **320**

References **322**

List of figures

Figure 3.1.	The composition of the textbook database	64
Figure 3.2.	A screenshot of Concord in WordSmith Tools showing concordance lines of *well* from the academic sub-corpus of the HKCSE orthographic component	68
Figure 3.3.	A screenshot of iConc showing concordance lines of *well* from the prosodic transcription of the HKCSE	69
Figure 3.4.	A screenshot of the collocates of *so* in Concord	79
Figure 3.5.	A screenshot showing the spreadsheet database of *so* in the HKCSE with the application of multiple filters	80
Figure 4.1.	The distribution of *well* over the four domains in the HKCSE – D-use versus P-use	95
Figure 4.2.	The positional distribution of *well* in an utterance in the HKCSE – D-use versus P-use	99
Figure 4.3.	The positional distribution of *well* in a tone unit in the HKCSE – D-use versus P-use	101
Figure 4.4.	The distribution of D-use and P-use *well* by Hong Kong Chinese (HKC) and native speakers of English (NSE)	135
Figure 4.5.	The distribution of D-use and P-use *well* by female and male speakers in the HKCSE	138
Figure 5.1.	The distribution of *so* over the four domains in the HKCSE – D-use versus P-use	178
Figure 5.2.	The positional distribution of *so* in an utterance in the HKCSE – D-use versus P-use	183
Figure 5.3.	The positional distribution of *so* in a tone unit in the HKCSE – D-use versus P-use	186
Figure 5.4.	The distribution of D-use and P-use *so* by Hong Kong Chinese (HKC) and native speakers of English (NSE)	217
Figure 5.5.	The distribution of D-use and P-use *so* by female and male speakers in the HKCSE	220
Figure 6.1.	The comparison of the frequency distributions of *well* and *so* across the three functional domains in the corpus data	262

List of tables

Table 3.1.	The composition of the HKCSE by domain and text type	48
Table 3.2.	The composition of the HKCSE by the first language of speakers	50
Table 3.3.	The composition of the HKCSE by speaker gender	51
Table 3.4.	The intonation choices available in the discourse intonation framework (reproduced from Hewings and Cauldwell 1997:vii)	53
Table 3.5.	The composition of the customised corpus from the BNC	58
Table 3.6.	The list of upper-secondary English textbooks used in the study	62
Table 3.7.	The composition of the sample section in the textbook database	65
Table 3.8.	The list of features analysed in the HKCSE, the BNC (customised) and the sample section in the textbook database in the present study	70
Table 3.9.	The values for the feature “utterance position”	71
Table 3.10.	The values for the feature “tone unit position”	73
Table 3.11.	The values for the feature “prominence position”	74
Table 3.12.	A summary of the details of the features examined in the present study	74
Table 3.13.	The relationship between the value of Cramér’s V and the strength of the association (adapted from Healey 2007:263)	83
Table 3.14.	An example of a statistically different result owing to a particular category (reproduced from Zar 1999:503)	83
Table 3.15.	The 2 x 3 contingency table formed by omitting column 3 and keeping columns 1, 2, and 4 of Table 3.14 (adapted from Zar 1999:503)	84
Table 3.16	The 2 x 2 contingency table formed by keeping column 3 and combining columns 1, 2, and 4 of Table 3.14 (reproduced from Zar 1999:504)	84
Table 4.1.	The distribution of *well* in the HKCSE	94
Table 4.2.	The positional distribution of *well* in an utterance in the HKCSE	97
Table 4.3.	The positional distribution of *well* in a tone unit in the HKCSE	100

Table 4.4.	The prominence pattern of *well* in the HKCSE	102
Table 4.5.	The distribution of nuclear tones on *well* in the HKCSE	104
Table 4.6.	The comparison of tonal distribution on D-use *well* in Svartvik (1980:170) and in the HKCSE	104
Table 4.7.	The distribution of pauses before and after *well* in the HKCSE	105
Table 4.8.	The distribution of *well* within a span of five words after the two most common reporting verbs *say* and *think* in direct speech quotation	126
Table 4.9.	The positional distribution of the discourse functions of *well* in an utterance in the HKCSE	128
Table 4.10.	The positional distribution of the discourse functions of *well* in a tone unit in the HKCSE	130
Table 4.11.	The prominence distribution of the discourse functions of *well* in the HKCSE	132
Table 4.12.	The distribution of nuclear tones for the discourse functions of *well* in the HKCSE	133
Table 4.13.	The distribution of *well* in the HKCSE according to speakers' linguistic background (Hong Kong Chinese vs. native speakers of English)	134
Table 4.14.	The distribution of the discourse functions of *well* by linguistic background in the HKCSE (Hong Kong Chinese vs. native speakers of English)	136
Table 4.15.	The distribution of *well* in the HKCSE according to speaker gender (male vs. female)	137
Table 4.16.	The distribution of the discourse functions of *well* by gender in the HKCSE (female vs. male speakers)	139
Table 4.17.	The functional distribution of D-use *well* by domain in the HKCSE	140
Table 4.18.	The functional distribution of D-use *well* across some major text types in the HKCSE	141
Table 4.19.	The distribution of *well* in the BNC (customised)	143
Table 4.20.	The distribution of *well* in the HKCSE (reproduced from Table 4.1)	144

Table 4.21.	The distribution of the discourse functions of *well* in the HKCSE and the BNC (customised)	145
Table 4.22.	The distribution of *well* after the two most common reporting verbs *say* and *think* in direct speech quotation in the HKCSE and the BNC (customised)	148
Table 4.23.	The positional distribution of D-use *well* in an utterance in the HKCSE and the BNC (customised)	151
Table 4.24.	The positional distribution of the discourse functions of *well* in an utterance in the BNC (customised)	152
Table 4.25.	The functional distribution of D-use *well* by domain in the BNC (customised)	154
Table 4.26.	The functional distribution of D-use *well* across some major text types in the BNC (customised)	155
Table 4.27.	The comparison of *well* in the textbooks, the HKCSE and the BNC (customised)	158
Table 4.28.	The comparison of the discourse rate of *well* in presentations in the HKCSE, the BNC (customised) and the textbook database	159
Table 4.29.	The comparison of the discourse rate of *well* in discussions in the HKCSE, the BNC (customised) and the textbook database	160
Table 4.30.	The positional distribution of D-use *well* in an utterance in the HKCSE, the BNC (customised) and sample texts in the textbook database	161
Table 4.31.	The functional distribution of *well* in the sample presentation and discussion texts in the textbook database	165
Table 5.1.	The distribution of *so* in the HKCSE	177
Table 5.2.	The positional distribution of *so* in an utterance in the HKCSE	180
Table 5.3.	The positional distribution of *so* in a tone unit in the HKCSE	184
Table 5.4.	The prominence pattern of *so* in the HKCSE	187
Table 5.5.	The distribution of nuclear tones on *so* in the HKCSE	189
Table 5.6.	The comparison of tonal distribution on D-use *so* in the HKCSE and D-use *oh* in the London-Lund Corpus (Aijmer 2002:109)	190
Table 5.7.	The distribution of pauses before and after *so* in the HKCSE	191

Table 5.8.	The positional distribution of the discourse functions of *so* in an utterance in the HKCSE	210
Table 5.9.	The positional distribution of the discourse functions of *so* in a tone unit in the HKCSE	212
Table 5.10.	The prominence distribution of the discourse functions of *so* in the HKCSE	214
Table 5.11.	The distribution of nuclear tones for the discourse functions of *so* in the HKCSE	215
Table 5.12.	The distribution of *so* in the HKCSE according to speakers' linguistic background (Hong Kong Chinese vs. native speakers of English)	216
Table 5.13.	The distribution of the discourse functions of *so* by linguistic background in the HKCSE (Hong Kong Chinese vs. native speakers of English)	218
Table 5.14.	The distribution of *so* in the HKCSE according to speaker gender (male vs. female)	219
Table 5.15.	The distribution of the discourse functions of *so* by gender in the HKCSE (female vs. male speakers)	221
Table 5.16.	The functional distribution of D-use *so* by domain in the HKCSE	222
Table 5.17.	The functional distribution of D-use *so* across some major text types in the HKCSE	224
Table 5.18.	The distribution of *so* in the BNC (customised)	227
Table 5.19.	The distribution of *so* in the HKCSE (reproduced from Table 5.1)	227
Table 5.20.	The distribution of the discourse functions of *so* in the HKCSE and the BNC (customised)	229
Table 5.21.	The positional distribution of D-use *so* in an utterance in the HKCSE and the BNC (customised)	234
Table 5.22.	The positional distribution of the discourse functions of *so* in an utterance in the BNC (customised)	235
Table 5.23.	The functional distribution of D-use *so* by domain in the BNC (customised)	237

Table 5.24.	The functional distribution of D-use *so* across some major text types in the BNC (customised)	238
Table 5.25.	The comparison of *so* in the textbooks, the HKCSE and the BNC (customised)	241
Table 5.26.	The comparison of the discourse rate of *so* in presentations in the HKCSE, the BNC (customised) and the textbook database	242
Table 5.27.	The comparison of the discourse rate of *so* in discussions in the HKCSE, the BNC (customised) and the textbook database	243
Table 5.28.	The positional distribution of D-use *so* in an utterance in the HKCSE, the BNC (customised) and sample texts in the textbook database	244
Table 5.29.	The functional distribution of *so* in the sample presentation and discussion texts in the textbook database	249
Table 6.1.	Comparison of discourse functions of *well* and *so*	253
Table 6.2.	The comparison of the frequency distributions of the discourse functions of *well* and *so* in the corpus data	261

List of abbreviations

BNC	British National Corpus
BNC (customised)	Customised corpus from the British National Corpus
CANCODE	Cambridge and Nottingham Corpus of Discourse in English
D-use	Discourse use(s)
D-rate	Discourse rate
DP	Discourse particle
HKC	Hong Kong Chinese
HKCSE	Hong Kong Corpus of Spoken English
NSE	Native speakers of English
P-use	Propositional use(s)
POS	Part-of-speech
SPSS	Statistical Package for the Social Sciences

Transcription notation

The Hong Kong Corpus of Spoken English (HKCSE)

The orthographic transcription of the HKCSE mainly follows Jefferson's system of transcribing speech to written text (for details, see Atkinson and Heritage 1984). The prosodic transcription of the HKCSE is based on the discourse intonation system developed in Brazil (1985; 1997). The transcription conventions used for data in the HKCSE are explained as follows:

Symbol	Remarks
…	parts of an utterance which have been omitted
*	onset of simultaneous speech produced by the current speaker
**	onset of simultaneous speech produced by an interlocutor other than the current speaker
(.)	a brief, unfilled pause roughly lasts for the length of a syllable
(pause)	a unit, unfilled pause which is longer than a brief pause and normally lasts for a few seconds
(( ))	a non-linguistic feature such as laughter, coughing, throat clearing and applause
((inaudible))	unintelligible speech
A:	female native speaker of English
B:	male native speaker of English
a:	female Hong Kong Chinese
b:	male Hong Kong Chinese
x:	female speaker of a language other than English and Cantonese
y:	male speaker of a language other than English and Cantonese
u:	unknown speaker
{ }	tone unit boundary
/	rise tone
\/	fall-rise tone
\	fall tone
/\	rise-fall tone
=	level tone
?	unclassifiable tone
CAPS	prominent syllable
[ ]	key
< >	termination
^	high pitch level
_	low pitch level

The customised corpus from the British National Corpus (BNC)

Examples cited from the BNC are taken verbatim, with most of the tags removed, including the part-of-speech (POS) tags and tags marking segment number and speaker identification (for details of the mark-up of the BNC, see Bernard 2000). Paralinguistic information is retained and provided in round brackets. As in the HKCSE, utterances in the BNC which are reported only in part are indicated by "…". All speakers in the BNC are marked as N: in the present study to denote native speakers of British English. A simple example of how a segment of text in the BNC is represented in this study is provided below:

Original text segment in the BNC

<u who=PS1DP>
<s n="1936"><w AV0>Right<c PUN>, <w DTQ>what <w VDB>do <w PNP>you <w VVI>want <w PRP>for <w DPS>your <w NN1>dinner<c PUN>?
</u>
<u who=PS1DS>
<s n="1937"><w AV0>Well<c PUN>, <w PNP>I<w VHB>'ve <w VVN>got <w TO0>to <w VVI>look<c PUN>.
<s n="1938"><pause dur=7> <w DTQ>What <w VDB>do <w PNP>you <pause> <w VVI>suggest<c PUN>?
</u>
(BNC customised, kbj, sn1936-38)

Representation in this study

N1: Right, what do you want for your dinner?
N2: Well, I've got to look.
(pause) What do you (pause) suggest?
(BNC customised, kbj, sn1936-38)

Chapter 1

Introduction

Abstract

In this introductory chapter which sets the scene for the present study, the background which stimulates this research work is first provided (1.1). The rationale for the focus of the study is then explained, highlighting the importance of discourse particles in communication in general, their relation to prosody, their value in intercultural communication, and their role in pedagogy (1.2). This is followed by a statement of the purpose and research questions of the research (1.3) and a general outline of the overall structure of the present study (1.4).

1.1 Background to the study

> You should be careful not to make too frequent use of particular phrases which are largely without meaning, such as 'you know', 'et cetera', 'I guess', 'O.K.', which can be very distracting and irritating to listeners (and to the examiners!) (Hong Kong Examinations Authority 1998:13, quotes in original)[1]

Surprising as it may seem, the present study originates from the rather off-putting remark above regarding a number of lexical items such as *you know* and *okay* which are sometimes referred to as 'discourse particles'. Two interesting points are made in the above remark: that discourse particles are largely without meaning and that they are very distracting and irritating to hearers if they are used too frequently. Readers of the statement above cannot help but wonder: Are discourse particles largely without meaning[2]? Are they so detrimental in discourse that they need to be dismissed and their use discouraged?

---

[1] This is a short extract from the *Guidelines for Candidates: Use of English Oral Examination* (1998), a document published by the Hong Kong Examinations Authority for students at the upper secondary level to prepare for the Use of English Oral Examination in the AS-level English examination which students are to take upon the completion of their upper secondary education.

[2] While the word 'meaning' is typically associated with semantics, the present study explores the *pragmatic* functions of discourse particles to determine whether or not they are pragmatically 'meaningless' in discourse.

The present research sets about answering the above questions. In particular, its main interest lies in the use of discourse particles in naturally-occurring texts. In other words, the present study chiefly concerns itself with how discourse particles are actually used by interlocutors in real-life communication.

1.2 Rationale for the study

Discourse particles have been approached from a variety of perspectives by different scholars in linguistic research. The present study in particular deals with three aspects which are very often neglected: their relation with prosody, their use in intercultural communication, and their role in pedagogy. The following section highlights the significance of discourse particles and discusses the reasons why the present research pays special attention to discourse particles in these three areas.

1.2.1 Why study discourse particles?

While it might remain questionable what the examination body meant by "too frequent use" in its advice above on particular words and phrases such as *you know* and *okay*, the high frequency of discourse particles, in the spoken language in particular, is indisputable. In a 50,000 word sample from the London-Lund Corpus of Spoken English (LLC), for example, discourse particles together with greetings, thanks and apologies make up about one-tenth of all word-class tokens (Altenberg 1990). Grouped collectively under the class "discourse items", they are the fourth largest word-class in the sample surpassed in rank only by verbs, pronouns and nouns (Altenberg 1990:185). Individually, discourse particles such as *yeah* often occupy high ranks in the word frequency lists of English general corpora. In the British National Corpus (BNC), *yeah* is the 19th most frequently occurring word form and is significantly more popular in speech when compared with writing (Leech, Rayson & Wilson 2001:144).

Since discourse particles are abundant in daily communication, it is not unreasonable to speculate that they have an important role to play in the language. However, negative comments such as the one above made by the Hong Kong Examinations Authority are plentiful. Not only are they seen as fillers which are "exasperating" and "unnecessary" by ordinary native speakers as reported in Stubbe and Holmes (1995:63), they are also considered in many sentence-based language studies to be "a form of linguistic detritus unworthy of close attention" (Schourup

1999:228). It has also been pointed out that language purists perceive them as “a sign of bad or sloppy usage” (O’Keeffe, McCarthy & Carter 2007:39). In everyday conversation, discourse particles are often associated with lack of proficiency and even “imprecise, uncertain or uneducated ‘lower class’ speech” in the minds of the general public (Stubbe and Holmes 1995:63, quotes in original). Indeed, stylistic stigmatisation is listed as one of the basic features of discourse particles (‘pragmatic markers’ in original) by Brinton (1996). As pointed out by Zwicky (1985:292, quotes in original), “the class of ‘particles’ is distinguished entirely negatively: particles are the words left over when all the others have been assigned to syntactic categories”.

Given that such pejorative remarks about discourse particles are common, are discourse particles verbal debris which are meaningless in the sense that they are superfluous? For many who study these elements in great detail in linguistic research, it is definitely not the case. As early as in the 1950s, Quirk (1955:178-9) discussed the use of *you know*, *you see* and *well* as “sharing devices and intimacy signals in our everyday talk” and noted that they are of “considerable importance”. This is echoed by Watts’ (1988) comment that they are an important tool to facilitate the processes of interpretation and social involvement in spoken interaction. More recently, Leech and Svartvik (2002) emphasized that conversational cooperation is largely made possible by the use of discourse particles. In fact, although the omission of discourse particles often renders a text grammatically acceptable, many have pointed out that the text may become inarticulate and incoherent, resulting in a greater possibility of communicative failure (see, for example, Fraser 1990). Without such particles in the communicative context, discourse would be “unnatural”, “awkward”, “unfriendly” or “dogmatic” (Brinton 1996:35-6). The views presented above suggest that discourse particles do not occur in a haphazard way. On the contrary, they are indispensable in lubricating interactions to make sure that they go on smoothly. Hence it is no coincidence that discourse particles are ubiquitous in discourse.

1.2.2 Why study the prosody of discourse particles?

Discourse particles are often considered a common feature of natural speech and one of the defining characteristics of speech is the transmission of sound and the presence of intonation. It has often been proposed that discourse particles have distinctive prosodic patterns and intonation is a significant aspect in the study of discourse particles. In his comment on these linguistic items, Altenberg (1990:180) argues that

"although their function is generally signalled both positionally and prosodically, the prosodic distinction is often the more important one". This highlights the importance of prosodic information in the identification of discourse particles from their form-sharing counterparts. Yet despite such importance, the prosodic properties of English discourse particles have only been given a patchy treatment so far. Little empirical work has been done, let alone large-scale studies based on a substantial number of instances of discourse particles in authentic data, with only a few exceptions (see Chapter 2 for a review of related studies). Aijmer's (2002) comprehensive account of discourse particles, for example, demonstrates the value of prosody in the interpretation of functions and the expression of emotion by drawing on a large quantity of examples from corpus data. However, the relationship between discourse particles and prosody is still far from being thoroughly investigated. More research work thus needs to be devoted to this area which has long been acknowledged as being fundamental but inadequately studied.

1.2.3 Why study discourse particles in intercultural communication?

The last decade has seen the rapid increase in intercultural communication because of factors such as globalization, the popularity of the Internet and the upsurge of budget airlines. This is particularly so in multicultural societies where the population is made up of different ethnic groups. As one of the world's major languages for international communication, English still holds a leading position as a lingua franca among the global languages, though its dominance is gradually diminishing (Graddol 2006). Given the prevalence of intercultural encounters in English, it is imperative that communication in English between different cultures should no longer be considered exceptional or incidental. The estimation that second language users of English outnumber native speakers (Graddol 1997) shows that intercultural studies probably have wider applications than generally assumed.

In Hong Kong where the present study is carried out, communication between people of different cultural backgrounds has always been common as a result of international trade and historical influences. The transformation of the city from a manufacturing-based society to a serviced-oriented metropolis in the past few decades has called for more professionals who are able to communicate effectively in English in an ever-expanding multicultural business environment (Bilbow 1997). Historically, Hong Kong had been a British colony for more than a century and a half

until the year 1997 when it was returned to Chinese rule. Owing to this former colonial history, English was and still is one of the two official languages in Hong Kong. Children start to learn English as early as in kindergarten, when they are approximately three years of age. In primary and secondary education, English is a compulsory subject. It is also the medium of instruction in many secondary schools and tertiary institutions. Apart from the educational sector, English also plays an important role in other professional settings such as the legal and medical sectors, and as mentioned earlier, it is being widely used by the business community (see, for example, Evans and Green 2001). Naturally, it follows that English is the predominant language in intercultural interactions in Hong Kong (Cheng 2003).

Despite the remarkable growth of intercultural communication all over the world including Hong Kong and the importance given to cultural influences, features of language such as discourse particles which impart cultural values have not been sufficiently examined. As pointed out by Wierzbicka (1991:341), "[t]here are few aspects of any language which reflect the culture of a given speech community better than its particles". The study of discourse particles in intercultural communication, therefore, provides us with an opportunity to look at the similarities and differences between cultural groups in the use of these lexical items which may in turn further our understanding of their cultural values.

1.2.4 Why study discourse particles in relation to pedagogy?

In the pedagogical setting, the examination body in Hong Kong is not alone in making unfavourable judgements on discourse particles. Disparaging comments by relevant stakeholders such as teachers and materials designers are common (Erman 1987) and discourse particles are often dismissed as a sign of dysfluency. This derogatory attitude towards discourse particles may account for the general impression that discourse particles are by and large overlooked in pedagogy, though it has been noted in many studies that the use of discourse particles contributes to the pragmatic and communicative competence of speakers (see, for example, Müller 2005; Wierzbicka 1991) and hence is an essential aspect of language that learners should master.

However, is the general impression that discourse particles are shunned in the academic sector validated by research? In the literature, very few studies investigate whether discourse particles are part of language teaching, and if so, in what ways

they are taught. At the same time, it appears that little attempt has been made in the field of second language acquisition to concentrate on discourse particles as a linguistic resource for learners. Research which offers a link between discourse particles and pedagogy by examining both the use of discourse particles in 'real' English and their descriptions in textbook English should therefore be valuable and could have implications for further improvement of teaching materials for enhancing the pragmatic competence of learners. This is especially relevant in the context of Hong Kong, where "English is almost entirely a 'learnt' language" (Bolt and Bolton 1995:3, quotes in original; see also Bacon-Shone and Bolton 1998; Tsui 2004) which is chiefly learned in the classroom setting with textbooks constituting a major part in English language teaching (Evans 1997).

1.3 Purpose, research questions and scope of the present study

Based on the rationale above, the present study attempts to explore the use of English discourse particles in naturally-occurring spoken interactions from corpus data and in textbooks. Specifically, the following research questions will be addressed:

1. Are English discourse particles largely without meaning?
2. Do discourse particles have distinctive prosodic profiles and is there a correlation between their prosodic features and discourse functions?
3. How do the linguistic features of discourse particles, the sociolinguistic background of speakers and the context of communicative events influence the use of discourse particles in authentic speech?

To tackle this question, factors including collocations and positions of discourse particles, language background and gender of speakers, as well as domain and text type of communicative events will be examined. The final research question relates the present study to pedagogy:

4. Do textbook descriptions and presentations of discourse particles conform to how they are used as observed in corpus data?

Two discourse particles, namely *well* and *so*, constitute the central focus of the present study. They are selected for both practical and theoretical reasons. On the practical side, *well* and *so* are two of the most frequently occurring discourse particles in the data examined and in the English language in general. In other words, not only are they commonly found in the corpus data and sufficiently represented in the pedagogical materials currently under investigation, they are also among the most

frequent items in English and constantly occupy top rankings in frequency lists of major English corpora (see Leech, Rayson & Wilson 2001; O'Keeffe et al. 2007). They are thus representative examples of discourse particles in terms of their high frequencies. On the theoretical side, the two discourse particles have not been studied to the same extent before in the literature. While *well* has been extensively researched and thus perceived as an archetypal example of discourse particles, *so* has not been given so much emphasis in particle studies. An analysis of these two items could therefore provide results which are comparable to previous studies as well as potentially revealing.

1.4 Outline of the study

Following this introductory chapter, a literature review of some key areas in the field of discourse particles and of issues which are most relevant to the present study is presented in Chapter 2. Chapter 3 gives a detailed description of the data and introduces the research methodology adopted in the study. Chapters 4 and 5 provide in-depth analyses of the two particles *well* and *so* respectively based on empirical results drawn from corpus data and textbooks. Major findings on *well* and *so* are compared in Chapter 6, which have implications for the study of other particles and for the development of a functional framework for discourse particles as a class. Chapter 7 summarises the main findings of the study and suggests areas for further research.

Summary

At the beginning of this chapter, the background which instigates the present study is described. Reasons for the particular focus of the study are then described, which highlight the importance of discourse particles, the value of intonation in particle studies, the prevalence of intercultural communication and the relation between discourse particles and pedagogy. This is followed by a discussion of the purpose of study and the questions to be addressed in this research and a brief outline of the structure of the present study.

Chapter 2

Literature review

Abstract

This chapter discusses some of the key issues in the study of discourse particles and reviews some previous research work in areas which are particularly relevant to the present study. Following a brief introductory overview of the field (2.1), the chapter first considers some controversial notions in particle research (2.2) and examines some common properties of discourse particles reported in earlier studies (2.3). Major approaches which have been used to analyse these lexical items are then described (2.4). This is followed by a functional account of discourse particles (2.5). In the review of previous studies, special emphasis is given to aspects which are often neglected in the field, including the prosodic patterns of discourse particles (2.6) and the relation between discourse particles and language learning (2.7), with a section devoted to their treatment in grammar books and dictionaries (2.8). The final section gives a short summary of the chapter and suggests how the present study could contribute to some aspects of research in the current literature which require further investigation (2.9).

2.1 Overview

Discourse particles have long been an intriguing subject in linguistic research. Since the 1970s, discourse particles have become a focus for a considerable number of studies, alongside the surge of interest in how discourse is interpreted in pragmatic and contextual aspects (Schourup 1999). Discourse particles have also been a challenging subject for researchers. As will be shown in the following sections, discussions on certain major issues concerning discourse particles have reached little consensus within the research community, even down to the basic definition of what discourse particles are and what term should be used for these elements.

In the literature, discourse particles have been analysed in different ways. Some research studies investigate a collective range of discourse particles (see, for example, Aijmer 2002; Fraser 1990; Lenk 1998; Schiffrin 1987; Schourup 1985) while others concentrate on individual ones (see, for example, Lakoff 1973; Erman 2001). In some cases discourse particles are examined within a single analytical framework

from a top-down approach, specifically concentrating on how they fit in a particular theoretical model (see, for example, Jucker 1993; Lenk 1995; 1998). In other cases, discourse particles are studied using a data-driven approach, with extra attention given to the inspection of each single instance in authentic texts prior to making generalisations (Aijmer 2002; Müller 2005). This bottom-up approach allows analysts to take data as the starting point and work their way through to a thorough description of discourse particles. This chapter reviews some of the major studies in the field and describes research outputs in areas which are especially relevant to the present study. Despite the fact that abundant scholarly effort has been made in the study of particles in many different languages, this chapter chiefly reports the key findings of research work in English discourse particles, which are the focus of the present investigation.

2.2 The notion of discourse particles and related items: an ongoing debate

There are at least three contentious issues concerning the notion of discourse particles. First, a variety of terms are used to represent these mysterious elements. In other words, there is a lack of a universally accepted terminology which most people working in the field give consent to. Second, even when the same term is employed among different researchers, no general agreement is reached upon what the term actually describes and how it should be defined. Finally, there is little consensus regarding which elements in a language are members of the class, or if it is at all possible to group these items into a single set (Schiffrin 1987). This section takes a look at how these three challenges are dealt with in previous studies of discourse particles.

2.2.1 Terminology

Apart from the term ‘discourse particle’ (see, for example, Aijmer 2002; Fischer 2006; Goldberg 1980; Schourup 1985), there is no shortage of contenders which are used in similar and often partly overlapping ways in linguistic research. The name ‘discourse marker’ is probably the most commonly employed (see, for example, Schiffrin 1987; Schourup 1999; Stenström 1994; 2006). Other examples include but are not limited to ‘connective’ (see, for example, Bazzanella 1990; Fraser 1988), ‘discourse operator’ (see, for example, Redeker 1991), ‘pragmatic expression’ (see, for example, Erman 1987), ‘pragmatic marker’ (see, for example, Aijmer 2005;

Brinton 1996) and ‘utterance particle’ (see, for example, Luke 1987; 1990). Each is employed with subtle differences from the others, depending on the research design and interest. As shown in Brinton (1996), an inventory for naming the items under discussion could include more than twenty terms, all of which are subject to objections from other researchers.

Being two of the most popular terms in the list, ‘discourse particles’ and ‘discourse markers’ share the same connotation. Both refer to items which function at the discoursal level. However, some researchers state that by choosing one term over the other, there is a difference in approach. Risselada and Spooren (1998) remark that while studies on particles focus on the semantics and the role particles play in reflecting the attitudes of speakers and the illocutionary force of utterances, studies on markers are based on discourse analysis and oriented towards the use of corpus-based data. This may well be a proper distinction at the time of writing, but studies which prove otherwise abound in the current literature. Aijmer’s (2002) empirical study of discourse particles, for example, is based on data from the London-Lund Corpus. Not only does it examine the attitudinal meanings of particles but it also analyses their textual meanings. Supporters of the term ‘discourse marker’ also feel that the label ‘discourse particle’ is problematic because the term ‘particle’ has traditionally been associated with syntactic classes while the elements under discussion are more often regarded as a functional class (Schourup 1999). In addition, due to the indeterminate nature of the term ‘particle’, discourse particles tend to include expressions which are not ordinarily regarded as discourse markers (ibid.). Categories which are cited as non-discourse markers include interjections such as *oh*, pause markers like *uh* and so on (see, for example, Fraser 1988 and 1999; Schourup 1999). Nevertheless, research studies which deal with such expressions under the umbrella term ‘discourse marker’ are not uncommon (see, for example, Andersen, Brizuela, DuPuy & Gonnerman 1999; Redeker 1990). Admittedly, there may be a tendency to associate a particular term with certain senses and features, for example, by using the term ‘discourse particle’ to refer to particles in German, Dutch and Norwegian which have distinct functions from those elements in English. However, given that each term is employed by different researchers without much consensus, and each appears to refer to a wide range of linguistic items in various situations according to the design and approach of particular research, there is no strong reason

to suggest that one expression is more preferable and desirable than the other in absolute terms[3].

2.2.2 Definition

As mentioned above, there is no general agreement on how a particular term is used. Hence, even when the same term is employed, there is no guarantee that it is defined in the same way in different studies. This mainly arises from the fact that the topic of discourse particles has attracted interest from a rich variety of disciplines and thus is analysed within different approaches which represent divergent research concerns and goals (see Section 2.4 for details). The multifunctional nature of discourse particles, which results in multiple readings for a single item, also makes it difficult to resolve the definitional issue. Although discourse particles are usually defined from a functional perspective, taxonomies differ considerably[4]. Added to this difficulty is the complexity of generalising the defining features across different types of discourse particles in a single language, let alone across languages (Aijmer 2002).

In one of the most influential and widely-quoted studies of discourse particles, Schiffrin (1987:31) operationally defines discourse markers as "sequentially dependent elements which bracket units of talk". After a thorough analysis of her set of markers, she modifies the definition by stating that "markers propose the contextual coordinates within which an utterance is produced and designed to be interpreted" (Schiffrin 1987:315). Such contextual coordinates, according to Schiffrin, operate on various levels of discourse[5]. In another study, Lenk (1998:52) defines discourse markers as "short lexical items" which are used "to signal for the hearer how the speaker intends the present contribution to be related to preceding and/or following parts of the discourse". This definition is, apparently, oriented towards how these elements function in discourse organisation. Under the same expression 'discourse marker', however, these elements can also be defined as signals of attitudes and intentions (see, for example, Andersen et al. 1999; Redeker

[3] For a detailed discussion on terminology between the two labels 'discourse particle' and 'discourse marker', see also Fischer (2006).
[4] Attempts to define a special syntactic class of discourse particles (see, for example, Őstman, 1982), on the whole, appear to be problematic. In terms of form and structure, it seems that no single features or set of features could define such a word-class (Aijmer, 2002).
[5] For a more detailed discussion of Schiffrin's (1987) framework, see Section 2.4.

1990), which is inclined towards the interpersonal aspect of communication. As a result, items bounded by one definition may just overlap minimally with those designated under another definition using the same term (Schourup 1999). Thus far, as pointed out by Lenk (1998:37, italics original), "not one single definition of the term *discourse marker* remained undisputed or unaltered by other researchers for their purposes". The same is true for the term discourse particle and other similar terms.

2.2.3 Members of the class

As with terminology and definition, the issue of what linguistic items are considered to constitute the set of discourse particles (or whatever names they are given) is also disputable. While units such as *well*, *now* and *I mean* are often cited as examples of discourse particles with little disagreement from the research community[6], elements such as connectives (for example, *moreover* and *after all*) and sentence adverbials (for example, *frankly* and *confidentially*), as described in Blakemore (1987) and Rouchota (1998) respectively, are more doubtful[7]. It appears that there are features which disparate discourse particles have in common but there are also characteristics which are not shared by all putative items of the group. In this respect, it is useful to consider the class of discourse particles as a continuum with prototypical members on one end and peripheral members on the other. Following the Roschean concept of prototype (Jucker and Ziv 1998), elements which exhibit more properties matching the criteria used to define discourse particles are oriented towards the core end of the cline. On the other hand, items which display fewer characteristics could be treated as peripheral members of the class. However, since words in the traditional word classes could develop discourse functions over time while some discourse particles fall out of fashion, and new markers or particles could come into existence (see, for example, the use of *innit* in Stenström and Andersen 1996), the composition of the membership could be subject to fluctuation. As a result, it is difficult, if at all possible, to generate an exhaustive list of discourse particles (see, for example,

[6] An exception could be found in Fraser (1988) where *I mean* is not categorised as a discourse marker, based on the view that it does not signal a relationship between "the current utterance and the foregoing context" (1988:26).

[7] Schourup (1999), for example, does not consider elements such as *frankly* and *confidentially* which he terms "illocutionary adverbials" (1999:231) to be discourse markers on the grounds that they do not connect utterances or other discourse units.

Bazzanella 1990). Although the issue of what particular linguistic items should be considered discourse particles remains ambiguous and fuzzy, there are basic features which are generally accepted as properties of the set. The idea of placing members of discourse particles on a scale thus accounts for the possible variations and overlapping in features among different items in the class.

As evidenced from the above discussion, these terminological and definitional issues present challenges to researchers working in the field. The present study does not intend to complicate the labelling debate by adding one more candidate to the already bulky list. Until more is known about the nature of these items and what they constitute as a group, any term used to represent them is merely a pre-conceptual expression for reference. As one of the more popular labels in the literature, discourse particle "suggests a focus on small, uninflected words that are only loosely integrated into the sentence structure, if at all" (Fischer 2006:4). This appears to be a suitable cover term for the two lexical items *well* and *so* under investigation in this research. The fact that this study follows closely the corpus-linguistic approach used by Aijmer (2002) in her study of discourse particles also makes the selection of this term appropriate for the present purpose. For the sake of consistency, the name 'discourse particle' will be used throughout the present study.

2.3 Properties of discourse particles[8]

With the above disputes in mind, it is now helpful to move towards what properties are, in general, taken to be associated with discourse particles. Similar to the way that discourse particles could be prototypical or peripheral, the characteristics of discourse particles could also be classified as either more fundamental or marginal, depending on their popularity among the research community. As will be shown in the discussion below, certain features, as described in Sections 2.3.1 to 2.3.3, are generally considered necessary attributes of discourse particles. Other features, as in Sections 2.3.4 and 2.3.5, are considered common but non-defining properties. The remaining features in this discussion, as reported in Sections 2.3.6 to 2.3.8, are mostly descriptive and serve little diagnostic purposes (cf. Jucker and Ziv 1998;

[8] For a comprehensive list of characteristics of pragmatic markers, also see Brinton (1996).

Schourup 1999). The following illustrates some of the key properties reported in the literature[9].

2.3.1 No or little propositional meaning

One of the defining features which identify an element as a discourse particle is that it carries no or little propositional meaning. In other words, the presence of a discourse particle in an utterance does not affect its propositional content. This semantic characteristic possibly leads to the notorious reputation that discourse particles are useless and redundant. According to Östman (1982:153), for example, one of the principles for an element to be what he calls a 'pragmatic marker' is that it "does not directly partake in the propositional content of an utterance" but "implicitly anchors" the propositional meaning. This view is echoed by Schiffrin's (1987:321) claim that discourse particles do not create propositional connections; rather they help hearers "select, and then display, structural relations between utterances". While this property remains one of the key selection criteria for discourse particles (see, for example, Fraser 1988; Jucker and Ziv 1998; Lenk 1998; Schourup 1999), it is nevertheless not a criterion unanimously embraced. Erman (1987), for instance, argues that although the literal meaning of discourse particles may only be partially present, they are not in any sense void of meaning. Wierzbicka (1986:612) even goes so far as to say that taking such a stance merely reflects "analytical failures". Apparently, the difference in opinion concerning whether discourse particles are semantically empty may partly be related to which discourse particles are examined.

2.3.2 Syntactic optionality

Discourse particles are in most studies seen as syntactically optional in the sense that the grammaticality of an utterance is not affected by the removal of discourse particles. Therefore, a text is still grammatical and intelligible if a discourse particle is deleted. This is related to the observation that discourse particles are often situated "outside the syntactic structure or loosely attached to it" (Brinton 1996:34). Since

[9] It may be worth noting that studies on individual particles may present properties of discourse particles in general based on the characteristics of the specific particles under examination. As commented by He and Lindsey (1998:151), "… one may raise concerns as to the nature of the phenomenon we are examining – is it the inherent properties of "you know" or properties of discourse particles as a class in general?"

discourse particles are “syntactically detachable” from a sentence (Schiffrin 1987:238), it follows that they have the “privilege of absence” (Fraser 1988:23), i.e. they have the status of being optional, which makes it difficult to analyse them syntactically (Stenström 1990). The view that discourse particles are grammatically optional does not, however, imply that their removal has no effect on the interpretation of the utterance. As pointed out by Fraser (1988), for example, the omission of discourse particles from an utterance eradicates a powerful clue about the relationship between the current and prior utterances which the speaker wants to express. Similarly, Schourup (1999) also states that this removes the signpost for signalling to the hearers how the utterance should be interpreted. Therefore, while discourse particles may well be syntactically optional, they play essential roles in performing pragmatic functions.

2.3.3 Indexicality

The pragmatic functions that discourse particles are said to fulfil are summarised under the notions of indexicality (Aijmer 2002) or connectivity (Schourup 1999), that is, to serve as deictic signals or cues for hearers’ interpretation. On the one hand, some studies are largely concerned with how discourse particles indicate the linkage between utterances or discourse units (see, for example, Lenk 1998; Schourup 1999). The primary function that discourse particles are seen to perform is thus as a connective, either relating the current utterance to its immediate linguistic co-text, often known as local coherence, or to segments further apart, often known as global coherence. As remarked by Schiffrin (1987:315), “markers index the location of an utterance within its emerging local contexts” and “contribute to the integration of discourse – to discourse coherence”. On the other hand, there are studies which not only focus upon the connective property of discourse particles in relating utterances but also the link between discourse particles and the attitudinal aspect of language. In this respect, discourse particles are considered devices which index the speaker’s orientation towards the hearer or the utterance. Examples include expressing stance such as certainty or uncertainty to the propositional content of an utterance or conveying affective meaning towards the participants. Following Halliday (1985), the above distinction could be broadly classified as textual and interpersonal uses respectively. In either case, the ‘indexing or guiding’ feature remains an important

functional property of discourse particles and will be revisited in Section 2.5 when the functions of discourse particles are extensively addressed.

2.3.4 Multi-functionality

Closely related to the above issue is the general acknowledgment that discourse particles are multi-functional. The concept of multi-functionality could be divided into two aspects: "categorially multi-functional" (Aijmer 2002:27) and pragmatically multi-functional.

Discourse particles are said to be categorially multi-functional because very often items referred to as discourse particles also operate in other grammatical categories such as adverbs and conjunctions. A differentiation between the discourse function and the grammatical function thus is common, though it is generally acknowledged that in some cases the differentiation may prove fuzzy and difficult (see Brinton 1996; Schourup 1999)[10]. The extent to which discourse particles are multi-categorial can be described in terms of what Stenström (1990:161) called their "D (item)-function ratio", which is defined as "discourse function in relation to grammatical function expressed in per cent". A discourse particle with a high D-item ratio signifies that it is more frequently used on the discoursal level while an item with a low D-value indicates that the item may be more associated with its grammatical use. In the London-Lund Corpus, for example, the discourse particle *oh* is found to have a 100% D-item ratio, meaning it is always analysed as serving discourse functions (Aijmer, 2002:27) while for *actually*, the D-item ratio is only 57.4% (Stenström 1990:162). Although Östman (1982:153) states that for an element to acquire the status of discourse particles it should have the discourse function as its sole function (e.g. as it is the case of *oh* in the London-Lund Corpus), it is apparent that many items previously studied would not fit into the category if the "uniqueness criterion" he suggested above is adopted.

Discourse particles are also considered to be pragmatically multi-functional when only their discourse function is concerned. In other words, not only is it possible for them to serve more than one pragmatic function, but it is also common for them to fulfil more than one function concurrently in a particular instance. In addition, a discourse particle may operate on different discourse levels at the same time.

[10] For more about the distinction between discourse use and literal use, see also Section 2.3.7.

According to Schiffrin (1987:328), for instance, a qualifying factor for an element to be a discourse marker is that it should work “at both local and global levels of discourse”. As pointed out by Aijmer (2002:39), following the distinction between textual and interpersonal uses of discourse particles in Hallidayan terms, “textual and interpersonal domains are not mutually exclusive”, that is, a discourse particle can simultaneously perform both the textual and interpersonal metafunction in the same discourse. Similarly, it is suggested that discourse particles could operate in several domains such as textual, attitudinal, cognitive and interactional (Jucker and Ziv 1998). Furthermore, discourse functions could be analysed in terms of dimensions such as the concept of face and politeness (Aijmer 2002). Such a complex display of multi-functionality may account for the fact that discourse particles are not randomly interchangeable, as elements functioning similarly at one level might differ at another (Schiffrin 1987).

In the event of multiple functions of a discourse particle in a single usage, however, there is no general agreement about whether there is a main function. While it may be the case that no single function necessarily predominates in a particular context (Schiffrin 1987), it is maintained elsewhere that the identification of a primary function, given sufficient contextual information, is possible (Holmes 1984). In addition, there is the suggestion that the primary and secondary function(s) of particles could be distinguished if one function dominates or is most frequent (Kroon 1995).

Such controversy extends to the related issue of the meaning of discourse particles. Given the multi-functionality of discourse particles, is it possible to determine a core meaning for them? When a number of functions are associated with a discourse particle, could the seemingly different senses of the item be subsumed under a single meaning or is the item homonymous? This problem has been a central topic for researchers working on discourse particles which results in two major opposing views. On the one side of this spectrum, some follow the monosemy approach and believe that it is possible to associate the phonological or orthographical form under study with “a single invariant meaning” (Fischer 2006:13). Central to this approach is the search for a “unitary word sense” for each discourse particle (Aijmer 2002:19), or put it simply, its core meaning. Fraser’s (1999:945) description of one of the basic features of discourse particles is a typical example of such a view: “every individual discourse marker has a specific, core meaning”. The

emphasis is thus on a single meaning which could fit the one form-one function formula (cf. Bolinger 1977). However, critics of the approach feel that the core meaning proposed is often too abstract or broad (Aijmer 2002) that it fails to capture and explain precisely how a discourse particle functions in a wide range of contexts and how it differs from other particles. In addition, the rather subjective process of identifying the core could lead to divergent and often conflicting outcomes (Schourup 1999). Given discourse particles exhibit a variety of functions in different contexts, it is difficult, if at all possible, to assign a core meaning (see, for example, Romero Trillo 2001). Instead, discourse particles should be seen as homonymous, with distinct senses working on different contexts in which the particle occurs. Such a description may turn out to be extremely complex as it has to account for the various functions operating on different levels (Aijmer 1996). A caveat for this stance, apparently, would be the possibility that there are so many different senses for a discourse particle that in the end it becomes an "elusive element lacking meaning" (Aijmer 2002:19). A balance between the two views therefore is proposed, in which discourse particles are said to have different senses which are related "either in a chain-like fashion through family resemblances, or as extensions from a prototype" (Hansen 1998:87). This approach, which assumes that the distinct readings of a discourse particle are generally related (Fischer 2006), is commonly known as polysemy. The subtlety of the relations between resemblances or extensions and the prototype remains, however, to be fully unveiled.

2.3.5 Non-truth-conditionality

Levinson (1983:87-8) was among one of the researchers who underlined the importance of discourse particles as a group and specifically pointed out that they have "at least a component of meaning that resists truth-conditional treatment". Non-truth conditionality, together with the lack of propositional meaning, is frequently considered to be the semantic property of discourse particles (see, for example, Jucker 1993; Schourup 1999). By 'non-truth-conditional' it basically means that the presence of discourse particles does not affect the truth conditions of the propositions in an utterance. However, the issue of whether discourse particles are non-truth-conditional is multifaceted and is not always straightforward. In Grice's (1989) influential account of discourse connectives, linguistic devices such as *but* and *so* (which are sometimes regarded as discourse particles) are seen as meaningful but

non-truth-conditional. Bach (1994) however disagrees with Grice's treatment of certain discourse connectives such as *but* and *therefore*. He argues that, like other 'content' words, they are indeed truth-conditional expressions which contribute to what is 'said'. Schourup (1999) also suggests that in some cases it may be possible to deny the truth value of the particle *so*, thus hinting that *so* could affect truth-conditions. In a study of the discourse particle *like*, it is found that the omission of the marker in some instances may lead to ungrammaticality and loss of propositional meaning, providing evidence that as a discourse particle the item is, in a few cases, truth-conditional (Andersen 1998). The situation is equally complex for sentence adverbials such as *frankly* and *seriously* which are generally considered non-truth-conditional. In this respect, there is the view that such linguistic expressions convey conceptual information which does not affect the truth-condition of the content but contributes to the higher-level explicatures of an utterance (Grice 1989). While members such as *well* and *oh* may appear to be less problematic in this regard, some other items in the class may present a taxonomic challenge for this characteristic of discourse particles.

2.3.6 Positioning

The observation that discourse particles often appear in initial position is generally recognised among researchers, though different units of talk have been considered in the discussion, including turn, utterance, tone unit, sentence and clause-element (see, for example, Jucker and Ziv 1998; Schiffrin 1982; Schourup 1999; Stenström 1990; Watts 1989). Schourup (1999:233), for example, argues that the position of what he calls 'discourse markers' should be understood "in relation to the central clause elements rather than to the position of the first word in an utterance" because of the clustering of these elements at the onset and other places of an utterance. Given the pervasiveness of discourse particles in spontaneous talk, however, it is not always easy to determine the central clause element of a sentence, which are notions more geared towards the written language than the spoken language. It should also be borne in mind that not a single definition of utterance, or any conversational unit for that matter, has been generally accepted. Indeed, the segmentation of spoken texts into units of talk has reached little consensus so far (Forsyth, Clarke & Lam 2008). Despite this controversy surrounding the concept of talk unit, the likelihood to occur in initial position is taken to be a syntactic property of discourse particles in many

studies (see, for example, Fung and Carter 2007; Schiffrin 1987; Schourup 1999). In a few cases, it is even considered to be a defining feature of discourse particles (Brinton 1996). Jucker and Ziv (1998), for instance, claim that discourse markers are restricted to sentence-initial position. Schiffrin (1982) also only focuses on utterance initial markers in her study on the grounds that utterance-internal or final particles do not show sequential dependence between utterances on a discourse level. In her study, an utterance is defined as "a prosodically closed unit of talk, completed by a terminal intonation contour and an optional pause" (Schiffrin 1982:38). Therefore, utterance initial instances as defined in her study may be counted as utterance medial in other models. Apart from the confusion arising from the definition of utterance and other speech units and hence the subsequent position of particles, it is also suggested that the tendency for discourse particles to occupy utterance initial position is a language-specific choice rather than a universal phenomenon. While such preference is evident in English, it may not be the case in other languages. In Chinese, one of the distinctive features of utterance particles as identified in many studies is their likelihood to occur in sentence-final position (Luke 1990). Further, as demonstrated in various studies, English discourse particles can also appear elsewhere, either in utterance internal or final position [11], though discourse particles appearing in utterance final position seem to be relatively infrequent. The relatively small number of discourse particles in utterance medial or final position could possibly be due to the ambiguity that arises from distinguishing the discourse function and the grammatical function in these positions (Fraser 1988). In contrast, the initial position which discourse particles often occupy is regarded as outside the core syntactic structure of an utterance and it therefore helps to dissolve such ambiguity. In addition, the initial position serves as a "pre-front field" which is loosely attached syntactically (Aijmer 2002:29). This syntactical looseness makes it a position especially tempting for interactional and textual planning (ibid.). Furthermore, the initial position allows the speaker to guide the hearer from the onset before "interpretation goes astray" (Schourup 1999:233). Therefore, there appears to be obvious reasons for discourse particles to affiliate with utterance initial position[12].

[11] As noted earlier, there are concerns surrounding the subject of what an "utterance" is and hence the obscurity of defining utterance-initial/internal/final positions (Fraser 1990:389).

[12] For a detailed explanation of the importance of utterance-initial position, see also Aijmer (2002:28-30).

Although it is generally agreed that discourse particles could appear in utterance initial as well as medial and final positions, the positional flexibility seems to apply only to some but not all members of the group. For instance, there are studies which show that items such as *however*, *actually* and *you know* enjoy a high level of freedom for their positioning (see, for example, Fraser 1988; Lenk 1998; Macaulay 2002). Elements such as *now* and *well*, on the other hand, appear to be more restricted in terms of their positions in an utterance. As observed in many studies (see, for example, Schourup 2001), the particle *well* is found to be absent in utterance final position. Given this variety of flexibility in positions, it is proposed that there are varying degrees of positional fixedness for discourse particles (Aijmer 2002).

2.3.7 Syntactic diversity

As noted in many previous studies, discourse particles are marginal forms which come from different traditional word classes. This "syntactic diversity" (Schiffrin 1987:64) could be reflected by the items which have been considered to be in the group: adverbs (e.g. *now*), verbs (e.g. *look*), interjections (e.g. *oh*), coordinate conjunctions (e.g. *and*), subordinate conjunctions (e.g. *so*), phrases (e.g. *the way I see it*) and also clauses (see Aijmer 1996; Brinton 1996; Fraser 1990; Schiffrin 1987). The fact that discourse particles do not originate from a single word class but are drawn from various grammatical categories is increasingly seen as an indication of grammaticalization that these items have been going through. According to Traugott (1995:32) who relates grammaticalization to particle studies, grammaticalization is the process "whereby lexical items or phrases come through frequent use in certain highly constrained local contexts to be reanalysed as having syntactic and morphological functions, and, once grammaticalized, continue to develop grammaticalized functions". Discourse particles are thus regarded by some as products of linguistic forms which have undergone diachronic and synchronic processes of grammaticalization (Aijmer 2002) and as a result develop in themselves a more grammatical status (Brinton 1996). In the gradual process of grammaticalization[13], items from a variety of traditional word classes evolve, in a unidirectional manner on the cline, as discourse particles. This involves a change of semantic meaning. More specifically, the meaning of these elements shifts from

[13] See Brinton (1996) for a detailed description of the various characteristics of grammaticalization.

propositional to textual and interpersonal, with an increasing orientation towards cohesiveness and subjectivity (Traugott and König 1991). Given the nature of such movement, some scholars argue that pragmaticalization is a more accurate description of the process undergone by particles as the change involves the emphasis of pragmatic significance in the meaning of these items. While pragmaticalization may not entail the change operating at the grammatical level, the process of grammaticalization described in various studies has included the component of pragmatic tendency (see, for example, Brinton 1996; Romero Trillo 2001; Traugott 1995).

Some researchers state that the shift of categorial status arising from grammaticalization also helps to explain the relationship between the literal use and discourse use of an item. As discourse particles are grammaticalized elements derived from their traditional word class 'equivalent', it is maintained that the pragmatic meaning of discourse particles is related to their original lexical/propositional meaning. More precisely, the discourse use of an item is a semantic extension of its propositional use, motivated by the speaker's desire to incorporate more subjectivity and emotion into the discourse (Brinton 1996). Therefore, discourse particles and their lexical counterparts are not homonyms sharing the same form but with different meanings. Through the historical process of grammaticalization, however, the original literal meaning may be altered which results in seemingly disparate meanings between discourse particles and their original sources (Schourup 1999). The loss in original meaning is often referred to as 'semantic bleaching' or 'desemanticization', which normally occurs at the last stages of grammaticalization (Brinton 1996). In such cases, it is argued that the original 'full' literal meaning of the item needs to be rendered in order to account for the common semantic core (Schourup 1999).

2.3.8 Orality

One stylistic property which is related to discourse particles but generally not considered to be a necessary criterion is their tendencies to occur in spoken discourse. Many studies have pointed out that discourse particles are largely an oral feature and their occurrences are due to the informality of speech and the shortage of planning time (Brinton 1996). Particles such as *by the way*, for instance, encode a component of on-the-spot planning which is applicable only to impromptu speech (Schourup

1999). Additionally, it is suggested that some of the meanings of discourse particles are normally associated with familiarity with the addressee (e.g. *you know*), which is not a typical feature of objectively addressed writing (Schourup 1999). While most items previously studied as discourse particles appear principally in spoken discourse (e.g. *oh*, *well* and *I mean*), there are no reasons to suggest that discourse particles are excluded in writing, though the elements used and the reasons behind the usage in the two channels might differ (Fraser 1990). *Moreover* and *nonetheless*, for example, are cited as formal discourse markers which are much more common in written English (Carter, Hughes & McCarthy 2000). Further, the prevalence of instant messaging on computers and short message service on mobile phones makes the boundary between written and spoken languages increasingly blurred. Therefore, it might be possible to say that there are spoken and written discourse particles, with the current research attention predominantly on those working in oral discourse.

2.4 Major approaches to discourse particles

With the increasing interest in discourse and the spoken language in the last few decades comes the growth in particle-related research. As discourse particles are a rather open class of elements which exhibit a wide range of characteristics and functions, it is not difficult to imagine that they trigger interest from different linguistic areas and languages. In fact, it has been suggested that the meanings of particles are so elusive that they could fit in almost every theory (Foolen 1997). Discourse particles have been described and analysed within a number of theoretical frameworks and models, most notably within relevance theory and coherence-based theory. They have also been studied in relation to speech act theory and politeness theory. In recent years, the incorporation of corpus evidence in the examination of discourse particles has also gained popularity. This section reviews these major approaches and findings in the field.

2.4.1 Relevance theory

Relevance theory has been influential in the description and analysis of discourse particles as it has provided an integrated explanation for the elements under discussion in a variety of particle studies. As outlined in Sperber and Wilson (1995), relevance theory is a general theory of communication based on cognitive principles. In view of Grice's (1975) Cooperative Principle which consists of the four Maxims

of Quantity, Quality, Relation and Manner, Sperber and Wilson (1995) propose that since there is only one maxim which cannot be violated "without an immediate breakdown of communication" (Lenk 1998:22) and that is the maxim of relevance, the four conversational maxims should be reformulated to one which governs human interaction: be relevant.

In relevance-theoretical terms, utterances are "inputs to inferential processes which affect the cognitive environment of the hearer" (Andersen 1998:150). Such inputs contain assumptions which are "treated by the individual as representations of the actual world" (Sperber and Wilson 1995:2). As interaction proceeds with the interchange of speaker and listener, new assumptions are added to the mutual cognitive environment. To interpret an utterance therefore is seen as identifying the assumption expressed in the current utterance as well as incorporating this assumption to the set of assumptions already processed. This involves working out the contextual effects of this new piece of information in relation to the context established. The concept of context in relevance theory is not confined only to the physical context or the linguistic co-text but also includes participants' beliefs and knowledge (Aijmer 1996). An utterance is said to be relevant if it achieves some kind of contextual effects (Sperber and Wilson 1995). In other words, the hearer will find the utterance relevant if s/he can correctly infer the speaker's meaning from the set of assumptions constantly negotiated in the course of the interaction.

Relevance theory is a cost-benefit model which weighs the processing costs of communication against its contextual effects. Under the principle of relevance which states that "every act of ostensive communication communicates the presumption of its own optimal relevance" (Sperber and Wilson 1995:158), speakers and hearers are expected to participate in interactions such that the greatest possible contextual effects are achieved from the smallest possible cognitive effort. To ensure an effective communication, the speaker's role is to maximize the contextual effects for the hearer to process while the hearer has to assess such effects from the set of existing assumptions with minimum processing costs.

Striving to reach a balance between the effects and the costs, participants in communication regularly look for devices which could make the process more cost-efficient. One way to achieve this is through the use of discourse particles. Within the framework of relevance theory, discourse particles are perceived to be signals which facilitate hearers' interpretation by reducing their processing effort. They are

seen as carrying procedural meaning as they provide processing instructions for listeners. Unlike linguistic elements such as nouns and verbs which convey conceptual meaning, discourse particles and other procedural linguistic items like pronouns and mood-indicators guide hearers in their comprehension of an utterance (Andersen 1998). More specifically, discourse particles impose constraints on the interpretation process so that hearers require less processing effort to arrive at the intended meaning of the utterance.

This description of how discourse particles work based on a relevance-theoretical approach appears to serve well in a number of studies. In a collective analysis of items under the label of 'discourse connectives', Blakemore (1987) proposes that elements like *so* and *after all* are employed to indicate how the relevance of one discourse unit depends on another. *After all*, for example, acts as a sign for hearers to regard the current utterance as a piece of evidence for some conclusion made earlier (Blakemore 1987). In terms of studies of individual discourse particles, Jucker (1993) focuses on the marker *well* and discusses how relevance theory could account for all the four distinct uses of the element.[14] In his view, *well* is a signpost which instructs the addressee to reconstruct the context for interpretation as "what seems to be the most relevant context is not appropriate" (Jucker 1993:438). Instead of understanding an utterance in its immediate context, hearers are advised to renegotiate the set of background assumptions. Jucker (1993:451) suggests that relevance theory is a "superior" framework for the analysis of discourse markers as he claims that it provides an explanation for "all the examples cited in the relevant literature". Schourup (2001) refutes this comment by giving counter-examples which do not seem to fit in Jucker's (1993) model. He criticizes Jucker's (ibid.) description as confining itself only to textual connections between consecutive utterances. Further, he adds that "relevance is, after all, not only a property of verbal stimuli, but also of unverbalized assumptions that enter into inferential processes" (Schourup 2001:1028). In this respect, Andersen's (1998) account of *like*, which is also built on relevance theory, appears to address the issue. In this study, *like* is not only seen as a textual link for relevance but it signals that "the utterance (that *like* is attached to) in some respect is a less-than-literal rendering of a speaker's thought" (Andersen

[14] The four uses are (1) as a marker of insufficiency; (2) as a face-threat mitigator; (3) as a frame; and (4) as a delay device (Jucker 1993:435).

1998:157). The item thus indicates a difference between the “internal representation of a concept and its external realisation which operates at a meta-communicative level” (Andersen 1998:160).

Although relevance theory has proved useful in a number of studies of discourse particles, it is suggested that relevance theory alone is not sufficient for explaining their use. As pointed out by Aijmer (1996), there are various factors which contribute to the production of discourse particles yet are not related to the subject of relevance. Examples of such factors include the consideration of politeness, the modification of illocutionary force and the management of speaker floor. Relevance theory also does not seem to fully elucidate why a certain particle is selected instead of the others and why it is not chosen in all possible situations where it could be used (ibid.). In addition, examples given in studies working within a relevance-theoretic approach are often contrived and many authentic instances cannot be explained in a satisfactory way under this single model. Therefore, the use of other approaches may be necessary for a comprehensive understanding of discourse particles.

2.4.2 Coherence-based theory

Another model which is commonly associated with the study of discourse particles is the notion of coherence. Coherence is generally regarded as an important attribute in communication but is extremely complex and difficult to define (Lee 1998). Although this concept has been extensively discussed in the linguistic community, there is still wide disagreement over the nature of coherence and how it is established in discourse (Lenk 1998). In the general sense of the word, ‘coherent’ is used to denote “parts that fit together well so that it is clear and easy to understand” (Cobuild 1987:265). In terms of discourse coherence, it concerns “how speakers and hearers jointly integrate forms, meanings, and actions to make overall sense out of what is said” (Schiffrin 1987:49). While the notion of coherence may not be fully understood thus far, it has been highlighted in many studies that coherence is not a text-inherent feature but it concerns participants’ interpretation in an actual context (Bublitz 1988). As it is naturally assumed that coherence is observed by conversationalists (Brown and Yule 1983), it follows that explicit linguistic cohesive devices might not always be necessary for achieving coherence.

In one of the pioneering studies of discourse markers, Schiffrin (1987:21) proposed a model of discourse coherence which consists of five “planes of talk”,

namely ideational structure, action structure, exchange structure, information state and participant framework[15]. For successful communication to take place, i.e. to achieve discourse coherence, the five components of the discourse model have to be integrated. According to Schiffrin (1987), discourse markers contribute to discourse coherence in two ways. First, they index utterances to prior and/or subsequent text and to the speaker and/or hearer. The marker *but*, for example, indexes the current utterance to the speaker as it "continues a speaker's action" and at the same time it indexes the current utterance to preceding discourse as it "returns a speaker to an earlier point of the text" (Schiffrin 1987:324). Second, they work as contextual coordinates which locate the current utterance on one or more of the five discourse components. *Oh*, for instance, is said to operate primarily on the information state plane and secondarily in the participation framework and the action structure (Schiffrin 1987). In Schiffrin's view, coherence is built upon relations between adjacent units of talk. Discourse markers facilitate the integration of the different discourse components by simultaneously proposing more than one contextual coordinates within which an utterance is interpreted (ibid.). To put it plainly, discourse markers serve as "a kind of discourse glue" (Fraser, 1990:385) which joins the different planes together, thereby making the discourse coherent.

Although Schiffrin's (1987) model has been regarded as influential and ground-breaking in the field, her approach is not without criticism. Her discourse model of five planes of talk, in particular, stimulates much objection. It is argued, for example, that the components of discourse coherence are not clearly defined and explained (see Kroon 1995; Redeker 1991). In addition, Schiffrin's (1987) framework is said to be applicable to only a selected set of expressions and it fails to differentiate between the markers (Redeker 1991). The absence of prosodic description and the fact that the findings are only based on interview data collected from a particular community in Philadelphia, moreover, make the study less comprehensive (Aijmer 2002).

---

[15] The following explanations for the five planes of talk are adapted from Fraser (1990:384): "Exchange Structure, which reflects the mechanics of the conversational interchange and shows the result of the participant turn-taking and how these alternations are related to each other;
Action Structure, which reflects the sequence of speech acts which occur within the discourse;
Ideational Structure, which reflects certain relationships between the ideas (propositions) found within the discourse;
Participation Framework, which reflects the ways in which the speakers and hearers relate to one another as well as orientation toward utterances; and
Information State, which reflects the ongoing organization and management of knowledge as it evolves over the course of the discourse."

In response to Schiffrin's framework, Redeker (1991) proposes a modified coherence-based model of discourse which retains only three of Schiffrin's (1987) five planes of talk. The two planes (participation framework and information structure) are abandoned because they are only loosely associated with coherence between utterances (Redeker 1991). In her study, the term 'discourse operator' is used to represent "linguistic signals of textual coherence links" (Redeker 1991:1139). Again, this is similar to Schiffrin's (1987) and other coherence-based definitions in which discourse particles are said to indicate coherence relations(s) either between adjacent units or between a discourse unit and some aspects of the communicative context (Risselada and Spooren 1998).

Largely following Redeker's (1991) definition, Lenk (1995; 1998) examines how discourse particles facilitate the establishment of coherence in spoken discourse. In particular, she is mainly interested in how global coherence is achieved through the use of certain discourse particles. In contrast to those markers which function on a local level to signal relationships between immediately adjacent utterances, globally-oriented markers are concerned with segments which are further apart. The particle *anyway*, for example, is suggested to have a major function of closing a conversational topic (Lenk 1995). By orienting participants among the topics which are being discussed, the word *anyway* facilitates the overall coherence of the entire discourse (ibid.).

In sum, within coherence-based theory, discourse particles are seen as cohesive devices which aid the achievement of coherence and the organization of discourse. They are employed when speakers intend to make explicit the coherence relations in discourse, be it a relation between adjacent utterances, between segments which are distantly related, or between utterances and speakers' attitudes or intentions. Similar to the relevance-theoretical approach, a coherence-based account seeks to explain how discourse particles operate under a single model. While a solution which fits all has its obvious appeal, it might give less emphasis to the effect of various linguistic and contextual factors in the interpretation of what discourse particles actually do in a large number of authentic examples.

2.4.3 Speech act theory

Many of the expressions now known as discourse particles are studied in great detail within speech act theory. They are of special interest to speech act theorists because

they are meaningful but non-truth conditional. As discussed in Section 2.3.5, discourse particles are non-truth conditional in the sense that they do not contribute to the truth or falsity of the utterance in which they appear. They are meaningful because instead of contributing to the propositional content of the utterance, discourse particles indicate "the force or intention behind the words" (Holmes 1995:49), i.e. the illocutionary force of an utterance. They are considered "illocutionary force indicating devices" which modify the pragmatic force specified by the mood of the sentence (Aijmer 2002:8). More precisely, they are signals which indicate the type of speech act performed by the utterance or the speaker's stance towards the proposition contained in the utterance (Rouchota 1998).

Discourse connectives, which are considered to be examples of discourse particles in many studies, are studied closely within speech act theory. Elements such as *but* and *so*, according to Grice (1989:362), indicate "higher-order speech-acts" and convey a conventional implicature. When compared with conversational implicatures which have to be inferred by the hearer based on a set of pragmatic principles, conventional implicatures make use of the conventional meaning encoded in a group of linguistic expressions to give strong indications as regards what is implicated. While segments which perform lower-order speech acts contain the truth-conditional content of an utterance, linguistic devices such as discourse connectives encode an implicit proposition. Specifically, they comment in a certain way how these lower-order speech acts should be interpreted. Thus, for example, expressions such as *but* and *on the other hand* indicate the higher speech act of contrasting; elements like *so* and *therefore* signal the speech act of explaining. These higher-order speech acts are non-central, as their existence depends on the performance of central speech acts such as saying and telling. Since implicatures only suggest an implied relationship between two propositions, they can be cancelled without altering the meaning of the original utterance.

Apart from indexing what kind of speech act is performed in a statement, discourse particles also signal the attitudes of speakers towards the proposition expressed. One of the attitudes which could be conveyed by the use of discourse particles is the level of commitment towards a proposition. In such cases, discourse particles are employed as hedges on illocutionary force when the speaker is aware that s/he fails to fulfil a maxim (Aijmer 2002). The discourse particle *I think*, for example, indicates the non-observance of the quality maxim and implies that the

speaker “is not taking full responsibility for the truth of his utterance” (Brown and Levinson 1987:164). Similarly, the word *like* is often used to signal what is said is not an exact rendition of a speaker’s thought (Andersen 1998), thus leading to the non-observance of the maxim of manner.

Although speech act theory has been a crucial framework in the analysis of discourse particles, it is suggested that it does not provide a comprehensive account of these elements. The limitation of the model lies in its failure to take into consideration the sequential context of utterances (Aijmer 2002), which is fundamental for discourse particles to be understood. In other words, the textual functional dimension of discourse particles beyond the limits of a single utterance is lacking in a description based on speech act theory. The account of conventional implicatures also seems to imply that the higher-order speech acts performed by discourse connectives are invariable regardless of contextual situations, which appears to over-simplify the multi-functional nature of these linguistic items.

2.4.4 Theory of politeness

Another aspect in which discourse particles are examined as indicators of illocutionary force is the theory of politeness. As a pragmatic phenomenon, politeness is concerned with the strategies used by speakers to maintain a harmonious interaction (Holmes 1995). Central to politeness is the concept of face. According to Brown and Levinson (1987:62), face is “the public self-image that every member wants to claim for himself” and can be subcategorised into positive and negative. Through exchange with other people, face could be lost, maintained or enhanced. Face-threatening acts (FTAs) are speech acts which could damage another person’s face. In order to minimize the face-threatening effect of those speech acts and maintain rapport between participants, one needs to attend to the hearer’s face in interaction.

Within the theory of politeness, hedges are seen as face-saving devices which aid in the mitigation of face damage. As discourse particles with a hedging function such as *I guess* and *I think* reduce speakers’ commitment to the proposition expressed, they are employed as a protective mechanism when speakers are performing face-threatening acts such as criticisms (Aijmer 2002). This indicates the speakers’ wish to distance themselves from such acts and makes the allegation or the opposition less assertive. This thus helps to alleviate the face damage of others. In other words, these

discourse particles are linguistic realizations of politeness strategies which attend to the negative face of listeners, i.e. the desire not to be impeded or imposed upon (Brown and Levinson 1987). Alternatively, discourse particles could also be employed to attend to the positive face of participants. Positive face is reflected in an individual's desire to be liked and respected. In a study of the use of *eh* in New Zealand English, for example, it is found that the word is employed overwhelmingly by speakers with the purpose of establishing common ground with others (Meyerhoff 1992). In another study, elements such as *you know* and *you see* are also suggested to enhance the interlocutor's positive face by trying to "involve the listener in the thinking process of the speech" (Romero Trillo 2002:777). As shown from the above findings, certain discourse particles could be taken as face enhancing or mitigating devices to promote the speaker's positive face or to alleviate the negative face damage of the hearer.

2.4.5 Corpus-linguistic approach

The rapid progress in computer technology in the last few decades has led to substantial growth in corpus linguistics. The advance in information technology and the availability of powerful machines have facilitated a steady increase in the compilation and use of computer corpora. The ease of accessing and processing a vast electronic collection of texts presents new research opportunities for scholars interested in the study of discourse particles, as these elements could now be closely scrutinised in a representative sample of texts with the help of concordancers and other computer software. In recent years, an ever-increasing number of studies of discourse particles have based their findings on computer corpora, making use of all the advantages offered by this approach to carry out analyses which are qualitative and at the same time quantitative.

Lenk's (1998) coherence-based study is a full-length research which makes use of the London-Lund Corpus, a text collection of roughly half a million words including conversation, interviews and speeches. To allow for a comparison between American and British English, a part of the Santa Barbara Corpus of Spoken American English is included in the analysis. Primarily working within the concept of coherence, Lenk (1998) illustrates, though mainly qualitatively, how corpus data could be employed to describe the role discourse markers play in marking discourse coherence. Her study also reveals the similarities and differences of usage of

discourse particles in the two major varieties of English. As reflected from the statistics, the item *anyway*, for instance, is found to be more frequently used as a discourse particle in American English than in British English (ibid.).

Aijmer's (2002) study on discourse particles is another large-scale research work using the London-Lund Corpus. Unlike Lenk's (1998) monograph which draws data from the corpus to support her examination of markers within a coherence-based model, Aijmer's (2002) in-depth analysis is guided by corpus data instead of any pre-existing framework. Drawing on a variety of discourse-analytic techniques such as the notions of move, act and exchange (Sinclair and Coulthard 1975), Aijmer provides a comprehensive functional account of the distributional patterns of discourse particles in a range of text types. One of the characteristics of her research which is largely absent in many studies is the attention to prosodic and collocational features of discourse particles. With the carefully transcribed and tagged London-Lund Corpus, the study shows that prosody serves as a cue for identifying the discourse use of certain items. The element *just*, for example, is commonly employed as a discourse particle when it occurs with a "step-up in pitch" (Aijmer 2002:162). She also explores how the prosodic properties of discourse particles are related to their different discourse uses. The use of *sort of* followed by a pause, for instance, is suggested to be associated with hesitation, difficulties in planning and with floor-holding purposes (Aijmer 2002). Similarly, the collocations of discourse particles are useful in the interpretation of their functions. When *sort of* is used in the corpus, it is found that the expression often co-occurs with elements which have a down-toning effect such as *pretty* and *probably* (ibid.). She therefore suggests that *sort of* acts as a device for modifying illocutionary force. In the analysis of utterance final tags such as *and all that sort of thing*, she proposes that they collocate with expressions like *you know* to coordinate participants' viewpoints and establish common ground. By means of an annotated corpus and the aid of computer tools, Aijmer (2002) demonstrates how a corpus could be fully utilized and the potential of computer technology completely exploited to enhance our understanding of aspects of discourse particles which are often ignored.

More recently, Müller's (2005) comparison of the use of discourse markers in native and non-native English discourse also adopts a corpus-linguistic approach. In this full-size analysis which makes use of data from the Giessen-Long Beach Chaplin Corpus, Müller compares the use of four selected discourse markers by German EFL

speakers and American native speakers in a movie retelling experiment. Findings from the study show that not only do German speakers and American English speakers differ in terms of frequency of use of discourse markers, discrepancies are also observed in the ways the four markers are used. Undoubtedly, these interesting findings provide valuable resources for inter-language research and intercultural communication.

Although the study of discourse particles with computer corpora has just begun to blossom, it can be seen that the use of corpus data could greatly enhance the quality and efficiency of particle research. As remarked by Aijmer (2002:3), "corpora represent actual performance and provide the opportunity to study the distribution and function of particles in extensive text extracts representing different registers". As a result, they are especially powerful in giving an authentic account of how a discourse particle is used in real-life situations.

2.5 Functions of discourse particles[16]

Following the previous discussions of indexicality and multi-functionality as properties of discourse particles, this section examines the major functions of discourse particles reported in the literature. Given the multi-functional nature of these elements and the various approaches used in different studies, an investigation into their functions often reveals an intricate picture with different focuses and taxonomies (Brinton 1996). A generally agreed functional typology of discourse particles is missing thus far. However, as mentioned in Section 2.3.3, discourse particles could be treated as operating on two main functional domains in general, namely textual and interpersonal. The third meta-function "ideational" which concerns what is being talked about (Halliday and Hasan 1976) is associated with the propositional content of language and thus is generally considered less relevant to the analysis of discourse particles. While there are particles which primarily serve functions in one domain, there are also particles which work on both, either simultaneously in one occurrence or in different contexts.

When discourse particles play a textual role, they are "text-structuring devices" (Jucker and Ziv 1998:4) which principally index linkages between discourse units either on a local or a global level. On a global level, particles signal relationships

[16] For a detailed list of functions, see also Brinton (1996).

between larger segments of talk. This includes initiating or closing talk, marking boundaries in discourse, signalling new and old information and restricting the relevance of neighbouring utterances (Brinton 1996). The expression *you know*, for instance, is found to signal the beginning of a turn when spoken with a falling tone in an analysis of the London-Lund Corpus (Romero Trillo 2001). For closing discourse, the marker *anyway* is reported to have the core function of indicating the end of a topic within the context of conversation (Lenk 1995). Discourse particles also fulfil a crucial function in boundary marking such as indexing topic shifts or digressions. In a study of classroom discourse by Sinclair and Coulthard (1975), items such as *now* and *well* are classified as "frames" which are employed to indicate the boundaries of transactions. In such uses, it is observed that the literal meaning of the items is suppressed and the items exhibit characteristic prosodic features (Coulthard 1977). Similar to Sinclair and Coulthard (1975), Aijmer (2002:41) studies the global coherence function using the term "frame function". In her schema, discourse particles not only fulfil the various global functions as discussed above, but they also introduce explanations, background information or direct speech (Aijmer 2002). Marking self-correction, which is a local feature in Brinton's (1996) view, is regarded as a global function in Aijmer (2002). In sum, discourse particles are employed to achieve global coherence when the speaker senses a need to signal the transition in interaction, be it a switch in topics, information content or modes of speech.

On a local level, discourse particles mark the relations between adjoining units of talk. These neighbouring units could be segments within a single utterance or adjacent utterances produced by different speakers. Discourse particles could indicate how the current message is in parallel to or in contrast with the immediately preceding context (Fraser 1988). The word *and*, for instance, is considered an example of parallel markers which serves as a coordinate conjunction. On the other hand, the word *but* is a contrastive marker which functions locally by indicating that the upcoming message is in some way contradictory to the expectations set up in the previous utterances (ibid.). Another local function that discourse particles are suggested to perform is listing. According to Aijmer (2002), discourse particles can be employed to indicate how adjacent discourse units are to be interpreted sequentially in a list of activities. In her study of *now*, for instance, it is reported that

the word was frequently employed to “show the order of events” in demonstrations such as the sequence in cooking (Aijmer 2002:83).

Within the interpersonal domain, discourse particles serve both interactive and attitudinal purposes (Östman 1982). On the interactive level, discourse particles are important resources for ensuring the smooth flow of communication. Speakers who encounter difficulties in their planning process, for example, may resort to these elements so that their contribution will not come to a complete halt. When speakers are searching for the right word or structure, they may employ expressions such as *I mean* and *you know* for time stalling purposes so that their processing phase can be lengthened (Erman 2001). Discourse particles also facilitate interactions by acting as turn-taking or floor-holding devices (Aijmer 2002). They may be used when the interlocutor intends to grasp the conversational floor from others or to hold the ground to speak when interrupted. A third function on the interactive side is securing comprehension (Erman 2001). For speakers, they may make use of discourse particles such as *okay* and *right* to ensure that they have been understood correctly. On the part of listeners, they could employ back-channels and reception markers such as *yeah* to indicate their reception of the speaker’s output (Brinton 1996; Jucker and Smith 1998) or use the particle *oh* as an “information receipt” (Heritage 1984:307). The last function which concerns the operative aspect of interaction is raising participants’ attention. The items *look* and *listen*, for example, are attention-seeking elements which highlight the forthcoming message (Romero Trillo 2002).

Another crucial area within the interpersonal mode in which discourse particles are analysed is attitudinal. The attitudinal aspect of discourse particles is often examined within speech act theory and politeness theory, as shown in Section 2.4.3 and Section 2.4.4. Discourse particles provide a rich resource pool for speakers to express their attitudes towards the subject matter and the other participants. Adverbials such as *frankly* and *amazingly* which are called “commentary pragmatic markers” by Fraser (1990:386), for example, signal the speaker’s position on the message uttered. As noted in Section 2.4.4, discourse particles such as *I guess* and *I think* indicate speakers’ level of commitment on the propositional content of the utterance and thus they express speakers’ tentativeness on the topic. In addition, discourse particles could upgrade or downgrade the effect of a claim (Smith and Jucker 2000). Expressions such as *just* and *sort of*, for example, are used as softeners to weaken the illocutionary force of the message (Aijmer 2002). The approximation

meaning associated with these expressions makes the proposed message less obtrusive and face-threatening, thus implying the speaker's desire to maintain a harmonious relationship with the other interlocutor(s). When speakers desire to establish intimacy with others, they employ face-saving devices such as down-toners like *only* and addressee-oriented markers like *you see* and *you know* to enhance involvement and make the interaction less intimidating (Jucker and Smith 1998). Conversely, the marker *in fact* is used to introduce a stronger claim and amplify the pragmatic force behind (Smith and Jucker 2000). Thus the use of such intensifiers might suggest an emphasis on the authority of speakers (Erman 2001). In a study of children's use of discourse particles, for instance, it is found that discourse particles are employed to mark social status roles (Andersen et al. 1999). In sum, discourse particles are extremely helpful in indicating attitudes and evaluations, both in the negotiation of viewpoints and social relationships with others.

2.6 Prosodic patterns of discourse particles

It has long been claimed that the prosodic features of discourse particles provide important clues for interpretation (see, for example, Erman 1987; Schiffrin 1987). However, in comparison with the large number of particle studies, very few empirical systematic analyses of the role of intonation in the use of discourse particles have been carried out, possibly owing to the lack of prosodic information available. Although in the last decade there has been an increasing attempt to study discourse particles in relation to their prosodic features, the quantity of studies on the prosody of discourse particles is still meagre compared with particle studies in other areas. As remarked by Wichmann (2001), progress on prosodic studies has always been slower when compared with lexical and syntactic analyses in the field of corpus linguistics. The prosodic features of discourse particles are no exception. However, with the increasing accessibility of prosodically transcribed corpora in recent years, the prosodic properties of discourse particles have slowly been unveiled.

In the literature, the prosodic phrasing of discourse particles has been suggested to be a useful clue for distinguishing discourse functions from grammatical functions. Prosodic phrasing refers to whether or not the item in question is in a separate tone unit on its own (Aijmer 2002). While grammatical uses of a form are more likely to be integrated into the tone unit, discourse uses often constitute independent tone units (Jucker and Ziv 1998). At times, they are detached from the main segment by

"comma intonation" (Schourup 1999:233), that is, there is a break which prosodically separates the particle and the host utterance. This intonational autonomy often acts as a clue for identifying the discourse use of certain linguistic items[17]. Notice, however, that there are discourse particles which are intonationally integrated with the main clause, especially in utterance initial position (Schourup 1999). Therefore, phonological independence is suggested not to be a major defining quality of discourse particles. In such cases where there is an intonational integration, other prosodic features could be used to distinguish discourse uses from grammatical uses, as demonstrated by Aijmer (2002) in which the discourse uses of *now* are identified by the absence of tone. Stenström (1994) also suggests that tonicity and pitch direction are possible factors influencing whether a particular instance serves discourse functions.

Apart from being prosodically independent, discourse particles are often said to be phonologically reduced or unstressed (Brinton 1996), though empirical evidence based on a large number of instances is rare. Phonological reduction is also claimed to have functional correlations. Bolinger (1989:333), for example, discusses the difference in pragmatic functions between "full" *well* and phonologically reduced "*w'l*". Prosodic analyses also focus on the typical tones of individual particles. Svartvik's (1980) article on *well*, for example, suggests an association between the use of level tone with the discourse use of the word. In an analysis of *you know* and *you see*, Stenström (1984) observes from quantitative data that both expressions are more frequently spoken with rising tone. Moreover, there are studies which attempt to associate intonation contour with specific functions of discourse particles. In an interesting article describing a mathematical model for the analysis of variation in discourse particles, Romero Trillo (2001) shows, by mathematical means, how the various functions of discourse particles such as *look* and *I mean* are realised by the use of different tones. Aijmer's (2002) study on *actually* indicates that there may be correlations between prosody and discourse functions in different syntactic positions. While the fall-rise tone in initial position is associated with the goal of marking

[17] See, for example, Aijmer's (1988) study on *now*, Ferrara's (1997) study on *actually* and Macaulay's (2002) study on *you know*.

contrast, the "compound fall-plus-rise tone"[18] in final position appears to fulfil a social function (Aijmer 2002:263).

Another area which deserves attention in the prosodic analysis of discourse particles is their co-occurrences with pauses. As pointed out by Stenström (1990) based on a selection of monologues and dialogues from the London-Lund Corpus, lexical items such as *well* and *I mean* are frequently accompanied by either silent or filled pauses, or a combination of both. In particular, the words *anyway* and *now* have the strongest association with pauses proportionally. In most cases, they occur after the pause. This coincides with Svartvik's (1980) observation regarding the location of pauses in the neighbourhood of *well*. On the other hand, Jefferson (1983) explores the use of a number of items including *well*, *but*, *so* and *uh* which she calls 'conjunctionals' before pauses in serving interactional purposes. She points out that the use of such conjunctionals in combination with pauses can either signal floor-holding or turn-yielding. Following Jefferson's (1983) observations, Local and Kelly (1986) propose that whether such patterns are for keeping or giving away the turn depend on their phonetic realisations. In other words, their findings suggest a correlation between different discourse functions of these items and the kinds of pauses they associate with.

The limited but fruitful research outputs in the area show that the prosodic properties of discourse particles play an important role in distinguishing the discourse and grammatical uses and determining the discourse functions of these elements. The study of prosodic features should therefore be given more emphasis to enhance our understanding of discourse particles.

2.7 Discourse particles and language learning

Given the fact that discourse particles have received enormous attention over the years, it is indeed a surprise that little research effort has been devoted to this area in relation to language learning. This is reflected in the small number of particle studies on non-native speakers. Similarly, the area of discourse particles has been shunned by researchers working in second language acquisition (Müller 2004); an outcome that possibly arises from the notorious reputation discourse particles have as

[18] This refers to a tonal pattern which consists of "a sequence of a fall tone followed by a rising tone on actually" (Aijmer 2002:263).

discussed in Chapter 1. Research work which deals with discourse particles together with language learning issues, in addition, appears to largely focus on the implicational side. Despite the fact that relatively limited research has been undertaken in this field, there are studies which provide valuable insights into the use of discourse particles by language learners and its relation to pedagogy.

In a study comparing the use of discourse particles by native and non-native speakers of English, Romero Trillo (2002) investigates the development of certain pragmatic markers in the speech of Spanish speakers of English. By examining children and adult data for both groups, he shows that non-native adults' usage of discourse particles was even more limited than that of native children (Romero Trillo 2002). Although the Spanish adults in his study were English philology students who achieved a satisfactory grammatical level in English, he reported that there was a shortage of discourse particles in their speech, which might hinder them from effective communication (ibid.). More precisely, the limit in quantity and diversity of discourse particles used by these non-native speakers could possibly make their speech sound blunt and impolite. In the light of these findings, he called for the teaching of discourse particles in the curriculum in Spain so that learners could attain pragmatic competence (ibid.). He further proposed that research studies which describe the use of discourse particles following mathematical and statistical approaches could be incorporated in the teaching process (ibid.).

The use of discourse particles in Xhosa English, a sub-variety of Black African English, is the focus of a study by de Klerk (2005). Differences found in the use of *well* by Xhosa speakers of English and native speakers highlight the fact that discourse particles are largely overlooked in the educational system. With limited exposure to the English language, it is argued that teaching materials should give more emphasis to discourse particles.

Müller's (2004) comparison of the use of *well* by German EFL speakers with American native speakers is also associated with pedagogy. The study shows that non-native speaker data exhibit fewer functions of *well* when compared with native speakers. In addition, German speakers of English seem to have a preference for *well* over another particle, *so*, particularly when they are used to fulfil similar functions (Müller 2004). According to Müller, this is presumably due to the influence of German textbooks in English. In a brief analysis of the items *well* and *so* in German textbooks, the author reports that the ratio of occurrences of the two elements is not

consistent with the statistics shown in the native corpora. As textbooks are central to English teaching in Germany, speakers may be influenced to use *well* instead of *so* (ibid.).

Although there is a prevalent view that discourse particles are a form of linguistic stigma and therefore are not a commendable feature for language learning, it has been increasingly acknowledged that discourse particles are crucial for learners to communicate successfully at the pragmatic level of interaction. As pointed out by Wierzbicka (1991), the failure of language learners to grasp the meanings and functions of discourse particles will severely hamper their communicative competence. More importantly, if language learners are denied access to these critical pragmatic devices in their learning process, they may not be able to fully project their personality in the target language. As a result, although they may well be capable of attaining transactional goals, they could only operate in the target language with a "reduced" personality (McCarthy 1998:112). In other words, learners are deprived of the right to behave and express themselves in the same way as they do in their native language. The image that they could present in the second or foreign language is, at most, a partial alter ego. At worst, the dearth of discourse particles in their talk could leave them "potentially disempowered and at risk of becoming a second-class participant" (O'Keeffe et al. 2007:39). When the issue of critical language awareness has become part of the agenda for language learning, it is perhaps also appropriate to review the place of discourse particles in language learning.

2.8 Discourse particles in grammar books and dictionaries

Closely connected to the reporting of particle studies in relation to language learning, this section surveys how discourse particles are described in grammar books and dictionaries, which arguably play a key role in English language learning for students who predominantly learn English in the classroom. This section is sub-divided into three parts, which examines largely in a chronological order how reference materials for the English language deal with the group of lexical items under investigation in the present study. From traditional grammar through corpus-based grammar to the ground-breaking Linear Unit Grammar (Sinclair and Mauranen 2006), the evolution of the treatment of discourse particles in the course of time can be seen from the review below.

2.8.1 Traditional grammars and dictionaries

At the time when the term ‘electronic corpus’ was esoteric in the field of linguistics and the collection of sound recordings was extremely difficult, English grammars and dictionaries mostly focused on the written language. In particular, they were mainly concerned with the grammatical structure at the sentential level. The emphasis on written English was evidenced by taking the sentence as the centre of attention and by breaking down the sentence into a number of major components such as nouns, verbs and adjectives. It thus comes as no surprise that discourse particles, as a feature characteristic of the spoken language, are only given a cursory mention in most of these reference books. More importantly, even if these lexical items are discussed, they are more likely to be subsumed under other traditional word classes, mostly notably adverbs. In *A Grammar of Contemporary English* (Quirk, Greenbaum, Leech & Svartvik 1972:274), for example, a brief description on the discourse particles *well*, *oh* and *ah* as “reaction signals and initiators” is provided under the chapter on adjectives and adverbs. In *A Comprehensive Grammar of the English Language* (Quirk, Greenbaum, Leech & Svartvik 1985), again explanations related to discourse particles which are scattered across different grammatical categories remain limited and patchy. In these grammars, the term ‘discourse particle’ or ‘discourse marker’ is out of sight and collective accounts of discourse particles are absent. When lexical items such as *well* and *so* are individually discussed, the concentration is chiefly on their grammatical functions. The grammatical uses of these two words, including as an intensifier or as a pro-form, are extensively described, whereas their discourse uses are given far less attention. The same applies to other discourse particles which are categorially multi-functional. This disproportionate emphasis on grammatical functions seems to suggest that the discourse uses of these words are only minor, if not subordinate, which does not seem to be justified given the actual frequency of the discourse uses of these elements in authentic speech.

Similarly, in dictionaries where entries are arranged in alphabetical order, the grammatical functions of words are often given priority. Not only are grammatical functions always placed in front of discourse functions, the explanations for grammatical uses are often more detailed. Discourse uses, if they are ever included, are only presented very briefly (Luke 1990; Stenström 2002). This could be

attributed to the fact that most examples in these dictionaries are obtained from the written language. In *Practical English Usage* (Swan 1996), for example, the entry for the word *well* only contains two explanations. The first one is that *well* and *good* are semantically similar and the second one is the use of *well* meaning "in good health" (Swan 1996:614). Considering the fact that it is a dictionary which specifically deals with problems encountered by foreign language learners and their teachers, the information it contains regarding one of the most frequently occurring words in spoken English appears to be surprisingly inadequate.

2.8.2 Corpus-based grammars and dictionaries

The advance in technology over the last few decades has brought about dramatic changes in the ways in which language can be analysed. The accessibility of audio records in combination with the ease of processing computerised texts has opened up new opportunities in the study of spoken English, allowing researchers to get hold of a large quantity of speech data for analysis.

Since corpus work became a reality and gained growing recognition in the study of language, the orientation of grammar books and dictionaries has gone through a gradual shift from being wholly biased towards the written language to giving more rightful but long overdue consideration to the spoken language. In most corpus-based grammars today, a considerable proportion is devoted to the description of spoken English. The concept of a dictionary entry is redefined. Instead of specifying the prescriptive rules on the correct usage of a word, descriptive observations of how a word is actually used in authentic examples are given. As a result of the availability of spoken data, features which are typical in speech have been more thoroughly described than ever. Discourse particles are no exceptions.

In the first entirely corpus-based grammar of English *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan 1999), discourse markers are discussed collectively in various sections of the grammar of conversation, in addition to a separate section contributing to their use in detail. Some suggested lexical items in the category include single word inserts such as *well* and *now*, formulaic clausal forms such as *you know* and *I mean*, and some other multi-word constructions such as *good grief*, though it is noted that members in the category are subject to debate. Discourse markers are described as "words and expressions which are loosely attached to the clause and facilitate the ongoing

interaction" (Biber et al. 1999:140). According to the Longman grammar, their functions are both pragmatic and discoursal, which involve the expression of emotions and attitudes and the marking of links or transitions. By comparing the frequency distribution of certain linguistic elements such as *now* and *you see* across registers and dialects, Biber et al. (1999) show the contextual and regional variations in the use of some common lexical items which are primarily used as discourse particles.

The fully revised and redesigned third edition of *A Communicative Grammar of English* (Leech and Svartvik 2002) also incorporates corpus evidence into its descriptions. Partly based on *A Comprehensive Grammar of the English Language* (Quirk et al. 1985), it has an expanded section on spoken English when compared with older versions. Although its explanation of discourse particles is not as detailed as the one in the Longman grammar, the importance of these expressions especially to foreign learners is highlighted. In Leech and Svartvik's (2002:14) view, discourse particles reveal "speakers' attitude to their audience and to what they are saying" and serve interactive functions such as achieving cooperation and turn keeping in discourse.

The Cambridge Grammar of English (Carter and McCarthy 2006) is yet another fruitful result of corpus research. Mainly based on data from the CANCODE corpus, it gives equal importance to the grammar in spoken and written English. Like Biber et al. (1999), Carter and McCarthy (2006) treat discourse markers as a lexical category made up of different linguistic constructions, including words, phrases and clauses. Their inclusion of phrases such as *at the end of the day* seems to suggest a broader class of discourse markers than customary. In terms of discourse functions, the Cambridge grammar proposes that "they organise and monitor an ongoing discourse" and "mark the state of knowledge between participants" (Carter and McCarthy 2006:901). This seems to suggest that the functions of discourse markers are basically textual and epistemic.

As a pioneering corpus-based project in lexicography, the *Collins Cobuild English Language Dictionary* (1987) was the first dictionary to use authentic and representative examples from a corpus. By making use of frequency figures from *The Bank of English*, the dictionary provides users with indicators of how frequently the search word is used. This also makes sure that uses which are most common in the spoken language are fully described and given precedence over other uses. The word

well, for example, has six entries in the dictionary, with the first being its discourse uses. Thirteen different discourse uses are identified with authentic examples from the corpus given. This makes it clear to language learners that *well* is primarily used as a discourse particle and is mainly used in the spoken language, making a noticeable contrast to its treatment in dictionaries solely based on the written language.

2.8.3 Linear Unit Grammar (LUG)

An innovative descriptive model representing texts in a linear succession of units, Linear Unit Grammar offers a complement to conventional grammar by revealing meaning-bearing patterns in texts which could be subsequently analysed in traditional grammatical frameworks (Sinclair and Mauranen 2006). Being one step ahead of corpus-based grammar, Linear Unit Grammar incorporates the description of the written and spoken language within a single grammar, allowing for a thorough examination of a variety of texts and linguistic patterns which are mostly ignored in current grammatical terms.

In Linear Unit Grammar, it is proposed that there are two basic types of elements in discourse: “those which are concerned with that which is being talked about, and those which are primary concerned with managing the discourse” (Sinclair and Mauranen 2006:59). The former is called a message-oriented element while the latter is known as an organisation-oriented element. Given the fact that discourse particles generally contribute little to the incrementation of topics, they are normally categorised as organisation-oriented elements within this framework. The organisation units are further sub-divided into two kinds, depending on whether they are text-oriented or interactive-oriented, though overlap could exist between the two. In their analysis, discourse particles are thus mostly involved in the management of discourse by either indicating interrelations between segments of talk, expressing attitudinal stance of speakers or regulating turn taking and yielding activities. Under the holistic approach to language in Linear Unit Grammar, discourse particles are no longer treated as peripheral items which are mostly ignored in conventional grammars and dictionaries because they are hard to fit within a traditional grammatical framework. Instead, they are given due attention based on their functions in ensuring the success and efficiency of the increment of topics.

2.9 The present study

From the above discussions, it is obvious that there are still gaps in the study of discourse particles which remain unfilled. Apart from the unsolved controversies arising from terminology, definition and classification as discussed in Section 2.2, there are other issues surrounding discourse particles which have yet to be surveyed. Their defining properties continue to be debatable (Section 2.3) and some approaches to discourse particles seem to focus on a specific aspect while overlooking others (Section 2.4). The many discourse functions which could be achieved by discourse particles in a variety of settings, as outlined in Section 2.5, are still waiting to be completely unveiled. As described earlier in Section 2.6, the prosodic properties of discourse particles are far from being fully understood. Analyses on the use of discourse particles by non-native speakers and their relation to language teaching and learning, as evidenced by discussions in Section 2.7, appear to be minimal. Thorough examinations of how discourse particles are treated in pedagogical materials such as reference grammars and dictionaries, as briefly set out in Section 2.8, are generally lacking. Also scarce in quantity is the study of discourse particles in intercultural communication. How do different groups of speakers use discourse particles in intercultural interactions? Do they display characteristics which originate from their first language or culture? More comparisons of discourse particles in intercultural settings are needed before we could seriously consider the answers to the above questions (see, for example, Cheng and Warren's (2001) study of *actually* in a corpus of intercultural conversations).

In view of the above issues, the present study visits some of the largely deserted areas in the field. By looking at data in intercultural settings, this research offers a new perspective for understanding discourse particles in the ever-growing field of intercultural encounters. The examination of the prosodic features of particles aims to further our understanding of the role prosody plays in the use of these linguistic items. To provide a comprehensive description of discourse particles, other syntactic, lexical, sociolinguistic and contextual factors are studied. Further, the link between discourse particles and language learning is investigated through the analysis of teaching materials. In sum, the present study is a modest attempt to deepen our knowledge of discourse particles by bridging some gaps in the field through empirical analyses, the methodology of which will be thoroughly described in Chapter 3.

Summary

Starting off with an overview of the literature, this chapter begins by discussing some of the controversial notions in the field surrounding the terminology, definition and classification of discourse particles. It then moves on to describe some of the most common properties of these items. An examination of how discourse particles have been analysed from different approaches is provided, leading to a detailed account of their discourse functions. Some largely neglected areas are also discussed, including the prosodic profiles of discourse particles and their relation to language learning. To highlight the pedagogical aspect of the present study, a report is given on the coverage of discourse particles in grammars and dictionaries, which constitute an important reference source for foreign language learners. Finally the chapter identifies gaps in the current literature and suggests how the present study could address some of the largely overlooked issues and contribute to the better understanding of discourse particles.

Chapter 3

Data description and research methodology

Abstract

This largely descriptive chapter is divided into two main parts. The first part consists of a detailed discussion of the three sets of data upon which the findings of the present study are based (3.1). In the second part, some research methodological issues which are relevant to the present study are addressed. Specifically, the analytical tools which are employed to carry out the study are introduced and the procedures involved in the entire research process are described (3.2).

3.1 Data description

The data used for the present study come from three sources, namely an intercultural corpus, a tailor-made sub-corpus of British English and a textbook database. The main findings concerning the use of discourse particles and its relation to various linguistic, sociolinguistic and contextual factors are generated from the intercultural corpus and form the bulk of the present study. In order to investigate the influence of linguistic background of speakers in greater detail and to verify the main findings, a customised subset of texts from a general corpus of British English is carefully selected and compared with results from the intercultural corpus. The pedagogical aspect of discourse particles is studied at length through the examination of the textbook database, which provides a contrast in the description of discourse particles between authentic speech and teaching materials.

3.1.1 The Hong Kong Corpus of Spoken English (prosodic)

The Hong Kong Corpus of Spoken English (prosodic) (HKCSE hereafter[19]) is a one-million word collection of naturally-occurring speech (see also Cheng, Greaves & Warren 2005; forthcoming, for further details of the corpus). As the name suggests, the corpus is both orthographically and prosodically transcribed. Texts in the corpus are primarily intercultural encounters between Hong Kong Chinese and speakers of

[19] The Hong Kong Corpus of Spoken English (prosodic) is usually referred to as 'HKCSE (prosodic)' to differentiate it from the orthographic version. For the sake of convenience, it will be abbreviated as 'HKCSE' in the present study.

languages other than Cantonese, though in a few settings only Hong Kong Chinese are involved. The combination of intercultural data with the availability of prosodic information makes the HKCSE useful for the present study. Since its introduction, the HKCSE has been the source of data for a number of intercultural studies in Hong Kong (see, for example, Cheng 2007; Cheng and Warren 2006). As the main purpose of building the HKCSE is to investigate the linguistic features of spoken English in Hong Kong in a variety of academic, business, social and professional domains, the participants in the HKCSE are all adults. Compiled between the mid-1990s and the turn of the millennium, the corpus consists of more than 100 hours of recordings from 311 texts evenly spread across four major settings in which English is spoken in the context of Hong Kong, viz. academic, business, conversational and public. Table 3.1 gives a summary of the composition of the four sub-corpora in the HKCSE:

Table 3.1. The composition of the HKCSE by domain and text type[20]

	Number of words	**Proportion (%)**	**Number of texts**
Academic	213,204	22.44	29
Business	259,484	27.31	112
Conversational	258,882	27.25	71
Public	218,402	22.99	99
TOTAL	949,972	100.00	311
ACADEMIC			
consultation	17,808	8.35	5
lecture	62,315	29.23	9
seminar and tutorial	38,610	18.11	6
student presentation and Q&A	91,077	42.72	8
workshop for staff	3,394	1.59	1
SUB-TOTAL	213,204	100.00	29
BUSINESS			
announcement and Q&A	22,103	8.52	3
conference call/video conferencing	6,017	2.32	2
informal office talk	27,338	10.54	4
interview	80,443	31.00	25
meeting	36,272	13.98	9
presentation	20,120	7.75	10
presentation and Q&A	51,218	19.74	4
service encounter	14,457	5.57	52

[20] A detailed index of the HKCSE is provided in Appendix 1.

workplace telephone talk	1,516	0.58	3
SUB-TOTAL	259,484	100.00	112
PUBLIC			
discussion forum	6,699	3.07	2
interview	87,151	39.90	25
press briefing	3,771	1.73	7
press briefing and Q&A	10,111	4.63	2
radio announcement	88	0.04	1
speech	88,443	40.50	57
speech and Q&A	22,139	10.14	5
SUB-TOTAL	218,402	100.00	99

As shown in Table 3.1, the HKCSE contains a variety of text types in the four contextually-based categories which are of varying degrees of interactiveness ranging from the highly scripted monologic public speeches to the spontaneous uninstitutionalized conversations. Apart from conversations which are all dialogic, the other three domains comprise monologues as well as multi-party talk. Owing to the difference in the duration of texts, the number of texts in each domain varies. In terms of the number of words, however, each sub-corpus roughly constitutes a quarter of the HKCSE.

The presence of a wide range of text types in the HKCSE makes it particularly useful for the purpose of studying the influence of context on the frequency and function of discourse particles. Quantitatively, the frequency rates of a particular discourse particle in different text types can be compared to determine in which text type(s) the particle is most prevalent, though it is important to bear in mind that in the present case some text types only contain a small number of texts and thus findings based on these texts could be merely suggestive. Qualitatively, the distribution of functions of a discourse particle across various text types could reveal how the particle is used in different ways according to the situational context.

As regards the language background of participants in the corpus, they are mainly divided into three groups: Hong Kong Chinese whose first language is Cantonese, native speakers of English[21], and speakers whose first language is neither Cantonese

[21] Given the global role of English, the dichotomy between native and non-native speakers of a language is increasingly being challenged (see, for example, Prodromou 2003, on "successful user of English"). It is therefore important to stress that in the present study the term 'native speakers of English' is simply used to refer to speakers whose first language is English, as opposed to speakers whose first language is not English.

nor English. A breakdown of the composition of the corpus according to the first language of speakers is presented in Table 3.2:

Table 3.2. The composition of the HKCSE by the first language of speakers

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Hong Kong Chinese	669,431 (70.7%)	173,966 (82.1%)	180,485 (69.7%)	122,373 (47.5%)	192,607 (88.2%)
Native speakers of English	240,847 (25.4%)	25,655 (12.1%)	74,287 (28.7%)	120,084 (46.6%)	20,821 (9.5%)
Speakers of other languages	36,298 (3.8%)	12,244 (5.8%)	4,109 (1.6%)	15,082 (5.9%)	4,863 (2.2%)
TOTAL	946,576 (100.0%)	211,865 (100.0%)	258,881 (100.0%)	257,539 (100.0%)	218,291 (100.0%)

From Table 3.2, it is observed that the majority of the participants in the HKCSE are either Hong Kong Chinese or native speakers of English. Across the four sub-corpora, Hong Kong Chinese contribute the greatest number of words. They account for 70.7% of the total speech produced. About a quarter of the whole corpus consists of speech produced by native speakers of English who are from countries such as Australia, the United Kingdom and the United States of America. For the present study, no attempt is made to categorise the native speakers in the HKCSE into different groups in relation to the varieties of English they speak. This is because the present study is mainly concerned with the comparison of speakers in terms of the first language they speak in the setting of Hong Kong and it is assumed that native speakers of English would be more similar to each other linguistically when compared with Hong Kong Chinese (Cheng 2003). In addition, the presence of different ethnic groups in the corpus reflects the multi-cultural and cosmopolitan nature of Hong Kong. Since one of the focuses of the study is to look at the use of discourse particles in intercultural encounters between Hong Kong Chinese and people that they normally interact with whose first language is English, the distinction between different varieties of native English would be relatively less essential in the present case. After all, for a corpus which contains an extensive number of texts, it is inevitable that participants may come from different backgrounds and the varieties of English they speak may be influenced by various factors such as regional and social variation, even in places where English is

predominantly the first language. It is therefore not feasible, if at all possible, to compile a large collection of texts from a wide range of contextual settings with participants from a fairly homogenous speech community, especially in a city like Hong Kong where a mix of ethnic groups is present. Further, findings from the HKCSE related to the linguistic background of speakers will be further compared with those from a customised corpus of British English of a comparable size to verify the conclusions drawn. This renders the differentiation between different native speaker groups unnecessary in the relatively small sample of native speakers' speech in the HKCSE.

Apart from Hong Kong Chinese and native speakers of English, a small number of speakers of other languages, such as Mandarin and French, are present in the corpus. They contribute only 3.8% of the total number of words in the HKCSE. Since the first languages of these speakers are distinct from one another and none of the speaker groups constitutes a representative population in the corpus, they are omitted from the examination of how linguistic background relates to the use of discourse particles.

In addition to the first language of participants, the present study also investigates the effect of gender on particle usage. Table 3.3 shows the number of words produced by male and female speakers in the whole corpus and across the four sub-corpora:

Table 3.3. The composition of the HKCSE by speaker gender

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Female speakers	389,659 (41.2%)	98,426 (46.5%)	133,473 (51.6%)	103,625 (40.2%)	54,135 (24.8%)
Male speakers	556,815 (58.8%)	113,441 (53.5%)	125,366 (48.4%)	153,855 (59.8%)	164,153 (75.2%)
TOTAL	946,474 (100.0%)	211,867 (100.0%)	258,839 (100.0%)	257,480 (100.0%)	218,288 (100.0%)

From Table 3.3, it is observed that the total number of words produced by male speakers is comparatively more when compared with female speakers. While in the academic and business settings the number of words spoken by the two gender groups is fairly evenly distributed, more words are produced by men than by women

in the other two sub-corpora. The proportion is most unequal in the public domain, with a ratio of three quarters of male speech to a quarter of female speech. This mirrors the fact that relatively speaking there is a greater proportion of men serving in the public sector, for example, as government officials and councillors, in the context of Hong Kong.

As will be demonstrated in the presentation of findings on the two particles *well* and *so* in Chapters 4 and 5, the breakdown of number of words in terms of text type, language background and gender of speakers provides valuable information which is essential to the quantitative analysis regarding the contextual and sociolinguistic factors which may affect the use of discourse particles.

In the HKCSE, all data were either audio-recorded or video-recorded. Most of the recordings were collected using an MD-recorder or a video-recorder by the researchers involved in the project, though a number of recordings in the business and public domains were retrieved from the media including the radio, television and relevant websites from the internet. Once the recording was made, it was first orthographically transcribed by a research assistant. Each orthographic transcription was coded with respect to a number of contextual and sociolinguistic factors that described the participants and the speech event. These codes include the first language as well as the gender of the speaker[22]. Whenever possible, background information such as the occupation of the speaker, the relationship between interlocutors, and bibliographic specifications of the text such as the date and venue of the speech event was noted in a separate index for future reference. Paralinguistic features such as laughter, throat-clearing and coughing are indicated in the transcription. Other details which have been carefully transcribed include truncated words, repetitions, simultaneous talk and inaudible speech. Although the marking of unfilled pauses was largely impressionistic and did not indicate the duration of the pause, attempts were made to consistently distinguish between a brief pause signalling a small gap of silence and a unit pause which normally lasts for a few seconds. When the initial orthographic transcription was completed, it was cross-checked by at least two researchers who are highly-experienced in working with spoken data. After a thorough checking and revision process, the orthographic

[22] For information about the transcription conventions used in the HKCSE, please refer to the transcription notation, which is provided before the introductory chapter.

transcription was then passed to an academic who is fully-trained in prosodic transcription using Brazil's (1985; 1997) discourse intonation framework. The fact that the prosodic mark-up was manually transcribed by only one individual for the whole corpus ensures that a high level of consistency was achieved. Again, cross-checking was done by at least two individuals who have conducted extensive research in the field of prosody in addition to an external consultant who is familiar with the discourse intonation system.

According to Brazil (1997), the discourse intonation framework primarily deals with the communicative value of intonation in English. In other words, it is concerned with the functions of intonation in the English language. Unlike some other approaches to prosody which propose that intonation has a fixed link with grammar or attitude, Brazil (1997) argues that intonation choices are deliberately made by speakers on a real-time, case-by-case basis according to their perceptions of the situational context. The application of the discourse intonation framework is thus especially relevant to those who are interested in interpreting the pragmatic meanings of intonation in naturally-occurring speech.

Strictly speaking, there are four systems in the discourse intonation framework, namely prominence, tone, key and termination. Out of these fours systems, speakers can make a total of thirteen intonation choices depending on the extra layers of meaning they intend to add to their message (Brazil 1997). These intonation choices in the systems are presented in Table 3.4 below:

Table 3.4. The intonation choices available in the discourse intonation framework (reproduced from Hewings and Cauldwell 1997:vii)

System	**Choices**	**Number**
Prominence	prominent/non-prominent syllables	2
Tone	rise-fall, fall, level, rise, fall-rise	5
Key	high, mid, low	3
Termination	high, mid, low	3

Within the systems of discourse intonation, a tone unit is the basic element for analysis. It is defined by Hewings (1990:136) as "a stretch of speech with one tonic segment comprising at least a tonic syllable, but which may also extend from an onset (first stressed syllable) to the tonic (final stressed syllable)". For each tone unit, speakers make a set of intonation choices from the four systems which are

independent from each other and individually constitute an aspect of discourse meaning. The system of prominence, which is sometimes referred to as accent or stress in other intonation models, is concerned with the choice speakers make in context to give extra emphasis on words which are "situationally informative" (Cheng et al. 2005:55). In terms of tones, speakers have five choices to choose from, including the two referring tones (fall-rise and rise), the two proclaiming tones (rise-fall and fall) and the level tone. While the two referring tones are employed when speakers want to establish that the information being conveyed is common ground, the two proclaiming tones are used to indicate new information is being presented. As for the level tone, Brazil (1997) states that it is chosen when speakers have no intention to proclaim or refer, and thus it serves the function of extricating speakers from the present situation. Its occurrences are also commonly found in the immediate surroundings of a hesitation, and are sometimes used for "rhetorical effect" (Brazil 1997:140). The last two systems, key and termination, are related to the pitch level choices speakers can make on the first prominent syllable and the final prominent syllable in a tone unit respectively. By definition, the last prominent syllable in a tone unit, which is also known as the nucleus in other models, carries the tone. There are a total of three options, namely high, mid and low, which speakers can select for the pitch level of the first and final prominent syllables in a tone unit. The selection of high, mid or low key carries contrastive, additive or equative value respectively (Brazil 1985). The choice of termination, on the other hand, could seek to constrain the response made by the next speaker (Brazil 1985).

In the present study, the intonational patterns of *well* and *so* in the HKCSE are examined with respect to Brazil's intonation framework. In particular, the prominence patterns and the nuclear tones of the two particles are discussed in great detail to see if they indicate whether the word is used as a discourse particle and if there is a correlation between intonation and discourse functions. Apart from looking at discourse particles through the discourse intonation model, the study also investigates the association of discourse particles with pauses and the positional distribution of particles in a tone unit based on the traditional English intonation account (see, for example, Crystal 1969; 1975; Roach 1983). More details of the analysis are provided in Section 3.2, where research methodology is outlined in depth.

3.1.2 The customised corpus from the British National Corpus (BNC)

As the major source of data in the present study, the HKCSE is used to investigate the properties of *well* and *so* as discourse particles, their pragmatic functions and the various linguistic, sociolinguistic and contextual factors which may affect their use. However, as suggested by Stubbs (2000), it is inadvisable to rely on a single corpus as there are in-built biases in all corpora, whatever their size and design. Findings from one corpus should therefore be compared with and checked against an independent corpus. The aim of looking for and employing a reference corpus, therefore, is to provide a supplementary source of speech of native speakers of English which can be contrasted with the HKCSE in order to further examine the similarities and differences between first language speakers of English and Hong Kong Chinese in the use of discourse particles, and to determine the extent to which core findings generated from the HKCSE are valid.

In the making of such a comparable corpus, time and resources inevitably impose practical constraints on the project. While the compilation of a native speaker corpus based on the same corpus design and selection criteria as the HKCSE would be ideal, such an undertaking, which involves collecting and transcribing spoken language on a large scale, would be too time-consuming and unfeasible within the scope of the present study. Therefore, a hunt for a publicly available native speaker corpus which is largely comparable to the HKCSE appears to be the next best option.

While there are a growing number of corpora available for commercial and academic research, ready-made collections of speech are still difficult to come by. Given the complexity involved in the construction of a spoken corpus, it is not surprising that the choice available for research is more restricted in comparison with their written counterparts. The availability of a publicly accessible prosodically transcribed spoken reference corpus, in particular, is more limited, let alone a ready-made spoken corpus annotated with the discourse intonation framework (Brazil 1997) used in the HKCSE. Because of the immense input of time and expense in transcribing a corpus prosodically, no such corpus of a comparable size exists[23]. Therefore, it was decided at the outset that the comparison would not involve the prosodic properties of discourse particles.

[23] To the best of my knowledge, the HKCSE (prosodic) is the largest prosodically transcribed corpus applying Brazil's (1997) discourse intonation framework currently in existence.

After surveying the range of English spoken corpora accessible in the public domain, the British National Corpus [24] (BNC hereafter) was chosen for the comparison. More precisely, a subset of texts was selected from the BNC to form a customised corpus (BNC customised hereafter) for comparative purposes with the HKCSE (see also Lam 2007, for more details of the process involved in the building of the customised corpus and the problems encountered).

The BNC is one of the largest general-purpose corpora of the English language. It is generally taken to be a monolingual corpus of British English consisting of text samples which were produced in contemporary Britain from the latter part of the 20th century (Burnard 2000). The BNC contains 100 million words of written and spoken language from more than 4,000 text samples, of which 10 million words are from spoken sources. When the construction of the corpus was completed in 1994, the BNC formed the largest collection of spoken texts in English ever assembled at the time (Kennedy 1998). For the present study, all texts in the customised corpus were selected from the spoken section of the BNC.

The BNC spoken section contains 10 million orthographically transcribed speech samples from over 900 texts. It is sub-divided into two components, namely the demographic and the context-governed parts. The demographic component contains everyday talk recorded by 124 adult volunteers across the country using portable cassette-recorders as they went about doing their business in a period from two days up to a week. It comprises four million words representing the speech of four social classes from six different age groups. In order to complement the demographic component which contains mainly impromptu speech, the context-governed part was created. It consists of six million words of text samples which are chosen according to four equal-sized contextually-based categories, namely, educational and informative, business, public/institutional and leisure. For each category, the proportion of monologues and dialogues is 40% and 60% respectively.

Being one of the few large computerised collections of spoken native English which are publicly available, the spoken part of the BNC serves as a source for a large number of speech samples from a wide range of text types. As shown from the previous paragraph, the contextually-based categories in the context-governed component of the BNC spoken section are arranged in domains similar to the ones in

[24] In the present study, all references to the British National Corpus refer to the BNC World Edition.

the HKCSE. This facilitates the comparison of the two corpora in terms of the influence of text type on the use of discourse particles. Since text samples in the BNC are stored in different folders and each can be opened separately, end-users could easily select files individually from the corpus for the creation of tailor-made sub-corpora. The fact that the BNC is both available on a CD-ROM and via online subscription also means that the corpus can be accessed on a stand-alone personal computer, hence avoiding the technical problems associated with internet connection speed or server breakdown. Similar to the HKCSE, the BNC is also a static corpus. No new texts have been added since the completion of the project. Its collection of texts from the late 20th century thus provides a fairly comparable source of British English for the contrast with the HKCSE with respect to the date of the texts. Taking into consideration the various factors such as the composition, the ease of accessibility and the date of corpus outlined above, the BNC is considered more suitable for sub-corpus creation for the present study when compared with other corpora on the market.

For the purpose of direct comparison with the HKCSE, considerable effort has been made in the process to search and locate texts in the BNC spoken section which are most comparable to those in the Hong Kong corpus. A number of inventories and databases containing the bibliographic information of texts in the BNC have been consulted, including the official reference guide provided by the original corpus compilers (Burnard 2000) and the BNC Index (Lee 2003). Nevertheless, due to the different designs of the two corpora, it is impossible for the compositions of both sets of texts to be completely identical in terms of all the external factors concerned. Owing to the insufficient or incorrect provision of speaker details in many file headers of the BNC and the difficulty involved in identifying the sociolinguistic features of speakers, it is not always possible to take sociolinguistic factors into account when selecting texts from the BNC. As a result, the influence of gender on the use of discourse particles is not examined in the customised corpus. For the sake of convenience of comparison and a dearth of suitable texts, the actual make-up of the customised corpus from the BNC is roughly the same size as the HKCSE, arranged in a categorisation framework largely similar to the Hong Kong Corpus. Table 3.5 shows the composition of the BNC (customised):

Table 3.5. The composition of the customised corpus from the BNC[25]

	Number of words	Proportion (%)	Number of texts
Academic	212,532	22.40	25
Business	260,484	27.45	86
Conversational	258,391	27.23	122
Public	217,513	22.92	25
TOTAL	948,920	100.00	258
ACADEMIC			
consultation	15,641	7.36	3
lecture	63,395	29.83	9
seminar and tutorial	46,730	21.99	6
teaching sessions with more student talk	86,766	40.82	7
SUB-TOTAL	212,532	100.00	25
BUSINESS			
announcement and Q&A	31,235	11.99	4
informal office talk	26,586	10.21	19
interview	81,201	31.17	8
meeting	42,055	16.14	6
presentation	19,332	7.42	3
presentation and Q&A	47,263	18.14	6
service encounter	11,386	4.37	35
workplace telephone talk	1,426	0.55	5
SUB-TOTAL	260,484	100.00	86
PUBLIC			
interview	95,429	43.87	10
speech	89,148	40.99	10
speech and Q&A	32,936	15.14	5
SUB-TOTAL	217,513	100.00	25

By comparing Table 3.5 with Table 3.1 which shows the composition of the HKCSE, it is observed that there are some minor differences in text type in the HKCSE and the BNC (customised) owing to the lack of similar texts found in both datasets. Several text types which are found in the HKCSE are not present in the BNC. Close alternatives are selected wherever possible. Specifically, the absence of the text types "conference call/video conferencing", "press briefing", "press briefing and Q&A" and "radio announcement" in the BNC results in the increase in the inclusion of their

[25] A list of the files and extracts selected from the BNC for the creation of the customised corpus is provided in Appendix 2.

close alternative text types, i.e. “meeting”, “speech”, “speech and Q&A” and “speech” respectively in the customised corpus of the BNC. For text types such as “workshop for staff” and “discussion forum” in the HKCSE which do not constitute a large proportion in the category and do not have comparable substitutes, the proportions of other text types are slightly increased to make up for the difference. In the academic domain, teaching sessions with more student talk are selected in the BNC for comparison with “student presentation and Q&A” in the HKCSE, as no other suitable text types are available. This ensures that the composition of texts in terms of formality, contextual setting and interactive level is largely comparable in the two corpora.

As discussed earlier, the conversations in the demographic component of the BNC were collected when recruits participated in their daily activities. Admittedly, these spoken texts provide an authentic record of the language produced by people in their everyday life. However, it also means that the demographic section contains both transactional talk and interactional talk. Speech events included in the demographic part are wide-ranging, from talking to salespersons or going to dentists to having music lessons or attending training sessions at work. As noted by Lee (2001:59), the so-called “conversations” which make up the spoken demographic component of the BNC are “less well-defined” than the written texts. Although these texts often involve a diverse range of tasks and topics and as a result different communicative purposes, they are grouped under the umbrella term “conversation”, which is glossed as spoken interaction in general. This is different from the conversational sub-corpus of the HKCSE, where “conversation” is interpreted as “a speech event outside of an institutionalised setting involving at least two participants who share the responsibility for the progress and outcome of an impromptu and unmarked verbal encounter consisting of more than a ritualised exchange” (Warren 2006:11). Consequently, all texts in the conversational sub-corpus of the HKCSE are spontaneous speech events which take place outside specific institutional settings, with perceived equal rights and status between participants. Such a discourse type focuses on maintaining relationships and sharing background knowledge, rather than getting business done (Biber 1988). In order to ensure that the conversations in the BNC (customised) are largely comparable to those in the HKCSE, great care has been taken to select only segments from the conversation files in the demographic component of the BNC which are non-transactional. Therefore, some texts in the

BNC (customised) are excerpts of the full texts in the original corpus. Only parts of which are suitable for comparison are included in the customised collection.

It is worth noting from the above discussion that the selection of texts for the BNC (customised) focuses on retrieving texts which are closest matching to those in the HKCSE in terms of their socio-situational parameters. In this case, the BNC customised corpus is intended to be representative in the sense that it represents the collection of texts from the BNC which is most suitable for comparison with the HKCSE, instead of a freestanding representative spoken subsection of the original corpus.

While the HKCSE contains prosodic mark-up, the BNC is annotated with part-of-speech (POS) tags and other codes indicating paralinguistic information. However, for better illustrative purposes of the present study, all POS tags are omitted when examples are drawn from the BNC (customised). Other details such as the occurrences of pauses and paralinguistic features are included.

3.1.3 The textbook database

In an attempt to provide a description of discourse particles in teaching materials which is lacking in the literature and to compare particle usage in 'real' English and 'school' English, a database consisting of English textbooks collected in Hong Kong was created. In Hong Kong, the English language is almost entirely acquired through formal education. In the classroom, textbooks constitute an essential part of English language teaching. Therefore, an examination of how discourse particles are described and used in English language textbooks reveals to a large extent the attention given to these lexical items in the local pedagogical setting.

For the present study, English textbooks for upper secondary students in Hong Kong were chosen. There are three reasons for this decision. Firstly, students at the upper secondary level who are about 17-18 years of age have already been learning English for at least eleven years since they enter formal schooling. They are generally considered advanced learners of English who have reached a fairly satisfactory level of grammatical competence. At that learning stage, it is assumed that they should not only focus on whether their English is correct grammatically but also pragmatically. Since it has been pointed out in various studies that the appropriate use of discourse particles contributes to the pragmatic competence of speakers (see, for example, Müller 2005; Svartvik 1980), it would be interesting to

examine whether discourse particles are introduced as devices for achieving pragmatic competence in upper-secondary English textbooks. Secondly, discourse particles are given special emphasis in the syllabus for upper secondary schools in Hong Kong. In the official document published by the Curriculum Development Council (1999:20), it is stated under the area of listening skills that students in the sixth form should be able to "understand the use of discourse markers". Under speaking skills, students are expected to "use discourse markers appropriately" (ibid.). It is therefore not unreasonable to speculate that if some description of discourse particles is to be incorporated in English language teaching, it is likely to be included in English textbooks at that level. A final reason for choosing upper-secondary textbooks is related to the examination system. In the educational setting in Hong Kong, students sit public examinations for university entry upon the completion of their upper-secondary study which lasts for two years. For the assessment of the use of English, there is an oral component which tests the candidates' ability to use spoken English as it might be encountered in academic or vocational situations (Hong Kong Examinations and Assessment Authority 2003). Candidates are assessed in two parts for their oral skills, including a short presentation based on a passage provided and a group discussion on a specific topic with other candidates. Given the exam-oriented nature of the educational environment in Hong Kong, it is likely that a substantial amount of training in English oral skills is catered for in teaching materials. In fact, at the upper-secondary level, textbooks of the same series are usually published as individual components, each targeting a specific section of the English examination. As a result, English textbooks specially designed for oral practice are common. Presumably, textbooks focusing only on oral English should be more specialised in the teaching of features which are typical of speech when compared with general English textbooks. Accordingly, it was anticipated that they may contain a more comprehensive coverage of discourse particles, which are characteristic of spoken discourse.

At the time of data collection for the present study, there were 15 upper-secondary English textbooks available on the market which contained a component on teaching oral skills. In secondary schools in Hong Kong, off-the-shelf textbooks are widely used for English language teaching. School-based materials are uncommon and only serve supplementary purposes if present. It is thus reasonable to believe that these readily accessible textbooks provide a representative collection of

teaching materials which are actually used by teachers and students at the specified level of English language instruction at the time the study is conducted. Table 3.6 lists the 15 upper-secondary English textbooks which constitute the textbook database examined in the present study:

Table 3.6. The list of upper-secondary English textbooks used in the study

Book title	Year of publication	Publisher	Author(s)	Oral (O) / General (G) textbook
Expression: Oral Practice for AS-level	1994	Precise Publications	Duncan, J.	O
Skills for Success Section D Oral	2003	Macmillan		O
Speaking Precisely 1	1999	Precise Publications	Sutton, M. E. & Duncan, J.	O
Speaking Precisely 2	1999	Precise Publications	Sutton, M. E.	O
Steps & Skills Oral 6	2003	Witman	Potter, J.	O
Steps & Skills Oral 7	2003	Witman	Potter, J.	O
Teach & Practice: AS-level Oral English for Form 6	1999	Pilot Publications	Esser, D.	O
Teach & Practice: AS-level Oral English for Form 7	1999	Pilot Publications	Esser, D.	O
Use of English Oral Handbook	2002	Free Press	Leetch, P.	O
Use of English Oral Practice Papers Form 7	1993	Macmillan	Gran, B. J.	O
Skills Building for AS Use of English	1999	Longman	Lee, I. & Holzer, V.	G
Step by Step 6	2000	Ling Kee	Wong, W. & Etherton, A.	G
Step by Step 7	2001	Ling Kee	Wong, W. & Etherton, A.	G
Use English (volume 1)	2003	Macmillan	Li, A.	G
Use English (volume 2)	2003	Macmillan	Li, A.	G

The textbook database consists of coursebooks published from the year 1994 to 2003 which are used for the two-year duration of upper-secondary study in Hong Kong. Some textbooks are designed for the whole period, while others are only part of the series of two volumes, such as *Speaking Precisely* (1999). Altogether they represent the teaching materials from seven publishers. As shown in the last column of Table 3.6, two-thirds (10 out of 15) of the textbooks collected are specifically designed for the training of oral skills. The remaining one-third (5 out of 15) are general English textbooks which cover all skills (listening, reading, speaking and writing) and contain a component which focuses on the spoken language. For these general textbooks, only the oral section is examined. It is worth noting that all textbooks in

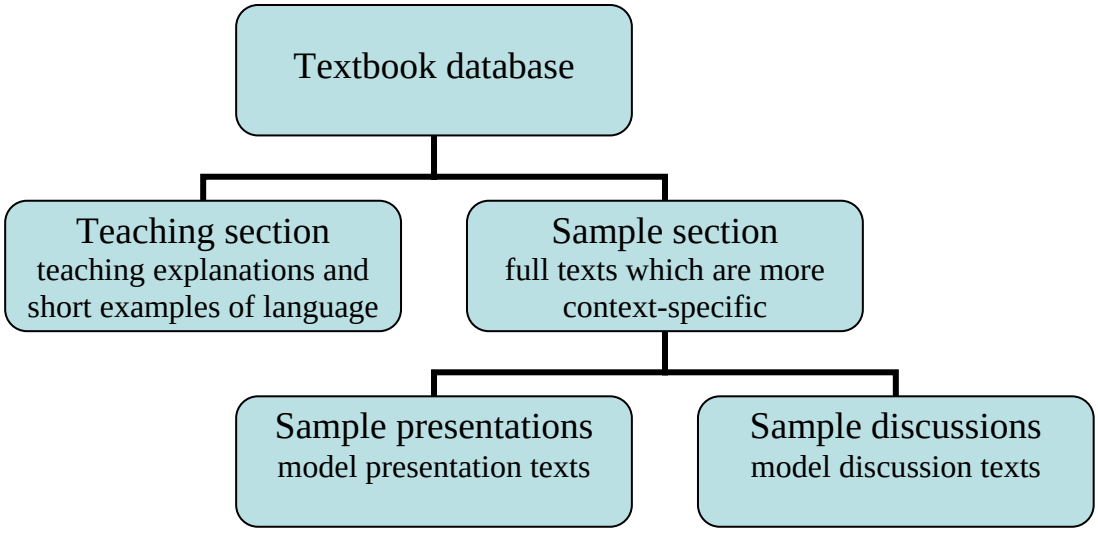
the database contain some teaching points or explanations. Drill exercise books which only consist of practice materials or exam papers are not included.

Unlike the two corpora which are electronic collections of text, the textbooks used in the present study are physical stacks of paper and thus need to be digitised before further analysis can proceed. Once the textbooks were collected, they were turned into machine-readable form by means of a scanner. Each textbook page was first scanned into the computer as an image file in tif-format. It was then processed with the optical character recognition (OCR) software *Microsoft Office Document Imaging* (version 11.0) to convert the image file into a text file. The success rate of the conversion principally depends on the complexity of the composition of the page. For textbook pages which are interspersed with all kinds of visual materials such as graphs and pictures as well as texts, the conversion results are normally less accurate. Therefore, manual checking of each text file generated from a textbook page was required to ensure that characters on the image file were correctly identified so that what was displayed in the text file matched the original textbook page. After the checking and editing procedures, each textbook page was saved in plain text format for subsequent data retrieval and could be processed in the same way as texts in an electronic corpus.

When all the textbooks were ready for electronic text analysis, their contents were thoroughly examined. Owing to the fact that these textbooks are used for the preparation of the public examination, it is not surprising that they mainly concentrate on presentation and discussion skills, which are the two areas being assessed in the oral examination component. Generally speaking, the contents of the textbooks could be divided into two parts: the teaching section and the sample section. The teaching section contains language teaching points which are explicit statements and explanations of what is being taught as well as short and often detached examples of language which are suggested to be used in presentations and discussions. These examples include dialogues and ‘speech bubbles’ in the textbooks. All the exercises which contain gaps for students to fill in are excluded. The sample section contains longer and often more context-specific texts which resemble the structures of authentic presentations and discussions. These sample texts are presented as ‘models’ for students to follow. Together these two sections make up

the textbook database of the present study[26]. Figure 3.1 shows the basic structure of the textbook database:

Figure 3.1. The composition of the textbook database



In previous studies of teaching materials, it appears that there have been very few guiding principles as regards what should be included in a textbook corpus (Nelson 2000). Decisions in most cases seem to be based upon the specific research purpose of the study and the contents of the textbooks examined (see, for example, Ljung 1990; Römer 2005). For the present study to give a comprehensive account of discourse particles in the textbooks examined, it is important to find out how discourse particles are described as well as how they are used in context in textbooks. In other words, both the descriptions and actual usage of discourse particles are studied. Accordingly, the teaching section of the textbook database is used to investigate whether explanations and short examples in textbooks overtly teach how discourse particles should be used. The analysis of this section is only qualitative. It aims to see if textbook descriptions of particles truly mirror usage in authentic data. The sample section, on the other hand, provides a source for examining how discourse particles are realised. As these sample texts are suggested by textbook writers as the 'models' of what presentations or discussions in the examination should be like, they normally have a rather complete structure similar to a

[26] For illustrative purposes, two textbook pages from the teaching section and two textbook pages from the sample section of the textbook database are provided in Appendix 3.

presentation or a discussion in naturally-occurring discourse. Consequently, these sample texts are analysed both qualitatively and quantitatively. In other words, they are treated the same way as texts in a corpus to study the frequency, positional and functional distributions of discourse particles. This draws a direct comparison of the use of discourse particles between invented texts in textbooks and authentic data. Admittedly, one could argue that textbook sample texts are constructed purely for pedagogical purposes and hence it is impossible to find text types in the real world which are completely analogous. Nevertheless, a comparison of the textbook model presentations and discussions with some similar text types in naturally-occurring data still offers useful indicators concerning the extent to which textbook data converge or diverge from real-life examples as regards the use of discourse particles. Table 3.7 shows the compositional details of the sample section:

Table 3.7. The composition of the sample section in the textbook database

	Sample presentations	**Sample discussions**
Total number of words	10,817	13,032
Total number of texts	38	11

In total, the sample section is made up of 49 texts with 23,849 running words. It is a rather modest size when compared with the two corpora used. However, one has to bear in mind that textbook corpora, which are an example of specialised corpora, are usually much smaller than general corpora (cf. for example Nelson 2000; Römer 2005). As pointed out by Sinclair (2001:xi), "[a] small corpus is seen as a body of relevant and reliable evidence" which can be "interpreted by the scholar directly". Its small size makes it more manageable, which in turn facilitates manual analysis. In the present case, the statistical study based on the sample section is supplemented by the qualitative analysis of the description of particles in the teaching section. In combination, the textbook database offers a balanced and representative picture of discourse particles in the textbooks examined.

3.2 Research methodology

This section describes in detail the procedures involved in the research and the analytical tools employed in the process. It also discusses some methodological issues which need to be addressed in the course of the present study.

3.2.1 Data selection and pre-processing

Before any actual analytic work on discourse particles could be done, much effort has been devoted to laying the groundwork for the study. This preparatory step involves the selection and pre-processing of data, which includes the creation of the customised sub-set of texts from the BNC, the building of the textbook database and the calculation of the number of words for each dataset and all their components. As the present research investigates a number of sociolinguistic and contextual factors in the use of discourse particles, it is necessary to work out the number of words produced by each speaker group under examination and the number of words for each text type before any further statistical analysis as regards how frequent the words being studied are used as discourse particles can be carried out. The most laborious process in this procedure entails segmenting every text in the HKCSE into utterances and categorising and grouping them according to the sociolinguistic groups of the speakers. At the time the data pre-processing of the study was conducted, this segmentation process had to be done manually. After the texts were divided into segments, the total number of words in the corpus for each sociolinguistic factor examined could be determined.

In the word counting procedure, it was discovered that the word counts in texts could vary depending on the computer program used. For example, the word processing software Microsoft Word uses an algorithm which is different from the one used in the linguistic software package WordSmith Tools (Scott 2005), resulting in a different word count. For the files selected from the BNC, word count information could be obtained from the file header, the official BNC reference guide (Burnard 2000) as well as the BNC Index (Lee 2003). However, all these references yield different totals. In order to maintain a level of consistency, the present study uses a single algorithm for the calculation of words for all the three datasets examined. The word count function in the program Wordlist of the package WordSmith Tools (version 4.0) is used to calculate the number of words for all the texts in the HKCSE, the BNC (customised) and the textbook database.

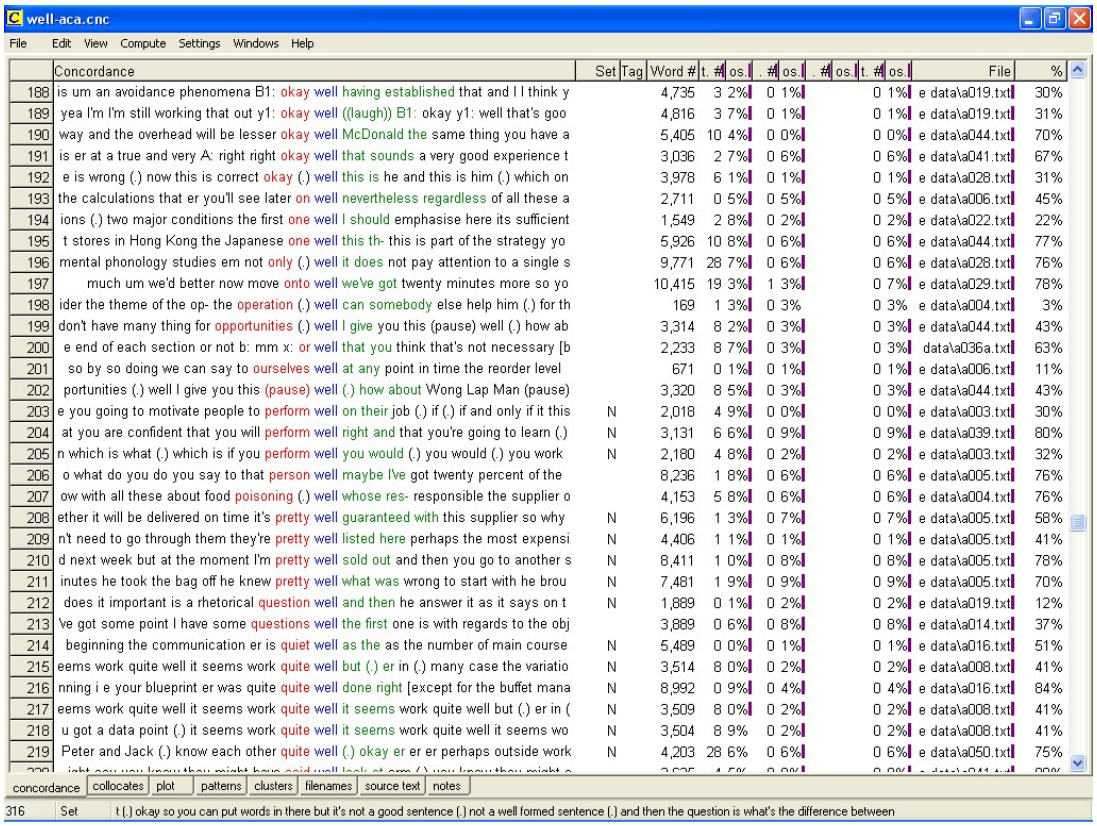
3.2.2 Data retrieval / extraction

When the three sources of data the HKCSE, the BNC (customised) and the textbook database were set for analysis, word lists were generated by the program Wordlist in

WordSmith Tools to find out the frequency of linguistic elements such as words and phrases which could be used as discourse particles. As the present research aims to incorporate both qualitative and quantitative analysis in the study of discourse particles, the lexical items chosen need to occur in sufficient numbers in all the three datasets for the quantitative approach to be meaningful. The two discourse particles *well* and *so* were selected from among other potential candidates because of their high frequency in the two corpora in addition to them being considered to be adequately represented in occurrences in the textbook database (see Chapters 4 and 5 for details of frequency information). In other words, they are not only characteristic of authentic spoken data but also commonly found in teaching materials. Their prevalence thus ensures that a reasonable number of occurrences of the two items can be studied in the quantitative analysis. Apart from frequency, the varying amount of attention given to *well* and *so* in previous studies also makes them an interesting pair of items for research. In the literature, *well* is generally considered a prototypical example of discourse particles. It has probably received more attention than any other particle (Aijmer 2002; Schourup 2001). Findings on *well* from the present research could therefore be readily compared with those from earlier studies. On the other hand, the status of *so* as a discourse particle is more marginal, as the number of studies of *so* has been relatively limited (Müller 2005). Results generated from the present work concerning *so* thus could enhance our understanding of a largely overlooked particle. Further, a survey of the literature has shown that the two words share some similarities as discourse particles and yet they also display different characteristics (see also Chapters 4 and 5 for previous studies of *well* and *so*). In sum, the present study serves to provide a comprehensive portrayal of two of the most frequently occurring discourse particles which have not been studied to an equal degree before.

Once a decision was made concerning the lexical items to be focused upon in the research, all instances of *well* and *so* had to be retrieved from the datasets. For the present study, the Concord program in WordSmith Tools was used to generate concordance lines of all occurrences of the two search words in the orthographic transcription of the HKCSE, the BNC (customised) and the textbook database. Figure 3.2 displays some concordance lines of the word *well* from the academic sub-corpus of the orthographic transcription of the HKCSE:

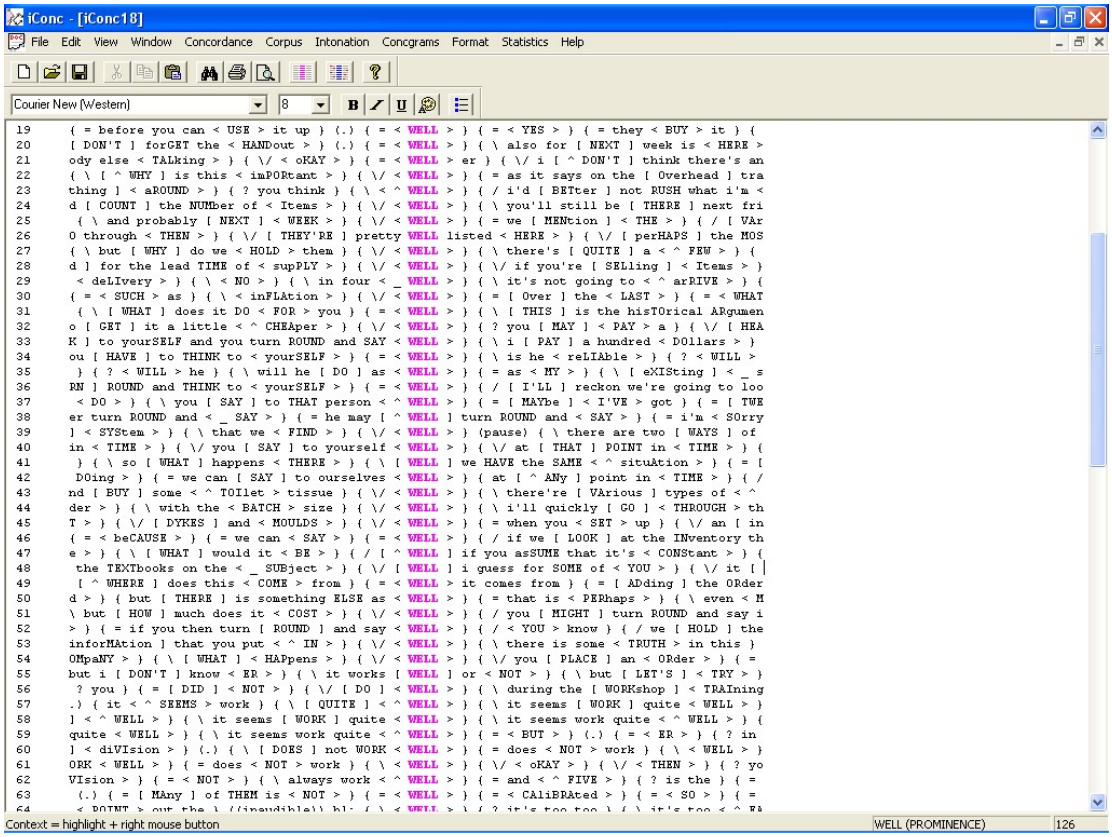
Figure 3.2. A screenshot of Concord in WordSmith Tools showing concordance lines of *well* from the academic sub-corpus of the HKCSE orthographic component



For every concordance line generated, Concord provides the file name in which the search word is located. A word number is also given based on the chronological sequence in which the word appears in the text. The file name and the word number together thus serve as a unique code for identifying a particular instance of the search word. This greatly facilitates the retrieval of a certain occurrence of the particles examined out of thousands of concordance lines. For all the examples cited in the present study, their source will be provided for ease of reference. When examples from the HKCSE and the textbook database are referred to, the citation consists of three parts, which includes the name of the dataset, the file or the section from which the search word is found, and finally the word number. For examples extracted from the BNC (customised), the sentence number of the example is added before the word number in accordance with the reference stipulation required by the BNC compilers.

For the prosodic transcription of the HKCSE, a specially designed search engine iConc© (Greaves 2005) was used to retrieve concordance lines of *well* and *so* showing all the prosodic features in their neighbouring linguistic context. Figure 3.3 shows a screenshot of iConc with concordance lines of *well* from the HKCSE:

Figure 3.3. A screenshot of iConc showing concordance lines of *well* from the prosodic transcription of the HKCSE



Once generated, these concordance lines were saved for subsequent analysis. In total 8,244 concordance lines of *well* and 14,710 of *so* were retrieved from the HKCSE, the BNC (customised) and the textbook database which were then studied in various ways to examine the relations between the use of discourse particles and various linguistic, sociolinguistic and contextual factors.

3.2.3 Codification

Although the concordancing programs Concord and iConc are useful for looking at the linguistic patterning of words, they are not designed for handling multiple variables and the relations between parameters. As a result, electronic databases were created for each particle examined using the software Microsoft Excel (version 2003). The first step in the creation of these databases involved the export of all the concordance lines generated from Concord and iConc to Excel so that they are in an Excel-compatible format. After the incorporation of all concordance entries into

spreadsheets, a set of features were studied with respect to every concordance line and each occurrence of the search word was coded in terms of these features. Depending on the dataset under investigation, the number of features being examined varied. Table 3.8 lists the types of linguistic, sociolinguistic, contextual and prosodic features which are analysed for each concordance line in the three datasets of the study:

Table 3.8. The list of features analysed in the HKCSE, the BNC (customised) and the sample section in the textbook database in the present study[27]

	HKCSE	**BNC customised**	**Sample section in textbook database**
LINGUISTIC FEATURES			
discourse particle status	√	√	√
utterance position	√	√	√
SOCIOLINGUISTIC FEATURES			
speaker gender	√	X	X
speaker linguistic background	√	√	X
CONTEXTUAL FEATURES			
contextual domain	√	√	X
text type	√	√	X
PROSODIC FEATURES			
tone unit position	√	X	X
nuclear tone	√	X	X
prominence	√	X	X
prominence position	√	X	X
key pitch level	√	X	X
termination pitch level	√	X	X

For every concordance line generated, two linguistic features were examined regardless of the source of data they came from. The feature “discourse particle status” involves the determination of whether the search word in the concordance line is used as a discourse particle. In the present study, the lexical items *well* and *so*

[27] A tick √ indicates that the feature is analysed in the given dataset and a cross X indicates the feature is not analysed in the given dataset.

are considered discourse particles if they can be omitted without changing the propositional content of the utterance containing them. As will be shown in subsequent chapters, other linguistic and prosodic features also provide useful indicators of the discourse particle status of a particular occurrence of the two words examined. The second linguistic feature, "utterance position", concerns the placement of a particular occurrence of the search word in an utterance. In the past, various ways of segmenting and dividing chunks of talk in the spoken language have been proposed. In this study, the concept of 'utterance' is used to better describe the positions of discourse particles which occur in overlapping speech. Following Stenström (1994:226), the present study defines an utterance as "anything that a speaker says". Under this definition, single words such as *yea* and *right* which are produced as feedback tokens while the other interlocutor is speaking are regarded as utterances. Consequently, for the present study the feature "utterance position" has four values: initial, medial, final and stand-alone, which are explained in Table 3.9:

Table 3.9. The values for the feature "utterance position"

Values	**Explanations**
Initial	First word of an utterance which contains more than one word, or only preceded by particles and/or fillers[28]
Medial	Anywhere not initial or final of an utterance which contains more than one word
Final	Last word of an utterance which contains more than one word
Stand-alone	Single-word utterance, or only preceded and/or followed by particles and/or fillers

The following examples illustrate the four possibilities of utterance position. Example 1 shows the use of *well* in initial position:

(1)

A: { = what are the [< PLANS >] } { = for [< reDUcing >] } { = the mountains of [< WASTE >] } { = [here] in < HONG > Kong }

a: { = ***well*** we [< ^ HAVE >] } { \ [< STARted >] the } { = the [< SEparation >] of } { = [< ^ BOTTles >] } { = [< PAper >] } …

(HKCSE, P077, 247)

[28] In this study, the position of a particular instance of *well* or *so* is taken to be initial if it is only preceded by other particles and/or fillers. This is because items which are supposed to be discourse particles frequently cluster in discourse, including at the beginning of an utterance (Schourup 1999).

Example 2 shows the use of *well* in medial position:

(2)

b: { \ so the [< COURSE >] } { = [< I >] } { = [I] I I < I > } { \ [< WENT >] } { \ [THIS] < WEEK > } { \ [< _ ***WELL*** >] } { = [NAMED] < ER > } { = [< MAnagement >] } { \ [DEvelopment] < COURSE > }…

(HKCSE, C109, 980)

The use of *so* is found in final position in example 3:

(3)

a: …{ = [< IF >] } { \ [THERE] is a muTAtion at c < LOcus > } { = [IT] will FIND to < BE > } { = [< ASsociate >] } { = [< WITH >] the di } { = [< DIminished >] } { \ [TYrosinase] < acTIvity > } { = and [< ***SO*** >] }

(HKCSE, A025, 119)

Finally, example 4 indicates the use of a stand-alone *so*:

(4)

A: ((laugh))

b: { / [< ***SO*** >] }

A: { = [< I >] } { = [< I >] } { \ i [< DON'T >] think } { ? there's any } { \ [WAY] < aROUND > that }…

(HKCSE, B058, 1635)

As mentioned before, the two sociolinguistic features "speaker gender" and "speaker linguistic background" are monitored in the HKCSE. The feature "speaker gender" contains three values: male, female and unknown. For the feature "speaker background", there are four values: Hong Kong Chinese (HKC), native speakers of English (NSE), others and unknown. In the HKCSE, only the use of discourse particles by the two major speaker groups, i.e. Hong Kong Chinese and native speakers of English is compared, for reasons given earlier in this chapter. Being a sub-set from the collection of contemporary British English, the BNC (customised) constitutes another source of speech of native speakers of English. It is used to compare with data in the HKCSE in terms of the first language of speakers. The two contextual features "contextual domain" and "text type" are concerned with the

situational context of the texts. As discussed earlier, texts in the BNC (customised) were specially chosen to be largely comparable to texts in the HKCSE. Therefore, the two corpora share a similar structure comprising four contextual domains containing mostly equivalent text types. The values of the contextual feature "contextual domain" are academic, business, conversational and public accordingly. For the feature "text type", the numbers of values identified in the HKCSE and the BNC (customised) are 21 and 15 respectively. Details of these values are summarised earlier in Table 3.1 and Table 3.5. For data in the textbook database, all sociolinguistic and contextual features are not applicable and thus are not studied.

Six prosodic features are analysed for data in the HKCSE. They are namely tone unit position, nuclear tone, prominence, prominence position and the pitch levels of key and termination. The values for the four features concerning tone, prominence, key and termination are described earlier in Table 3.4 in the discussion of intonation choices available in the discourse intonation framework. The values for the remaining two features "tone unit position" and "prominence position" are explained below in Table 3.10 and Table 3.11 respectively:

Table 3.10. The values for the feature "tone unit position"

Values	**Explanations**
Pre-head	An unstressed syllable in a tone unit preceding the first stressed syllable
Head	Anywhere from the first stressed syllable up to (but not including) the tonic syllable in a tone unit
Tonic syllable (in a shared tone unit[29])	A syllable which carries the tone in a tone unit where it is not the only element
Tail	A syllable between the tonic syllable and the end of a tone unit
Separate tone unit	A syllable constituting a tone unit on its own
Unknown	Prosodic pattern of a tone unit is indeterminable
Not applicable	Lack of prosodic information

[29] The term 'shared tone unit' is employed in this study to refer to tone units where there are more than one element. It is chosen instead of labels such as 'multi-word tone unit' to avoid the meaning laden with 'word', as this might exclude items such as fillers which occur frequently in speech but have no proper word status within the language system.

Table 3.11. The values for the feature “prominence position”

Values	Explanations
Unstressed	An unstressed syllable in a tone unit
First	The first stressed syllable in a tone unit
Middle	A stressed syllable anywhere between the first and last stressed syllables in a tone unit
Last	The last stressed syllable (also the tonic syllable) in a tone unit
Sole prominent syllable (in a shared tone unit)	The only prominent syllable in a tone unit containing more than one syllable
Sole prominent syllable (in a separate tone unit)	The only prominent syllable in a tone unit containing only one syllable

In the codification process, every concordance line of the search words was studied in relation to the set of features discussed in this section. A value was then given to every feature examined and entered in the Excel database for further analysis. A review of the twelve features is given in Table 3.12:

Table 3.12. A summary of the details of the features examined in the present study

Features examined	Number of values	Remarks
LINGUISTIC FEATURES		
discourse particle status	2	whether it is used as a discourse particle
utterance position	4	initial/medial/final/stand-alone
SOCIOLINGUISTIC FEATURES		
speaker gender	3	male/female/unknown
speaker linguistic background	4	HKC/NSE/others/unknown
CONTEXTUAL FEATURES		
contextual domain	4	academic/business/ conversational/public
text type	21/15	please refer to Table 3.1 and Table 3.5 for details
PROSODIC FEATURES		
tone unit position	6	please refer to Table 3.10
nuclear tone	5	rise-fall/fall/level/rise/fall-rise
prominence	2	prominent/non-prominent syllables
prominence position	6	please refer to Table 3.11

key pitch level	3	high/mid/low
termination pitch level	3	high/mid/low

3.2.4 Functional analysis

When all the concordance lines generated were examined with respect to the feature "discourse particle status", a detailed functional analysis was carried out for those occurrences of the search words which were regarded as discourse particles. As discussed in the previous chapter, various functional approaches have been used in the study of discourse particles which could be subsumed under two major categories, namely top-down and bottom-up, depending on whether they start from a given theoretical framework or from individual occurrences. The present study follows a bottom-up corpus-linguistic approach in the sense that the analysis is solely based on the evidence shown in the corpus data examined. It looks at every occurrence in the data and accumulatively arrives at a functional taxonomy. In other words, the functional categories developed in the analysis are derived from the recurrent patterns observed in the three datasets. Various linguistic, contextual and prosodic features outlined in the section above assist in the interpretation of the functions of discourse particles. The study is data-driven and it does not intend to explain the functions of discourse particles using a single existing theory or model. This does not mean, however, that the analysis does not take into consideration functions which have been proposed in earlier studies. After all, functions which have been identified in previous work may well appear in the data of the present study. If a particular function observed in the data conforms to what is reported in the literature, there is no reason why a new term should be devised. Likewise, functional categories formulated from the data might be explained with respect to well-developed concepts and theories such as face and politeness if they are deemed appropriate in the discussion. Therefore, while data take priority in the functional analysis, findings from previous research are constantly compared with those from the present investigation to see if functions identified in this study resemble the ones described in the literature or if they are functions which have yet to be discovered by other studies.

In taking a corpus-analytical approach, one must anticipate that there might be the chance of "lack of fit" between data which remain to be studied and categories which are developed based on data already examined. As remarked by Sinclair

(2004:9) on his experience in building the Cobuild dictionary: “I grossly underestimated the effect of the new information that the corpus supplied, and in particular the total lack of fit between the evidence coming from the corpus and the accepted categories of English lexicography”. As a result, the functional categories in this study had to be constantly adjusted in the process to fully capture the range of functions observed in the data. Revisions were especially vigorous in the early stage of the functional analysis, when the classification system was still evolving and rounds of modifications were required in order to account for the large number of examples examined. Eventually, an appropriate set of functional categories has been established for each particle which is presumed to be broad enough to accommodate all the “actual uses” (Kroon 1995:98) in the data but at the same time each function is specific enough for it to be distinct from the others.

Another methodological issue surrounding the functional analysis in the present study concerns multi-functional occurrences. Multi-functionality has been a well-acknowledged property of discourse particles and it has presented challenges for researchers working in the field who are interested in assigning examples to pre-defined functional categories. However, the occurrence of discourse particles doing more than one thing at the same time is by no means uncommon. In fact, this phenomenon abounds in many language units, which “merely manifest the multifunctional potential of human language resources” (Sinclair and Mauranen 2006:67). As pointed out by Aijmer (2002:28), “the multifunctionality of discourse particles does not seem to cause problems in communication because of the presence of linguistic and contextual clues functioning as interpreting strategies”. In the present study, when a discourse particle fulfils more than one function at a time, it is usually possible to determine the primary role it plays according to the various functional indicators available from the transcriptions and/or the recordings.

Given the inherent multi-functional property of discourse particles and the corpus-linguistic approach followed in this research, it is scarcely surprising that the present study makes no attempts to identify a core meaning for the two particles examined. In some previous work, especially those which are based upon a single pre-existing theoretical framework (see, for example, Blakemore 1988; Schiffrin 1987; Schourup 1985), there has been a working assumption that each discourse particle has a semantic core (see also Chapter 2). However, even in such studies a range of discourse uses are often suggested and it is only a matter of generalising the

commonalities between these uses by means of semantic extensions, metaphors and etymological relations. This frequently renders a core too abstract to account for all the instances found in authentic data (Müller 2005). As pointed out by Sinclair (1999:158), “the so-called core meanings are relatively infrequent and unimportant, and some hardly occur at all”. This is especially the case for the very frequent multi-functional words, of which discourse particles are a sub-group, “because few of them have a clear meaning independent of the co-text” (ibid.). Naturally, it follows that the issue of core meaning is irrelevant to the present discussion and, therefore, to discern the semantic core of the particles examined in this research is hardly necessary. More importantly, the identification and subsequent comparison of discourse functions in terms of various linguistic, sociolinguistic and contextual factors are deemed more fitting for the contrastive purposes of this study.

In order to facilitate the process of function determination, those instances which were irrelevant to the analysis of discourse functions were left out. This was achieved by means of data filtering. By applying the auto-filter function in Excel, a user could select and display only the records which suit specific criteria. Since all occurrences of the search words were coded for the feature “discourse particle status” in the previous procedure, instances which were not used as discourse particles could be excluded easily in the viewing of a spreadsheet database by using the filter function. Subsequently, only the concordance lines with the search words used as discourse particles were selected and shown for this functional analytic procedure. Apart from the functional analysis, the filtering function was also used in the following step of the study, when a number of factors were investigated in relation to the use of discourse particles.

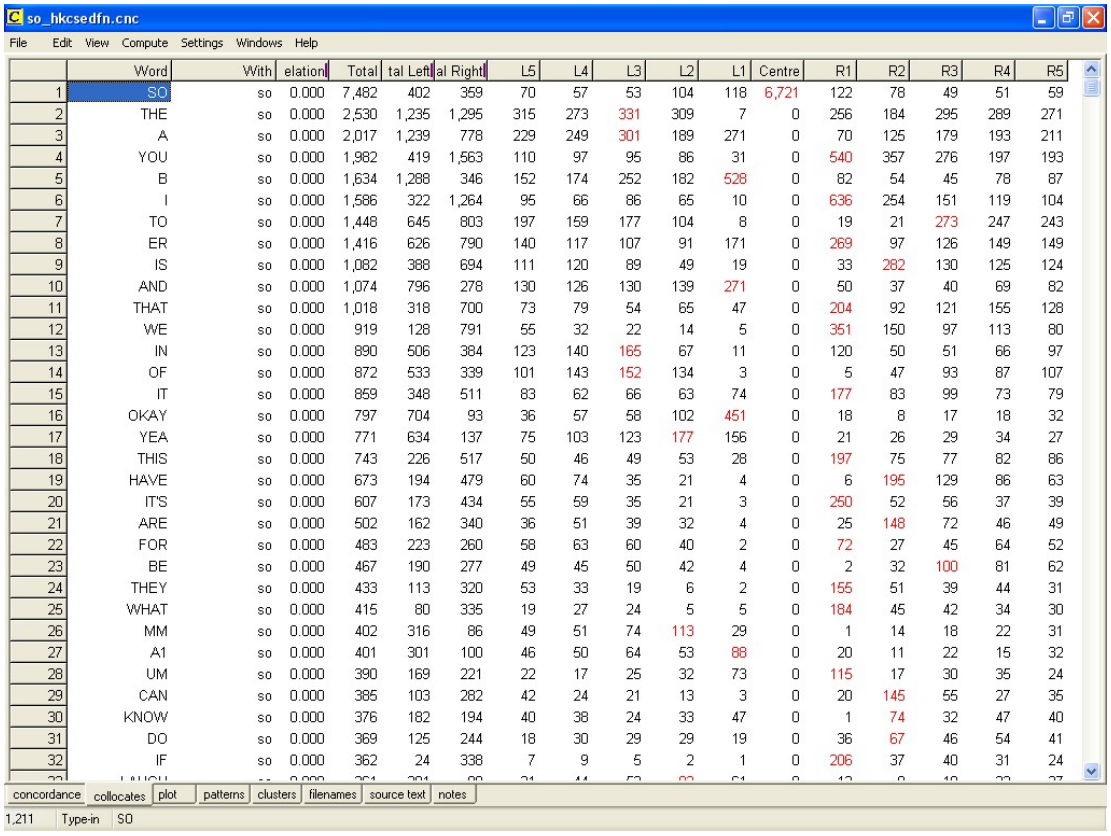
3.2.5 Data evaluation

Once all the occurrences of the two search words had been annotated according to the list of features discussed and the discourse functions interpreted, they were subject to data evaluation. This was carried out in two parts, with the help of two software packages, WordSmith Tools and Excel.

As discussed before, the lexical co-occurrences of the search words with other linguistic items could serve as useful clues to their functional interpretation. In addition to examining each concordance entry individually for the linguistic co-texts of discourse particles, the present study also made use of the Concord program in

WordSmith Tools to collectively investigate the lexical combinations of the particles *well* and *so* in the large number of instances found in the data. The vertical and horizontal views of concordance lines in Concord make it possible to carry out qualitative and quantitative analyses of the co-occurrences of discourse particles at the same time. In order to ensure that only the lexical patterns of particles were studied, occurrences which were not discourse uses were omitted. This was done by marking each propositional instance of the two words on the concordance file with a specific code in the "set" column in Concord and subsequently deleting them by performing the "zap" function. After the removal of these irrelevant entries, the lexical patterns of the two particles could be studied in two ways. Firstly, the concordance lines could be sorted by the alphabetical order of the words adjacent to the search word. This is especially useful for identifying the lexical items which frequently occur immediate to the left or right of the particles examined. By means of setting different sorting criteria, the lexical combinations of the two discourse particles could be visualized easily. Secondly, the frequency statistics of the words which occur in the neighbourhood of the particles could be obtained by consulting the "collocates" page in Concord. By specifying the horizons, a user could determine the size of the neighbouring linguistic context to be included in the collocation search. In the present study, a span of five words to each side of the search word was selected. Figure 3.4 shows a screenshot of the "collocates" page of *so* in Concord, with the most frequent position of a given collocate signalled in red:

Figure 3.4. A screenshot of the collocates of *so* in Concord



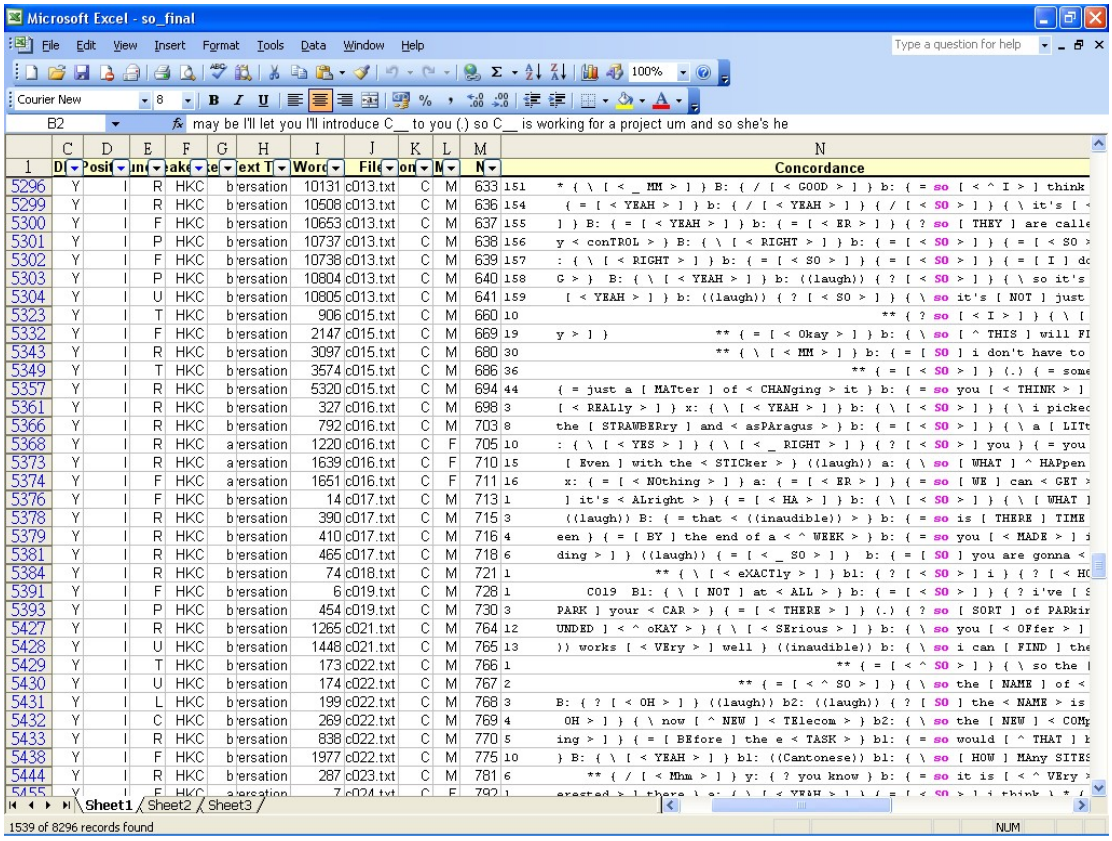
	Word	With	elation	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	SO	so	0.000	7,482	402	359	70	57	53	104	118	6,721	122	78	49	51	59
2	THE	so	0.000	2,530	1,235	1,295	315	273	331	309	7	0	256	184	295	289	271
3	A	so	0.000	2,017	1,239	778	229	249	301	189	271	0	70	125	179	193	211
4	YOU	so	0.000	1,982	419	1,563	110	97	95	86	31	0	540	357	276	197	193
5	B	so	0.000	1,634	1,288	346	152	174	252	182	528	0	82	54	45	78	87
6	I	so	0.000	1,586	322	1,264	95	66	86	65	10	0	636	254	151	119	104
7	TO	so	0.000	1,448	645	803	197	159	177	104	8	0	19	21	273	247	243
8	ER	so	0.000	1,416	626	790	140	117	107	91	171	0	269	97	126	149	149
9	IS	so	0.000	1,082	388	694	111	120	89	49	19	0	33	282	130	125	124
10	AND	so	0.000	1,074	796	278	130	126	130	139	271	0	50	37	40	69	82
11	THAT	so	0.000	1,018	318	700	73	79	54	65	47	0	204	92	121	155	128
12	WE	so	0.000	919	128	791	55	32	22	14	5	0	351	150	97	113	80
13	IN	so	0.000	890	506	384	123	140	165	67	11	0	120	50	51	66	97
14	OF	so	0.000	872	533	339	101	143	152	134	3	0	5	47	93	87	107
15	IT	so	0.000	859	348	511	83	62	66	63	74	0	177	83	99	73	79
16	OKAY	so	0.000	797	704	93	36	57	58	102	451	0	18	8	17	18	32
17	YEA	so	0.000	771	634	137	75	103	123	177	156	0	21	26	29	34	27
18	THIS	so	0.000	743	226	517	50	46	49	53	28	0	197	75	77	82	86
19	HAVE	so	0.000	673	194	479	60	74	35	21	4	0	6	195	129	86	63
20	IT'S	so	0.000	607	173	434	55	59	35	21	3	0	250	52	56	37	39
21	ARE	so	0.000	502	162	340	36	51	39	32	4	0	25	148	72	46	49
22	FOR	so	0.000	483	223	260	58	63	60	40	2	0	72	27	45	64	52
23	BE	so	0.000	467	190	277	49	45	50	42	4	0	2	32	100	81	62
24	THEY	so	0.000	433	113	320	53	33	19	6	2	0	155	51	39	44	31
25	WHAT	so	0.000	415	80	335	19	27	24	5	5	0	184	45	42	34	30
26	MM	so	0.000	402	316	86	49	51	74	113	29	0	1	14	18	22	31
27	A1	so	0.000	401	301	100	46	50	64	53	88	0	20	11	22	15	32
28	UM	so	0.000	390	169	221	22	17	25	32	73	0	115	17	30	35	24
29	CAN	so	0.000	385	103	282	42	24	21	13	3	0	20	145	55	27	35
30	KNOW	so	0.000	376	182	194	40	38	24	33	47	0	1	74	32	47	40
31	DO	so	0.000	369	125	244	18	30	29	29	19	0	36	67	46	54	41
32	IF	so	0.000	362	24	338	7	9	5	2	1	0	206	37	40	31	24

While the term ‘collocation’ is generally defined as the co-occurrences of lexical items in corpus linguistics (Sinclair 1991), it is quite commonly taken in studies of discourse particles and related items as a convenient label for any frequent word combinations observed. In the present study, collocation is used as a form of shorthand to refer to any frequent co-occurrences of the particles examined as observed from the collocation facilitates in Concord. It is worth noting, though, that at times the total number of the search word found on the “collocates” page may deviate from the actual number of concordance lines. This is because the “breaks” option in the Concord setting allows users to take various breaks into account during the computation of collocates. For example, if the setting is for the program to stop at sentence breaks, collocates will be counted only within the boundary of a sentence. Unfortunately, given the fact that the use of symbols and special layouts for representing simultaneous speech and other spoken features is common in transcriptions, none of the settings available from Concord provides statistics which are completely suitable for the present analysis. This means that manual checking is occasionally required by means of sorting the concordance lines and counting the

specific lexical pattern in question. As a result, figures displayed on the “collocates” page were merely treated as suggestive and were supplemented and further verified by evidence from the concordance lines.

Apart from lexical patterns, the present study also examines the interrelationship between other linguistic, sociolinguistic and contextual factors and the use of discourse particles. As discussed before, the auto-filter function was employed to investigate the relations between various features and the use of discourse particles. By using multiple filters, concordance entries of the search word which fulfil a certain set of criteria could be retrieved easily. For example, occurrences of *so* as a discourse particle in initial position produced by Hong Kong Chinese could be found quickly through the application of filter on the related features. This is illustrated in Figure 3.5, which displays a screenshot of part of the spreadsheet database of *so* in the HKCSE with the application of the filter function matching the criteria above:

Figure 3.5. A screenshot showing the spreadsheet database of *so* in the HKCSE with the application of multiple filters



At the bottom left corner of the window, the number of records satisfying the specified criteria is provided. This figure supplies a handy reference for frequency information of the various factors examined in the present study.

3.2.6 Statistical analysis

While comparisons were drawn through the examination of the spreadsheet databases during the course of data evaluation, some differences might be observed in the three datasets or between different components of the datasets under investigation. However, one could not be sure whether these differences are purely a consequence of chance. In order to provide statistical evidence for the quantitative differences found in the present study, results concerning frequency distribution of various factors were subject to a series of statistical analyses including statistical hypothesis testing and the determination of effect size.

In the investigation of statistical significance, the choice of which statistical test to employ is often not a straight forward issue. Each statistical hypothesis test has its own merits and pitfalls. Decisions regarding the selection of a particular test generally depend upon whether a normal distribution of the data is assumed as well as the scale of measurement of the data (Oakes 1998). Since the present study makes no assumption of the distribution of data, a nonparametric procedure which demands no distributional assumption is required. In statistics, a number of such procedures have been proposed, including the sign test, the Mann-Whitney U test, the median test, and the chi-squared (or chi-square) test, to name just a few of the popular options (see, for example, Butler 1985; Cramer 1998; Oakes 1998, for a detailed review of possible measures). Many of these tests, such as the Mann-Whitney U test and the median test, are used with non-categorical data which can be ranked (Cramer 1998). As variables in the present study involve nominal categories, a test which can work with frequency data on a nominal measurement scale is required. Being one of the most frequently employed tests of hypothesis not only in linguistic studies but also in social sciences which fit the above criteria (see Butler 1985; Healey 2007), the chi-squared test is chosen for the present purpose of statistical hypothesis testing.

As a statistical measure, the chi-squared test is well-known and widely used in corpus analysis. It has been used in various studies for the comparison of frequency distributions in different sociolinguistic groups or in two or more corpora (see, for example, Andersen 2001; Römer 2005). It is considered suitable for quantitative data

which mostly consist of nominal categories and for a large sample size (Oakes 1998; Siegel and Castellan 1988). In addition, it is generally considered a flexible test given its ability to handle variables which have more than two categories (Healey 2007). Therefore, it is deemed appropriate for the present study, which involves comparative analysis of a similar nature. By means of the chi-squared test, it is possible to determine whether differences found are statistically significant. In addition, the p-value generated from the test represents the significance level of the result, i.e. whether it is significant ($p < 0.05$), highly significant ($p < 0.01$) or very highly significant ($p < 0.001$).

Like any other statistical technique, however, statistical hypothesis tests are limited in the range of questions they can answer. The determination of effect size and the exact pattern of the relationship, for example, are two issues of interest to the present study which cannot be satisfactorily addressed by the calculations of chi-squared alone. Firstly, as discussed before, statistical hypothesis tests are only concerned with the likelihood of the observed difference to occur by chance. Therefore, the results of these tests do not indicate how large the effect is in the sample studied. In other words, although a significant chi-squared value indicates that there is *some* association between the variables, its magnitude does not reveal if the results are important in any other sense. To determine the effect size and hence to quantify the strength of association between variables, an index of association is employed in the present study. This is in response to the general observation that the inclusion of some index of strength of relationship together with information about statistical significance has been increasingly accepted as a recommended practice in reporting statistical findings (American Psychological Association 2001). As is the case with statistical hypothesis tests, a number of measures of association are available. Kirk (1996), for example, has discussed over 40 such measures, though only a minority of them are applicable to associations between nominal variables. In the present case, the statistic Cramér's V is employed because it is a commonly used chi-squared based measure of nominal association which works with data regardless of the size of the contingency table, unlike other alternatives such as phi (Cramer 1998). Its fixed range of values between zero and one also makes it an easier measure to interpret than T^2 and the contingency coefficient C (Healey 2007). Further, its symmetrical nature means that the strength of a relationship between two variables can be assessed without considering which variable is dependent and which

is independent. For the value of Cramér's V, the closer it is to one, the stronger is the relationship. Table 3.13 outlines the conventions for describing the magnitude of association used in the present study. Notice these descriptors and their division points are arbitrary and are only taken as general guidelines for interpreting the strength of the correlation in the present study[30]:

Table 3.13. The relationship between the value of Cramér's V and the strength of the association (adapted from Healey 2007:263)

Value of Cramér's V	**Descriptions**
Less than 0.10	Weak association
Between 0.10 and 0.30	Moderate association
Greater than 0.30	Strong association

As mentioned, the other issue which cannot be settled by chi-squared is concerned with the question of what exactly accounts for the difference found. Although the result of the chi-squared test shows the likelihood of a given difference to occur at random, it does not show "where in a contingency table the important discrepancies are" (Siegel and Castellan 1988:194). In other words, it is not possible to tell where the statistical difference is located in the data examined based on the result of the chi-squared test, which is arguably an important piece of information in itself. On some occasions, it is obvious that the statistical difference is largely due to a particular category in the data. As an illustration, Table 3.14 shows a simple example of the comparison of hair colour of the two sexes which exemplifies such a scenario:

Table 3.14. An example of a statistically different result owing to a particular category (reproduced from Zar 1999:503)

	Hair colour				
Sex	*Black*	*Brown*	*Blond*	*Red*	**Total**
Male	32	43	16	9	100
	(37%)	(40%)	(20%)	(36%)	
Female	55	65	64	16	200
	(63%)	(60%)	(80%)	(64%)	
Total	87	108	80	25	300

[30] A number of other crude descriptors for interpreting the magnitude of measures of association for nominal-level variables are available. For another example, see Rea and Parker (1992:203).

In the example above, the difference in hair colour between the two gender groups is found to be statistically significant. By examining the two-by-four contingency table, it can be deduced that the distribution of hair colour for the two groups is most different for the blond group. In other words, the statistical difference observed in the data is speculated to be mainly a result of the discrepancy observed in the blond category. However, the significant chi-squared for the contingency table above does not suggest which part of the table contributes most to this difference, not to mention supporting the speculation, which is probably of more interest to researchers.

In the present study, when a result was found to reach statistical difference and it was hypothesized by observing the result that the difference was largely due to a single category in the result, two chi-squared tests were performed with the data. One was calculated based on all the categories except the one which was suspected to mainly give rise to the statistical difference for the result. The other was calculated using the category which showed the most remarkable degree of discrepancy and all the other combined categories. To illustrate with the example just discussed concerning hair colour, Table 3.15 and Table 3.16 show respectively how the two chi-squared tests described above are performed based on the data in Table 3.14 in the present discussion:

Table 3.15. The 2 x 3 contingency table formed by omitting column 3 and keeping columns 1, 2, and 4 of Table 3.14 (adapted from Zar 1999:503)

	Hair colour			
Sex	*Black*	*Brown*	*Red*	**Total**
Male	32	43	9	84
Female	55	65	16	136
Total	87	108	25	220

Table 3.16. The 2 x 2 contingency table formed by keeping column 3 and combining columns 1, 2, and 4 of Table 3.14 (reproduced from Zar 1999:504)

	Hair colour		
Sex	*Blond*	*Nonblond*	**Total**
Male	16	84	100
Female	64	136	200
Total	80	220	300

If the result of the chi-squared test for Table 3.15 is not significant while a significant chi-squared is found for Table 3.16, the above speculation that blond chiefly accounts for the gender difference in hair colour can be confirmed. In statistics, this is called the sub-division of a contingency table. Sub-dividing chi-squared statistics allow researchers to see whether a significant difference for the overall table is largely a result of the difference only between certain categories and/or groups of categories. Although it is argued that this divided chi-squared analysis offers limited statistical value, it has been used in statistical analyses as a guide to test hypotheses concerning which part of the result contributes most to the statistical difference (see, for example, Zar 1999). As the above example demonstrates, this simple post hoc test has practical value in testing and specifically pointing out which category displays the greatest degree of divergence.

In the present study, the statistical package SPSS for Windows (version 15.0) and a python-based chi-squared program (Forsyth 2007) available in the public domain were used to facilitate the calculations in the statistical analysis. The chi-squared testing procedure and the determination of effect size were mainly performed through SPSS. This involved transferring the spreadsheets created in Excel to SPSS and saving them as databases which are readable in the SPSS environment. The data could then be processed by the software and analysed statistically. Apart from the computations of the chi-squared and Cramér's V values, SPSS was also used to generate tables and charts for the present study. For the results of divided chi-squared, the python-based chi-squared program was employed.

In applying statistical methods, one should bear in mind that hypothesis tests only give the user an estimation of how likely the claims made are wrong. The smaller the probability, the greater the confidence we have in the hypothesis made concerning the data. For the chi-squared test, statistical significance is easier to reach when the sample size increases (Oakes 1998). In other words, even small and non-notable differences can be found to be statistically significant if the sample is sufficiently large. Extra caution should therefore be taken to avoid an over-reliance on testing of statistical significance, especially in linguistic studies. After all, in natural language, words are rarely chosen at random, instead are picked based on the "idiom principle" (Sinclair 1991). Therefore, statistical significance should only be treated as further support for patterns observed in language, rather than the sole guidance for

identifying results which deserve to be reported. Just as one should not be too excited by indications of statistical significance, one should not ignore interesting findings which are not supported by statistical testing but are of practical importance from the linguistic point of view.

Summary

In this chapter, the three datasets used in this empirical study are first described in detail. Reasons for the selection of the three sources are given and the composition of each of them explained. The chapter then moves on to discuss the overall process of the research and outlines at length the procedures involved in each step with the application of various analytic instruments. In the discussion, some methodological issues encountered in the study are addressed.

Chapter 4

Study of *well*

Abstract

This chapter begins with a review of some previous work on *well* (4.1). The criteria distinguishing between the discourse use and the propositional use of *well* are then described, with illustrating examples of the different uses (4.2). This is followed by an in-depth analysis of *well* in the HKCSE, which consists of an examination of the distribution of *well* in terms of its overall frequency of use, positions, prosodic features and functions, and an investigation into the various possible factors which affect the use of *well*, including linguistic, sociolinguistic and contextual variables (4.3). The use of *well* in the BNC will then be described and compared with the findings in the HKCSE (4.4). The chapter will then move on to report on the use of *well* in the textbook database and make a comparison between the use of *well* in the authentic data and the teaching materials (4.5). To round off the chapter, a discussion of the pedagogical implications arising from the study of *well* will be provided (4.6).

4.1 Previous studies of *well*

Well has probably attracted more attention than any other discourse particles in linguistic research because it is a highly versatile word which occurs frequently in the spoken language. In the spoken part of the British National Corpus, *well* ranks 32nd in the frequency list, and its rounded frequency in speech is seven times higher than in writing (Leech et al. 2001:144). It is thus hardly surprising that there appears to be a lively and enduring interest in this linguistic item. This section reviews the treatment of *well* in the literature. In particular, it describes some major discourse functions of *well* which are discussed in previous studies. It also examines how the particle is analysed from different perspectives. Special attention is given to areas which have not been thoroughly examined in the field, including the prosody of *well* and the use of *well* in the speech of non-native speakers. As the present study is on discourse particles, the review only focuses on research work related to the pragmatic use of *well* but not its propositional use.

R. Lakoff's (1973) article on the use of *well* in question and answer pairs is often considered one of the earliest descriptions of the particle. In her attempt to examine

the conditions in which *well* could occur in responses, Lakoff points out that when *well* precedes an indirect answer, it indicates that the answer can only be deduced from the response given. On the other hand, when *well* is used to preface a direct response, the speaker is implying that "the rest of the answer is not to be taken as a complete reply giving all the information necessary" (1973:459). She concludes by saying that *well* as a particle is used in case of insufficient responses. Similarly, Owen (1981) looks at the use of *well* in answers to questions. She criticizes Lakoff's account by citing examples which do not fit in under the concept of insufficiency and calls for a corpus-based approach which relies less heavily on intuition. In her study, the use of *well* in second pair-parts is interpreted as one of the strategies to signal an impending face-threatening act. In other words, *well* reduces the subsequent loss of face, echoing Brown's (1977:117) view that the function of *well* is "to maintain (or establish) social relationships".

Compared with Lakoff (1973) and Owen (1981) who mainly focus on the responsive aspect of *well* in answers, the comprehensive paper by Svartvik (1980) discusses a number of discourse functions which are found in the London-Lund Corpus. In particular, he identifies the qualifier function and the frame function (see also Sinclair and Coulthard 1975) as the two major uses of *well*. As a qualifier, not only can *well* mark incomplete or indirect answers to questions as suggested by Lakoff (1973), it can also preface replies to statements which convey the sense of incompletion. As a frame, *well* acts as a structural device to facilitate the textual organisation of talk. Apart from these two key functions, Svartvik (1980) suggests that *well* also serves other additional functions associated with discourse management such as floor-holding and hesitation. Despite the many functions of *well* he found in the corpus, Svartvik (1980:176) concludes that in fact the particle "displays a much wider range of functions than those discussed" in his paper.

While Svartvik's (1980) study highlights the multi-functional nature of *well*, other studies attempt to search for a single meaning of the item. Schourup, for example, describes *well* as "primarily an evincive indicating consultation by the speaker of his or her current thoughts" (1985:64). He argues that all the uses of *well* can be subsumed under this single concept of "internal consultation" and thus his approach has the advantage of "being simpler and more comprehensive than existing treatments" of *well* (1985:65), which examine the particle from a functional perspective. His characterisation of *well* remains similar in a paper published more

than a decade later, where he maintains that *well* signals that "the speaker is actively considering whatever it is relevant to consider in determining what should now follow" (2001:1058). However, many of the examples in his studies (1985; 2001) are invented with minimal contextual information provided, leading to the question about the extent to which his conclusion can be applied to instances in naturally-occurring data.

For Jucker (1993) and Blakemore (2002), the core meaning of *well* can be formulated using Sperber and Wilson's relevance-theoretical approach (1995). In Jucker (1993:450), *well* signifies that "the context created by an utterance may not be the most relevant one for the interpretation of the next utterance". In other words, there is something wrong in the discourse which might increase the processing effort. In Blakemore (2002), however, *well* indicates that the message conveyed by the utterance it contains is consistent with the principle of relevance. In other words, the information preceding *well* is relevant to the interpretation of the information after it, meaning things are going smoothly in talk. This is in accordance with Carlson's (1984) view that *well* indicates acceptance of a move in dialogue games. Although Jucker (1993) claims that relevance theory is the only theory that can give a satisfactory explanation of all the discourse uses of *well*, the fact that the theory renders two opposite interpretations of the particle and that the two studies discuss mainly data from other studies to support their assertions casts doubt on the robustness of the theory in giving an integrated account of *well*.

Similar to the above studies which follow a particular framework, Schiffrin (1987) examines the role of *well* within coherence-based theory. Based on authentic data from her interviews, Schiffrin (1987:103) concludes that *well* contributes to conversational coherence by signalling to the hearers that the upcoming discourse is not "fully consonant with prior coherence options". By extending her idea of conversational coherence from pairwise interactions to embedded discourse, she discusses the use of *well* not only in question/answer pairs and requests but also in self-responses. Although Schiffrin (1987) and Jucker (1993) make use of two different theories in their studies, their conclusions are similar in the sense that both view *well* as a signal of some sort of inconsistency in the neighbouring context, either under the notion of coherence or relevance.

Compared with some studies which argue for a single meaning of *well* with the support of mainly invented examples, Schiffrin's (1987) study has the merits of

analysing instances from real-life data. However, since her description of *well* is solely based on the sociolinguistic interviews she conducted with seven informants in a Philadelphia neighbourhood, only the uses of *well* in one text type by a rather limited number of respondents are being examined. Other studies which are confined to a particular genre include the discussion of *well* in a family gathering in Watts (1989) as well as commentaries on a snooker game in Greasley (1994). Although these studies provide useful findings as regards the use of *well* in specific contexts, the analysis of merely one text type prevents these studies from investigating the possible effect of text type on the functions of *well*, and thus they may overlook some functions which are not present in the text type under discussion.

From a contrastive perspective, the particle *well* could be looked at cross-linguistically to examine the similarities and differences in meaning with its translation equivalents in different languages. Previous research work in this area includes the comparison of *well* and its Italian equivalent *bene* (Bazzanella and Morra 2000), *well* and its Swedish equivalents (Aijmer 2003), *well* and its correspondences in Norwegian and German (Johansson 2006), and *well* and its close parallel *bueno* and *pues* in Spanish (García Vizcaíno and Martínez-Cabeza 2005; Stenström 2006). These studies all suggest that there are functional differences between *well* and its semantic cognate in another language. As demonstrated in Aijmer and Simon-Vandenbergen (2003), a comparative approach using translation corpora allows for a better understanding of the particle. The wide range of Swedish and Dutch translations of *well* in the paper gives supporting evidence for its uses as a connective and as a pragmatic marker with interpersonal function, thus highlighting the multi-functional nature of *well*.

As with the study of prosody of particles in general, there appears to be little work done on the prosodic profile of *well*. Svartvik's (1980) detailed account of the particle looks at the position of *well* in the utterance as well as the distribution of tones on *well* as the nucleus. His findings show that *well* as a discourse particle is most likely to carry a falling tone when receiving primary stress. However, no significant correlations are found between prosody and the meaning of *well* in his study. In the book *Intonation and Its Uses*, Bolinger (1989) devotes a whole chapter to *well* and presents a variety of conversational uses of the particle with reference to its intonational features. Although he suggests that different uses of *well* may be associated with their respective prosodic patterns and gives an example of the

difference in meaning between “full” *well* and its “fully reduced” form *w’l* (Bolinger 1989:333), his study does not provide any quantitative results to support his argument. Similarly, Local and Kelly (1986) observe a number of phonetic shapes of *well* from 34 instances produced by one speaker and suggest that its use in signalling reported speech and repair has distinctive phonetic patterns. As far as the prosodic phrasing of *well* is concerned, Hirschberg and Litman (1993) find that 14 out of 27 instances of *well* as a discourse particle are in a separate tone unit, giving support to the result reported in Altenberg (1987) that half of the tokens of *well* have a potential of being the nucleus in a text from the London-Lund Corpus. These studies provide valuable findings as regards the role of intonation in the analysis of *well* which has rarely been discussed in the literature, and serve as a starting point for establishing the prosodic pattern of the particle after a large number of *well* is examined.

In terms of the analysis of *well* in the data of non-native speakers of English, the small number of studies in the literature reveals disparity in the usage between native and non-native speakers. Cheng and Warren (2000), for example, find that there are differences between native speakers of English and Hong Kong Chinese in terms of the functions of *well* and the discourse types in which they employ the particle. Specifically, the Hong Kong Chinese speakers of English use *well* more often to signal repairs than native speakers. In addition, there is an overall higher frequency of the discourse use of *well* in the speech of native speakers. In the interviews and conversations examined, native speakers use *well* at least twice as often as the Hong Kong Chinese (Cheng and Warren 2000). Elsewhere, Müller (2005) discusses the difference in frequencies of the individual discourse functions of *well* by German speakers and American speakers. Her findings show that the Germans in her study use *well* more frequently to search for words or expressions and to introduce indirect answers. In addition, the uses of *well* to make a conclusion and to mark the continuation of their opinion are only present in the speech of German speakers. The two studies above similarly point out that the non-native speakers in question might encounter linguistic difficulties, leading to a higher tendency of *well* in remedial functions such as repair and word search. The comparison of *well* in native and non-native speech is also the focus in de Klerk (2005). She states that using *well* to indicate mental consultation is the most common use in the speech of Xhosa speakers. By looking at the various functions of *well* in Xhosa English, she concludes that different uses are observed in Xhosa English and other native varieties of English,

though her statistical distribution of functions is only based on the corpus of Xhosa English and no comparative quantitative results from a corpus of native speaker speech are provided to substantiate her claim for her functional analysis.

From the examination of the previous studies of *well* above, it can be seen that *well* is a multi-faceted particle which displays a wide range of uses, irrespective of whether a top-down or bottom-up approach to the meaning of *well* is taken. Much of the research described suggests that the functions of *well* are conditioned by various factors, such as the discourse event in which it occurs, its intonational profile, the linguistic background of the speakers and so on. It is thus imperative to examine *well* from different perspectives, taking into account all the variables which may influence its use.

With the aim of exploring the complex relationships discussed above between functions and factors, an analysis of *well* will be presented in the remaining parts of this chapter. The next section will describe how the discourse use and propositional use of *well* are distinguished in the present study. This is followed by a discussion of findings from the three datasets. The interrelationship between the discourse functions of *well* and the linguistic, sociolinguistic and contextual variables in the data examined will be elucidated and compared with relevant results from previous studies wherever applicable.

4.2 Distinguishing the discourse use and propositional use of *well*

The word *well* has been given various labels in grammar books and dictionaries. Many of these uses are non-discourse uses, including its use as the adverbial equivalent of *good*, as a "health adjective" (Leech and Svartvik 2002:232), as an emphasizer, as an intensifier and in fixed phrases. When *well* is used as an adverb corresponding to the adjective *good*, for example, it cannot be omitted without altering the propositional meaning of the utterance, as in this example:

(1)

b: …{ = [< YOU >] } { = [DID] < NOT > } { / [DO] < ***WELL*** > } { \ during the [WORKshop] < TRAIning > }…

(HKCSE, A008, 2219)

Similarly, when it is used to mean in good health, it is within the syntactic structure of the utterance and thus is not syntactically optional, as in:

(2)

B: { \ [NOT] feeling < ***WELL*** > } { \ [< NO >] } { \ [< _ Mhm >] }

(HKCSE, C040, 3888)

When *well* follows modal verbs such as *can*, *could*, *may* and *might*, it conveys the meaning of probability (Quirk et al. 1985). Thus it contributes to the propositional content of the utterance, as in example 3:

(3)

B: …{ \/ it [^ MIGHT] ***well*** BE that ONCE the machine has been set < UP > }…

(HKCSE, A006, 2478)

Likewise, the use of *well* which expresses a high level of standard or a great extent contributes to the propositional meaning of the utterance, regardless of whether it acts as an intensifier modifying another adverb or becomes an adjective when it is in combination with a past participle, as in this example:

(4)

a: …{ = [I] FEEL that I have < TO > } { \/ be [***WELL***] < prePARED > } { = [< FOR >] the interVIEW }…

(HKCSE, B063, 1192)

The use of *well* in phrases such as *well and truly*, *as well* and *as well as* again are not syntactically optional:

(5)

a: …{ \ and [THEN] they do WIDE old BINding as < ***WELL*** > } { \ it's [^ NOT] the < PLAStic > one }

(HKCSE, B113, 3320)

The use of *well* as a discourse particle, on the other hand, is syntactically detachable from an utterance and can be removed without changing its propositional content, as in the present case:

(6)

b: { / [< ***WELL*** >] } { = you [< WANT >] } { = to have a [< ^ MOnitor >] } { = [HAVE] a REsolution < OF > } { = [< ONE >] }

(HKCSE, A009, 2215)

While the following sections will only focus on the functions of *well* as a discourse particle, the propositional uses will also be discussed whenever appropriate to make comparisons. In fact, as will become apparent in the later part of the chapter, the propositional uses and discourse uses of the word *well* show contrastive patterns in relation to various linguistic, sociolinguistic and contextual variables.

4.3 *Well* in the HKCSE

This section describes in detail the uses of *well* in the HKCSE. It first presents an overview of the frequency of *well* in the corpus and moves on to discuss the positional distribution and prosody of *well*. The pragmatic functions of *well* identified in the data are then discussed at length, followed by an investigation into the various possible factors which may correlate with the discourse functions of *well*.

4.3.1 Frequency of occurrence

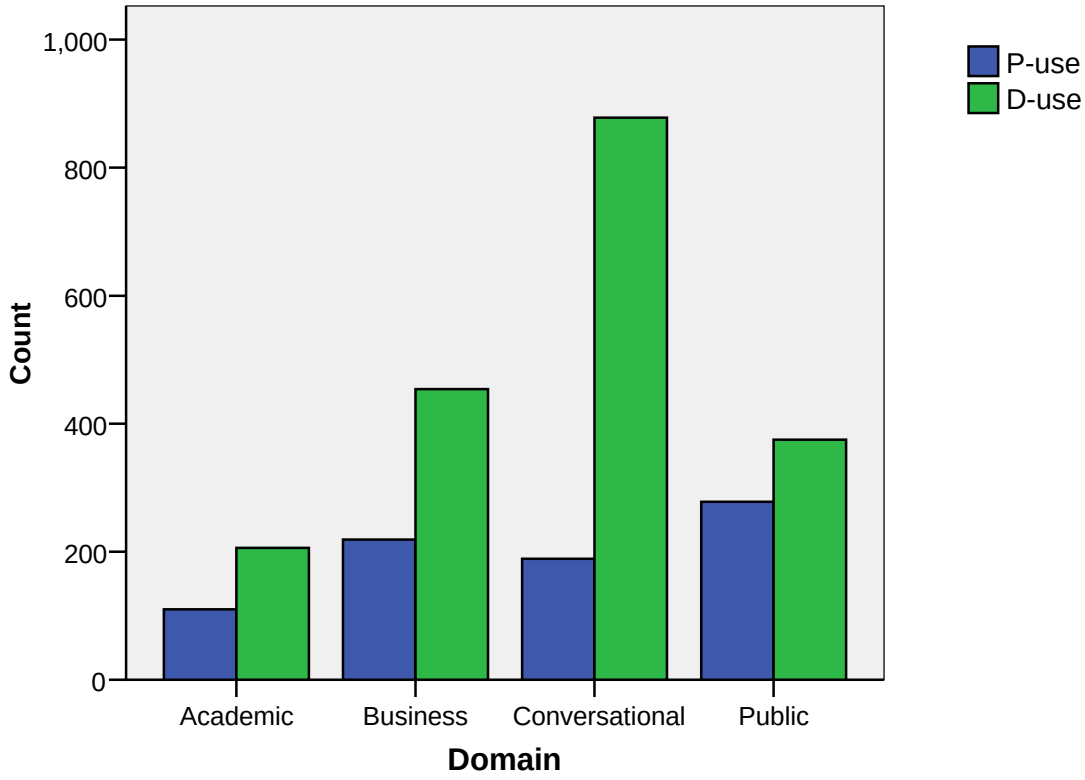
In the HKCSE which contains approximately 950,000 words in 311 texts, there are 2,714 instances of *well*, of which 1,913 are discourse uses (D-use). Propositional uses (P-use) amount to 796 instances, leaving five examples of *well* unclassified owing to insufficient contextual information. The distribution of the uses of *well* in the HKCSE is presented as follows:

Table 4.1. The distribution of *well* in the HKCSE

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Total number of words in HKCSE	949,972 (100.0%)	213,204 (22.4%)	259,484 (27.3%)	258,882 (27.3%)	218,402 (23.0%)
Total number of *well*	2,714 (100.0%)	316 (11.6%)	674 (24.8%)	1,071 (39.5%)	653 (24.1%)
Number of D-uses	1,913 (100.0%)	206 (10.8%)	454 (23.7%)	878 (45.9%)	375 (19.6%)
Number of P-uses	796 (100.0%)	110 (13.8%)	219 (27.5%)	189 (23.7%)	278 (34.9%)
Unclassified uses	5 (100.0%)	0 (0.0%)	1 (20.0%)	4 (80.0%)	0 (0.0%)
D-use / total use (%)	67.99	65.19	67.36	81.98	57.43
D-rate (per 10,000 words)	19.56	9.66	17.50	33.92	17.17

As shown in Table 4.1, the conversational sub-corpus contains more than 1,000 instances of *well*, contributing the highest number of instances of the word in the corpus (39.5%). The lowest number of *well* is found in the academic domain, with only less than one-third of occurrences when compared with the conversations (11.6%). Similarly, the highest discourse use of *well* is found in conversations (45.9%) and the lowest in the academic texts (10.8%). The discourse-function ratio of *well*, as calculated by dividing the number of discourse use by the total number of use of the word expressed in per cent, is a useful measure for investigating how typical the lexical item is used as a discourse particle. The overall discourse-function ratio of *well* in the HKCSE is 68%, meaning more than two-thirds of all instances of *well* in the corpus serve discourse functions. Again, there are variations in the ratio in the four domains, which is better captured graphically in Figure 4.1:

Figure 4.1. The distribution of *well* over the four domains in the HKCSE – D-use versus P-use



The comparison between the lengths of the two bars for each sub-corpus in Figure 4.1 shows that conversations have the highest discourse-function ratio in the corpus.

As presented in Table 4.1, more than 80% occurrences of *well* found in the conversations are used as a discourse particle. While the business and public sub-corpora contain approximately the same total number of *well* (n=674 vs. n=653), the business texts have a higher discourse-function ratio, as indicated by Figure 4.1. In the public sub-corpus, only slightly more than half of the instances of *well* are discourse uses. The differences in the discourse-function ratio between the four domains achieve very high statistical significance with a moderate association ($x^2 \geq 132.498$; d.f. = 3; $p < 0.0005$; $V = 0.221$), suggesting that the formality of the setting may have an influence on the rate at which *well* is used as a particle. In informal conversations where the main goal of interaction is the maintenance of relationships, the ratio of discourse use to propositional use of *well* is the highest. In the formal public events where texts are less spontaneous and more likely to be pre-planned, *well* is less likely to be a discourse particle. This is consistent with findings reported in other studies that there is an association between informal conversation and the use of discourse particles (see, for example, Biber 1988; Östman 1982).

The effect of context on the ratio of discourse use also appears to be supported by a comparison with the discourse-function ratio found in the London-Lund corpus. The higher proportion of discourse use of *well* (86%) in the London-Lund Corpus as reported in Stenström (1990) when compared with that in the HKCSE (68%) might be a result of the respective composition of the two corpora. In the London-Lund Corpus, about half (34 out of 87 texts) of the texts are informal face-to-face conversation. In the HKCSE, on the other hand, the conversational sub-corpus consists of just about a quarter of the total. Apart from the effect of formality of texts in the corpus which may result in a difference in the ratio of discourse use, other factors such as the linguistic background of the speakers in the two corpora could also be at work. While the London-Lund corpus consists of texts produced by native speakers of English, the HKCSE is an intercultural corpus containing speech of both native and non-native speakers. The lower discourse-function ratio in the Hong Kong corpus when compared with a native English corpus confirms findings from previous studies that *well* is less frequent in the speech of non-native speakers (cf. Cheng and Warren 2000).

Another indicator of how typical *well* serves as a discourse particle is its frequency rate in the whole corpus. The discourse rate (D-rate) of *well* indicates the number of discourse use of *well* in a sample of 10,000 words. While the discourse-

function ratio measures how likely the word will be used as a discourse particle when it occurs, the D-rate provides an estimate of the number of discourse uses in a given sample of words. As shown in Table 4.1, the overall discourse rate in the HKCSE is roughly 20 instances per 10,000 words. Again, the highest discourse rate of *well* could be found in the conversational sub-corpus, giving further evidence to the high occurrences of discourse particles in informal spontaneous talk.

4.3.2 Position (utterance and tone unit)

4.3.2.1 Position in utterance

In the HKCSE, *well* occurs in a range of positions in the utterance. The positional distribution of *well* in an utterance in the HKCSE is shown in Table 4.2:

Table 4.2. The positional distribution of *well* in an utterance in the HKCSE[31]

Position	**Total number of *well***	**D-use *well***	**P-use *well***
Final	137 (5.1%)	14 (0.7%)	123 (15.5%)
Initial	1,112 (41.0%)	1,107 (57.9%)	5 (0.6%)
Stand-alone	65 (2.4%)	64 (3.3%)	1 (0.1%)
Medial	1,395 (51.5%)	728 (38.1%)	667 (83.8%)
Total	2,709 (100%)	1,913 (100%)	796 (100%)

Table 4.2 shows that most instances of *well* occur in initial (41%) and medial positions (51.5%). When the discourse use and propositional use of *well* are considered separately, it can be observed that there are marked positional differences. As a discourse particle, about 58% of the instances of *well* occupy utterance initial position, as in:

(7)

A: { \ [***WELL***] you're really < BUsy > }…

(HKCSE, A041, 3169)

This is in contrast to the percentage of propositional use of *well* occurring at the beginning of a turn, which is a mere 0.6%. In fact, the five instances of initial *well* and one single instance of stand-alone *well* which are categorised as P-use are

[31] The five unclassified uses of *well* are not included in the quantitative analysis from this point onwards.

examples where speakers are discussing the word *well* itself in a meta-linguistic sense. In the example below, the student is citing *well* as an example of interjections when the teacher asks her what interjections are in a lecture on linguistics:

(8)

a1: { / [AND] < interJECtions > } { / everybody [KNOWS] what interJECtions < IS > } (.) { \ [< _ INterJECtions >] }

a2: ((inaudible)) { \ [< FEElings >] }

a1: { / [< eXAMples >] }

a2: { = [< UM >] }

a1: { = [< UM >] }

a2: { = [UM] < ER > }

a1: ((laugh)) { = [CAN] you GIVE me < SOMEthing > } { \ [MORE] < CONcrete > }

a7: ((laugh)) * { \ [< OH >] }

a1: ** { \ oh }

a7: { / [< ***WELL*** >] }

a1: { / [< ***WELL*** >] } { / [< YES >] } { / [< goodBYE >] }…

(HKCSE, A001, 5522; 5524)

Excluding these special cases of *well* where the word is used in a meta-linguistic sense, the utterance initial position and the stand-alone position are locations in which the propositional use of *well* rarely, if ever, occurs. The medial position, on the other hand, is the place where there is an overwhelming proportion of P-use *well*, as in the following example:

(9)

a: …{ \ here is the [BAGgage] < TAGS > } { \ and the airport [TAX] < reCEIPT > } { = [< _ ER >] } { = [< FOR >] } { \ one [< HUNdred >] } { \ and your [PASSport] as < ***WELL*** > } { = [< AND >] er } { = please [WAIT] for a < ^ WHILE > }…

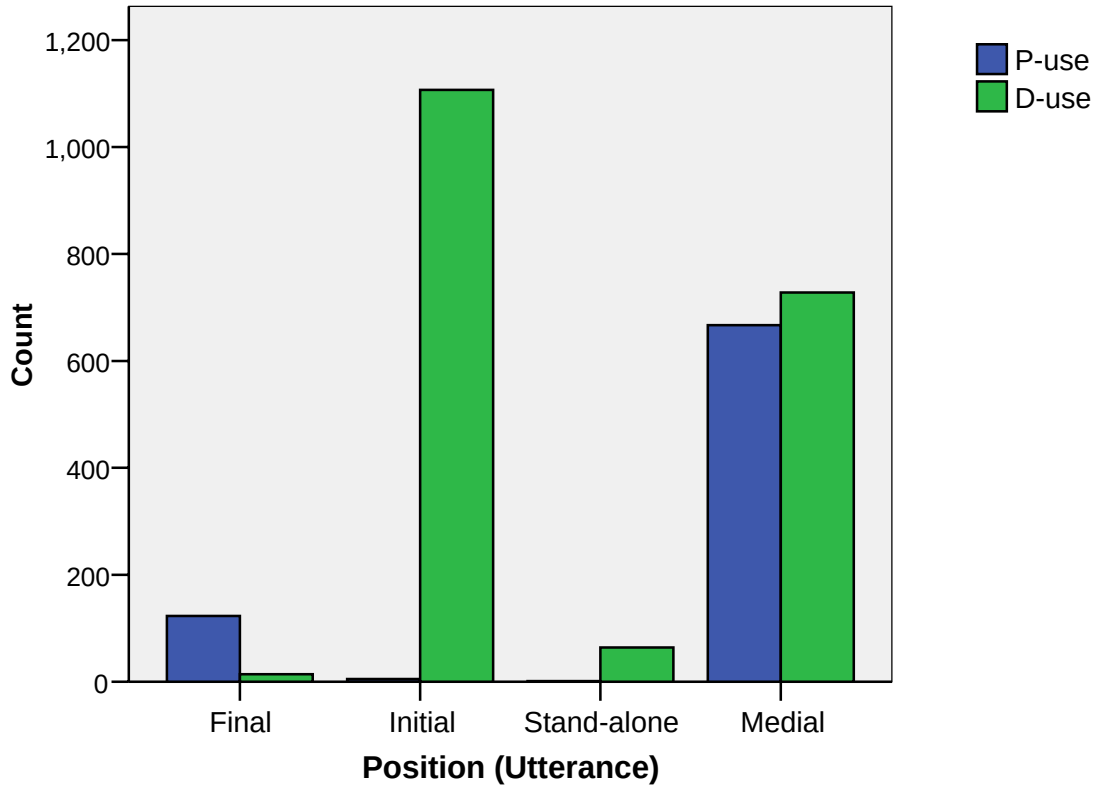
(HKCSE, B043, 154)

As a particle frequently associated with responses in previous studies, the D-use of *well* is often considered to occur in utterance initial position. In fact, some studies only focus on the use of *well* in front position, looking at those instances appearing at

or very near the beginning of turns (see, for example, Lakoff 1973; Owen 1981). However, an examination of the position of *well* in the utterance in the present study shows the positional flexibility of *well*. Although 57.9% of the instances of D-use *well* occupy front position, there are 38.1% of D-use *well* which occur in medial position, indicating that *well* does not only occur at the beginning of an utterance as a discourse particle, but it also appears quite frequently in medial position. Although some instances of D-use *well* are found in end position, the proportion is considerably lower than that in P-use *well* (0.7% vs. 15.5%), showing that the propositional use of *well* has a much higher preference for utterance final position.

The difference in the positional distribution between D-use and P-use *well* is statistically very highly significant with a strong association ($x^2 \geq 942.150$; d.f. = 3; $p < 0.0005$; V = 0.590), indicating a highly contrastive positional difference as regards whether *well* is used as a discourse particle. This contrast is represented as a chart in Figure 4.2 below:

Figure 4.2. The positional distribution of *well* in an utterance in the HKCSE – D-use versus P-use



4.3.2.2 Position in tone unit

Similar to the distribution shown in the position of an utterance, the D-use and P-use of *well* also display noticeable contrastive patterns in terms of their positional preference in a tone unit. The positional distribution of *well* in a tone unit in the HKCSE is shown in Table 4.3:

Table 4.3. The positional distribution of *well* in a tone unit in the HKCSE[32]

Position	**Total number of *well***	**D-use *well***	**P-use *well***
Head	340 (12.9%)	142 (7.7%)	198 (25.1%)
Pre-head	527 (20.0%)	504 (27.3%)	23 (2.9%)
Tail	103 (3.9%)	50 (2.7%)	53 (6.7%)
Tonic syllable (in a shared tone unit)	680 (25.8%)	190 (10.3%)	490 (62.2%)
Separate tone unit	985 (37.4%)	961 (52.0%)	24 (3.0%)
Total	2,635 (100%)	1,847 (100%)	788 (100%)

Of the 2,635 instances of *well* in the HKCSE for which prosodic information is available, more than one-third (37.4%) of all instances of *well* constitute a separate tone unit. In fact, out of the 985 tokens of *well* as a separate tone unit, a vast majority (n=961) of them are used as a discourse particle. This strongly suggests that the high level of prosodic independence is only applicable to the discourse use of *well* (see also Altenberg 1987). It is also noteworthy that the proportion of D-use tokens of *well* in a tone unit of its own found in the present study (52%) is exactly identical to the ratio (14 out of 27) reported in Hirschberg and Litman (1993) based on 27 instances of D-use *well*. Similarly, most of the uses of *well* (504 out of 527) found in the pre-head position are D-use, again indicating the syntactic looseness of the particle. An example of D-use *well* in the pre-head position is shown as follows, where *well* precedes the head *have to* and the tonic syllable *check*:

(10)

B: { \ [< Okay >] } { \ ***well*** we'll [HAVE] to < CHECK > }…

(HKCSE, B059, 1157)

[32] 19 instances of *well* are excluded in the prosodic analysis altogether due to the lack of prosodic information. 55 instances of *well* are excluded in Table 4.3 as the prosodic patterns of the tone units containing them are indeterminable. Therefore, their positions in the tone unit are unclassifiable.

Interestingly, when *well* is the tonic syllable but occurs together with other elements in the tone unit, it is more likely to convey propositional meaning. As a tonic syllable in a tone unit containing other items, 490 out of 680 are propositional uses. In the following example, the P-use of *well* is the tonic syllable inside the tone unit *as well as*:

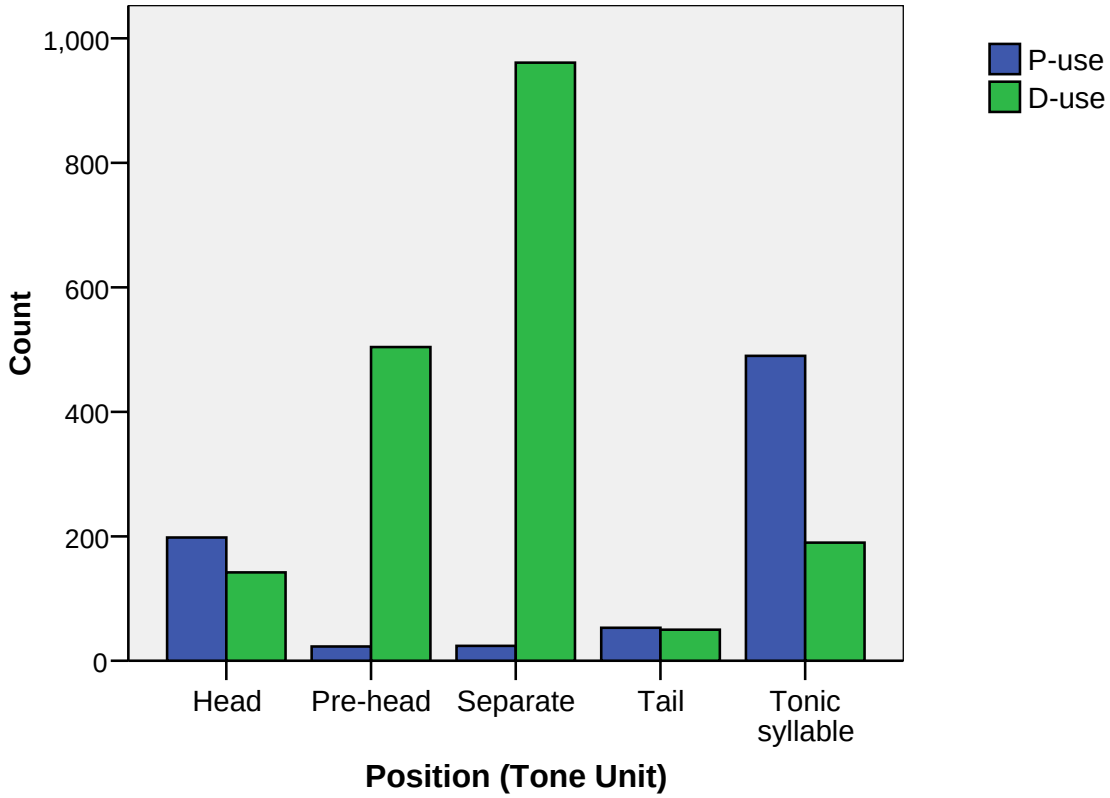
(11)

b: { = in the [PEARL] river < DELta > } { = as [< ***WELL*** >] as } { \ [THE] er < inCREASE > }…

(HKCSE, B099, 2277)

It is shown from the statistics in Table 4.3 above that the D-use and P-use of *well* display very different positional patterns in a tone unit. The difference again is statistically very highly significant with a strong association ($x^2 \geq 1247.985$; d.f. = 4; $p < 0.0005$; V = 0.688). Figure 4.3 shows a graphical distribution of the difference:

Figure 4.3. The positional distribution of *well* in a tone unit in the HKCSE – D-use versus P-use



While almost 80% of the discourse use of *well* are in a separate tone unit (52.0%) or in the pre-head position (27.3%), nearly 90% of the propositional use of *well* are either the tonic syllable inside a tone unit with other elements (62.2%) or in the head position (25.1%). The above findings provide empirical evidence that apart from the position in an utterance, the position in a tone unit is a crucial clue in determining the status of discourse use of *well*. In addition, they provide further argument for the syntactical detachedness and optionality of discourse particles (Altenberg 1987).

4.3.3 The prosody of well

Closely related to the position in the tone unit is the prominence pattern of the word. Of all the instances of *well* in the corpus, 26.7% are unstressed, leaving 73.3% with stress. The distribution of the prominence pattern of *well* in the HKCSE is presented in Table 4.4:

Table 4.4. The prominence pattern of *well* in the HKCSE

Prominence pattern	**Total number of *well***	**D-use *well***	**P-use *well***
Unstressed	703 (26.7%)	584 (31.6%)	119 (15.1%)
First	231 (8.8%)	111 (6.0%)	120 (15.2%)
Middle	36 (1.4%)	1 (0.1%)	35 (4.4%)
Last	305 (11.6%)	40 (2.2%)	265 (33.6%)
Sole prominent syllable (in a shared tone unit)	375 (14.2%)	150 (8.1%)	225 (28.6%)
Sole prominent syllable (in a separate tone unit)	985 (37.4%)	961 (52.0%)	24 (3.0%)
Total	2,635 (100%)	1,847 (100%)	788 (100%)

Of the total number of *well* which carries no stress, 584 out of 703 are discourse use. This indicates that when unstressed, *well* is more likely to be a discourse particle. In fact, this is the preferred stress pattern for D-use *well* when it shares the tone unit with other elements. In a shared tone unit, the discourse use of *well* rarely receives the primary stress, i.e. as the last prominent syllable, when there are other stressed syllables in the tone unit (2.2%). This is in sharp contrast to the propositional use of *well*, where about one-third (33.6%) of the instances are the nucleus when there are other stressed syllables. The occurrences of P-use *well* being the first stressed

syllable or the stressed syllable in the middle of a tone unit are also more frequent. When the word *well* is the sole prominent syllable in a tone unit, whether it is integrated with other items appears to correlate with its use. When it shares the tone unit with other elements, the word *well* as the only stressed syllable is more likely to express propositional meaning, as in example 12:

(12)

b: …{ = you [WORK] < HARD > } { = at the [END] of the < DAY > } { \ you will score [< ^ ***WELL*** >] in this subject }…

(HKCSE, A003, 944)

Conversely, the instances of D-use *well* make up 98% (961 out of 985) of all the instances of being the only element in a separate tone unit, thus also the prominent syllable, as in example 13:

(13)

b: …{ so [LET'S] talk about our < ^ FLOW > } { = [< ***WELL*** >] } { = hong [< KONG >] is }…

(HKCSE, B154, 2695)

When compared with the proportion of D-use *well* which are stressed in Svartvik's (1980) study, the ratio of stressed D-use *well* in the HKCSE is higher. In his examination of 303 examples of D-use *well*, Svartvik (1980) found that 56% of the instances of D-use *well* were stressed, compared with 68% in the HKCSE. However, it is not clear how many of the stressed instances constitute a separate tone unit in his study. In Altenberg (1987), the percentage of stressed D-use *well* is 70%, which is very similar to the result found in the present study. However, the findings in Altenberg (1987) are only based on twelve examples from a prepared monologue in the London-Lund Corpus. From the small number of research studies available, it appears that the particle *well* has a higher tendency of being stressed than unstressed, though it can be seen from the above analysis that the prosodic profile is complicated and depends much on whether *well* is a separate tone unit. Admittedly, more data from different sources are needed before conclusive statements can be made. Nevertheless, the present study provides some suggestive findings of the prominence pattern of *well* from a large amount of data which could be readily compared with future studies offering quantitative prosodic analysis of *well*.

Apart from the stress pattern of *well*, the tone is also an area of interest in prosodic analysis. In the corpus, there are altogether 1,665 stressed examples of *well*. The tonal distribution of *well* in the HKCSE is presented in Table 4.5:

Table 4.5. The distribution of nuclear tones on *well* in the HKCSE

Type of tone	**Total number of *well***	**D-use *well***	**P-use *well***
Fall	844 (50.7%)	543 (47.2%)	301 (58.6%)
Fall-rise	44 (2.6%)	21 (1.8%)	23 (4.5%)
Level	555 (33.3%)	427 (37.1%)	128 (24.9%)
Rise	174 (10.5%)	119 (10.3%)	55 (10.7%)
Rise-fall	0 (0.0%)	0 (0.0%)	0 (0.0%)
Unclassified	48 (2.9%)	41 (3.6%)	7 (1.4%)
Total	1,665 (100%)	1,151 (100%)	514 (100%)

Compared with the position and the stress pattern of *well*, the tone on *well* does not reveal much information about the status of *well* as a discourse particle. For both D-use and P-use, falling tone is the most frequent, followed by level tone. The use of falling and level tones amount to more than 80% of the total for both the discourse and propositional use of *well*, with a slight preference for falling tone rather than level tone for the P-use. The fact that a lower proportion of falling tone is found with D-use *well* than with P-use *well* can be seen as a result of the higher number of level tones with D-use *well*. In general, the profiles of tone on D-use and P-use of *well* are fairly similar.

In Svartvik (1980), only the tonal distribution of discourse use of *well* was analysed. His findings are compared with those in the HKCSE in Table 4.6:

Table 4.6. The comparison of tonal distribution on D-use *well* in Svartvik (1980:170) and in the HKCSE

Type of tone	**D-use *well* in HKCSE**	**D-use *well* (Svartvik 1980)**
Fall	543 (47.2%)	53 (62%)
Fall-rise	21 (1.8%)	3 (4%)
Level	427 (37.1%)	14 (16%)
Rise	119 (10.3%)	10 (12%)
Rise-fall	0 (0.0%)	5 (6%)
Unclassified	41 (3.6%)	0 (0%)
Total	1,151 (100%)	85 (100%)

A comparison of the intonational pattern of the 85 tokens of D-use *well* in Svartvik (1980) and in the HKCSE shows generally comparable results, though the proportion of falling tone is much higher in Svartvik (1980) and the level tone on the D-use of *well* is less than half of that found in the HKCSE. In addition, there are no instances of rise-fall tone on *well* in the HKCSE, regardless of whether or not it is a discourse particle. This is in accordance with Brazil's (1997) observation that the use of rise-fall tone is in general relatively uncommon in the data on which his discourse intonation framework is based. As the number of *well* studied in Svartvik (1980) is relatively small in size, more data on the tonal pattern of *well* are needed for comparison.

In addition to the prominence pattern and tonal pattern, the occurrences of pauses in the neighbourhood of *well* are also examined in the HKCSE. Table 4.7 shows the distribution of unfilled pauses before and after *well*:

Table 4.7. The distribution of pauses before and after *well* in the HKCSE

Type of tone	**D-use *well***		**P-use *well***	
	Before *well*	After *well*	Before *well*	After *well*
Brief pause	126	48	1	29
Unit pause	33	3	0	7
Total	159	51	1	36

As is apparent from Table 4.7, the occurrences of pauses display highly contrastive patterns for the D-use and P-use of *well*. While more than 200 pauses occur immediately before or after the particle *well*, only 37 pauses are found when *well* is employed for other non-discourse uses. In addition, the position in which the pauses occur is exactly the opposite. While more pauses are found before *well* when it is used as a discourse particle, the reverse is true for its propositional use. Of the 37 pauses found in the immediate context of P-use *well*, 36 of them occur after *well*. This contrast in the position of pauses can be seen in the following two examples. In example 14, the D-use of *well* is preceded by a brief pause:

(14)

b: …{ = the [< OpeRAtion >] } (.) { = [< ***WELL*** >] } { = can somebody [ELSE] < HELP > him }…

(HKCSE, A004, 169)

In example 15, on the other hand, the P-use of *well* is followed by a brief pause:

(15)

a: …{ \ < ^ ONE > day we [HOPE] that } { \ you'll be [^ JOINing] the < INdustry > } { \ as [< ^ ***WELL*** >] } (.) { \ so i [DON'T] not to TELL you it's not as good as it < ^ IS > }…

(HKCSE, B061, 5780)

The stronger inclination of D-use *well* to be preceded instead of followed by pauses in the present study is consistent with findings reported in Svartvik (1980), where 55 and 33 pauses were found before and after *well* respectively. The rather low proportion of D-use *well* to be associated with pauses (210 out of 1,913, 11%) in the HKCSE is also compatible with Stenström's (1990) remark that *well* is less likely to attract pauses than other particles.

As shown by the prosodic analysis above, the prosodic profile provides a rich dimension for the study of *well*. In particular, it offers useful clues to disambiguate the D-use and P-use of *well*. The prominence pattern and the occurrences of pauses, in particular, illustrate contrast between the different senses of the word. It has been suggested in various studies that the prosodic features may give important signals to the interpretation of the discourse use of *well*. This aspect will be examined in the functional analysis of D-use *well* in the HKCSE to investigate the role of prosody in the pragmatic functions of *well*.

4.3.4 Discourse functions of well *identified in the HKCSE*

In the corpus containing 1,913 uses of *well* as a discourse particle, six major functions are identified, serving functions in three different domains: textual, interpersonal and interactional. Instances of *well* serving functions in the textual domain are concerned with the structuring and organisation of discourse. They largely correspond to the text-oriented organisational unit (OT) in Linear Unit Grammar (Sinclair and Mauranen 2006). In Halliday's terms, they achieve the textual meta-function. Interpersonal functions are related to the expression of

attitudes, emotions and personal evaluations whereas interactional functions facilitate processes such as planning and turn management in the interaction. Tokens of *well* expressing functions in the interpersonal and interactional domains in the present study realise Halliday's interpersonal meta-function. They largely correspond to the interactive-oriented organisational unit (OI) in Linear Unit Grammar (ibid.). Since these three functional domains are not mutually exclusive, the functions categorised under each domain only present the key focus of the interpretation given the context but not all the potential meanings of the examples. As pointed out by Sinclair and Mauranen (2006:61), "natural language is not a sharp instrument with absolute or rigid boundaries, but is blurred at the edges". Multifunctional uses of *well* are hence not only common but also expected. In such cases, a dominant function is determined according to the linguistic and contextual details available. The six major functions identified are further divided into sub-functions which will be discussed at length with illustrating examples in the following.

Textual functions

4.3.4.1 Frame (framing)

One of the most frequently occurring functions of *well* is as a frame, i.e. to insert a point of division or transition for easy comprehension. There are 589 instances of *well* serving the function of a frame in the HKCSE, constituting 30.8% of all instances of D-use *well*. More than half (57.8%) of the instances of *well* in this function occur medially. This appears to be consistent with Svartvik's (1980) observation that this use is normally embedded in discourse. Although neither of the positions constitutes an overwhelming proportion, it will be shown later that there are positional preferences for the various sub-types of the frame function. In the literature, the frame function of *well* for textual organisation has been widely discussed in previous studies. In their analysis of classroom interactions, Sinclair and Coulthard (1975) categorise *well* alongside other markers such as *now*, *right* and *okay* as examples of the framing move to mark boundaries in teaching exchanges. In the present study, it is found that *well* can be used to signal a wide range of transitions in discourse. In the management of topics, for instance, *well* can be used to mark a topic shift, be it the change to an entirely different topic or a new aspect of the same topic. In example 16 in a student placement interview, the interviewer (speaker b) is first asking the student (speaker a) who is interested in getting a

placement in the hotel about her reading habit. Notice his use of *well* to shift the topic from reading to the education system in Hong Kong:

(16)

b: …{ [DO] you have TIME < TO > } { = [< READ >] your } (.) { = [LIKE] < NOvels > } { = [< OR >] } { = [< FICtions >] }

a: { = [< ^ ER >] }

b: { = [LOVE] < STOries > } { \/ [< Anything >] like this }

a: { \ [< ER >] } { / i [< ^ LIKE >] } { to [READ] < COmics > } { = [beCAUSE] i THINK er ^ AFter i REAding all < THE > } { / er [< TEXT >] } { = on [< THE >] } { / [< THOSE >] } { \ [< maTErials >] } { = i [< THINK >] er } { \ it's [< BETter >] } { \ for me to [READ] the < COmics > } { = because it will [< BE >] } { \ [QUITE] < reLAXED > }

b: { / [< Mhm >] } { \ i [< SEE >] } { \ ***well*** [^ TALking] about < INterest > } { \ and [< HOBbies >] } { = [I'M] just WONdering < UM > } { / [< ^ HOW >] } { \ do you [< THINK >] } { = [< UM >] } { = in the [< PROcess >] } { = [UM] when you're < STUdying > in the } { = in the [polyTECHnic] < uniVERsity > } { \ [MAYbe] as WELL as in the secondary < SCHOOL > } { = [< ER >] } { = we've [< ALways >] } { = [< UM >] } { / [SAID] THAT you < KNOW > } { \ in [HONG] kong the eduCAtion < SYStem > } { \ there're TOO much [< THINGS >] }…

(HKCSE, B066, 2740)

This kind of topic change can be announced by a meta-linguistic comment, which is a typical strategy in prepared talk such as presentations:

(17)

B: …{ = but [YET] at the same < ^ TIME > } { = [< NOT >] } { = [< VIolate >] } { = our [< GROUP >] } { \ [< HARmony >] } (.) { \ [< ^ YEA >] } { \ it [CAN] be < ^ DONE > } (.) { \ it [CAN] be < DONE > } { = [< ^ oKAY >] } { = ***well*** let's [TALK] a little bit < aBOUT > } { \ [< _ CONflict >] } { = [< ^ WHY >] } { = is [CONflict] < MANagement > } { = [< SO >] } { \ [< imPORtant >] }…

(HKCSE, B123, 8340)

In example 17, speaker B uses *well* in combination with the meta-comment *let's talk a little bit about* to ensure a smooth transition of topic. However, digression can also be introduced abruptly, as in example 18, where the teacher is using *well* to indicate a sudden change of his train of thought from talking about the next stage in the discourse event to the time left for the whole session:

(18)

B: { \ [< oKAY >] } { \ [THANK] you VEry < MUCH > } { = [< UM >] } { = we'd [< BETter >] } { = [NOW] move < ONto > } { \ [***WELL***] we've < ^ GOT > } { \ twenty minutes [< MORE >] } { = so you [SHOULD] be able to GET through the < MAIN > } { \ [< PREsenTAtion >] } { = [< UM >] } (.) { \ but we [PRObably] WON'T have much < _ TIME > } (.) …

(HKCSE, A029, 10415)

Apart from topic management, *well* can also be used to signal the development in discourse stage such as openings and closings. In public speeches, for example, it is not uncommon for speakers to use *well* to initiate talk at the beginning of the discourse, as in example 19:

(19)

B: …{ ? [< AS >] } { = [FInancial] < secreTARY > } { = he's [CHAIRman] of the exchange FUND adVIsory < ^ comMITtee > } (.) { = the [< GOverning >] body } { \ the [< HONG >] kong } { \ [MOnetary] < _ auTHOrity > (.) { = [< ANtony >] }

((applause))

b: { \ [< ***WELL*** >] } { = [THANK] you < SCOTT > } { = er good [< MORning >] } { = [LAdies] and < GENtlemen > } (.) { \ what a [WONderful] < SIGHT > }…

(HKCSE, P051, 8)

Similarly, *well* is used to mark the closing stage of a discourse event, or "opening up closings", as Schegloff and Sacks (1973:290) put it. The use of *well* in closings is especially frequent in conversations and TV interviews in the HKCSE. In the following example, the talk show host uses *well* to round off the TV programme:

(20)

b: { \ [< Okay >] } { = ***well*** [THANK] you both very < MUCH > } { \ and a very [HAPpy] new year to < YOU > } { = that's [ALL] we have for you this < WEEK > }…

(HKCSE, P127, 5543)

As seen from the above examples, the use of *well* to signal a new discourse stage or topic can occur either initially or medially. The occurrence of initial *well* is especially frequent in conversations and interviews in the HKCSE for the transition of topic and discourse stage. In fact, most of the instances of utterance initial *well* found in the frame function are related to topic or discourse management. Other sub-types of the frame function which are to be discussed in the following, on the other hand, are more likely to occur non-initially.

In addition to marking boundaries on the macro level of topic and discourse stage, the particle *well* can also indicate a break on the micro level, serving a role similar to punctuation marks in dividing words into sentences and clauses in the written language. The use of *well* to introduce direct speech, for example, is concerned with the shift of orientation from one speaker to another speaker or the same speaker in another place and time. In Schiffrin's (1987:122) words, *well* indicates a shift in "the deictic centre of talk". As discussed before, the typical prosodic profile of *well* as a discourse particle is either as a separate tone unit or as a pre-head. However, this sub-type of *well* in reporting clauses has a comparatively higher possibility to occur in the head and tail positions. It is also more likely to be the tonic syllable in a shared tone unit, both when compared with other frame functions and with other major functions in general. Of the 109 instances of *well* used to initiate direct speech in the corpus, only 47 of them (43%) are in a solitary tone unit or in the pre-head position, in contrast to 80% in such positions for D-use in general (see also Section 4.3.2.2). In example 21, the use of *well* to signal the beginning of direct speech quotation is in the head position, immediately preceding the nucleus *Jack*:

(21)

b: …{ = i [STILL] remember when we < WENT > to a } { \ a [BIG] company in hong < KONG > } { = [< ER >] } { \ we [WENT] to the C e o the C e o said ***well*** < ^ JACK > } { = you know [< ^ HOW >] can your guys } { \ who's see in their [EArly] < THIRties > } { \ [^ TELL] us how to BUY steel < BETter > }…

(HKCSE, B094, 3060)

Other than acting like an open quotation mark to signal a change in perspective, *well* can also indicate a break in question and answer pairs. In such cases, the speaker initiates a question which is followed by his/her own response. Before giving the response to the question s/he raises earlier, the speaker inserts *well* to signal the transition from the question to the reaction (cf. also Schiffrin 1987; Müller 2005). Example 22 shows this sub-type of *well* in a lecture. The teacher (speaker B) is explaining why it is necessary to hold different kinds of inventories:

(22)

B: …{ / so [THEY'RE] the various TYPES of < invenTORY > } { / that we are [TALking] < ^ aBOUT > } { \ but [WHY] do we < HOLD > them } { / [< ***WELL*** >] } { \ there's [QUITE] a < ^ FEW > } { = [REAsons] < WHY > } { \ [COMpanies] HOLD < INventory > }…

(HKCSE, A005, 4451)

For this sub-type of frame, speakers use *well* to initiate reactions to their own talk as if they are responding to another person. This sub-type of *well* is mostly found in monologues in the HKCSE. Of the 31 instances of *well* in question and answer pairs, 24 of them (77%) are from lectures, business presentations and public speeches. A possible reason is that in monologues, speakers are more likely to ask questions which are to be answered by themselves straight away as a tactic to raise listeners' attention. However, since the physical presence of a question mark is not possible in the spoken language, speakers need to resort to other resources to signal the end of the question as well as the start of the response which follows. The use of *well* and a rising intonation at the end of the question, for example, are thus employed to indicate this change.

Another sub-type of *well* working on the micro level involves the use of *well* in separating two parts of a text in a way similar to what commas do in the written language. In this respect, the particle *well* divides the talk into different sections which facilitates comprehension or emphasizes the topic currently discussed. This sub-function of *well* as a frame is normally associated with left dislocation, topic introducers (see also Aijmer and Simon-Vandenbergen 2003) and main clauses following if-clauses. Example 23 shows the occurrence of *well* right after left

dislocation in an informal office talk. The topic *five principles on the wall* is highlighted with the remainder of the message introduced by *well*:

(23)

a: …{ = [< AND >] } { ? the other thing } { = [< SO >] } { \ the [^ FIVE] < PRINciples > } { \ on the [< WALL >] } (.) * { = ***well*** we [< ^ HAVE >] it }

B: ** { \ [< YEAH >] }[33]

{ = [< BUT >] } { = you [< DON'T >] think } { \/ is that [CORporate] < CItizenship > right }

(HKCSE, B146, 7601)

Likewise, the topic *Hong Kong* is given extra emphasis in front in the phrase *and as for er Hong Kong* before the rest of the comment is brought out by *well* in the following example:

(24)

a: { = [< AND >] as for } { = [< ER >] } { \/ hong [< KONG >] } { \ [< ^ ***WELL*** >] erm } { = with [< THE >] er } { = [imporTAtion] of < THESE > mainland } { = [TAlent] and < proFESsional > } { = it would [HELP] hong kong to TAP the vast mainland < MARket > }

(HKCSE, P089, 1153)

In these examples above, *well* acts as a verbal comma to divide the texts into smaller chunks of talk. In real-time interactions where speakers are confronted with time and memory constraints, such uses of *well* contribute to the overall efficiency of exchange by giving useful clues to the hearers in processing as regards the division of units in discourse.

On the whole, the use of *well* as a frame facilitates the overall textual organisation of talk. When signalling the change of topic or discourse stage, the particle *well* smoothes the move by maintaining the conversational flow and making the transition less awkward. In other cases, it serves as verbal punctuation marks to divide the text into different parts for easier comprehension.

[33] When the overlapping speech goes beyond the end of a line, the line immediately follows the overlapping speech represents the continuing talk of the current speaker, if any (unless otherwise specified), and the next line represents the continuing talk of the speaker who interrupts, if any (unless otherwise specified). Hence in example 23, lines 4-5 are the continuing speech of speaker a.

4.3.4.2 Link (linking)

As a link, *well* introduces explanations and additional information to the preceding discourse. Although both the use of *well* as a frame and as a link contribute to the textual organisation of discourse, the linking function of *well* focuses on connecting the preceding and following parts of texts together to achieve coherence. In the *Collins Cobuild English Language Dictionary* (2002), this use of *well* is generally described as "to indicate that you intend or want to carry on speaking". This highlights the "continuative" aspect of *well* to signal that something is going to follow (Schourup 2001:1043). In their work *Cohesion in English*, Halliday and Hasan (1976:269) briefly describe this use of *well* as an introducer of "an explanatory comment". Similarly, the use of *well* to continue an opinion or an answer is found in Müller (2005). All these studies describe the linking property of the particle and regard what follows *well* as a continuation of what comes before.

When compared with the frame function, the use of *well* as a link is much less common in the HKCSE. Of all the instances of D-use *well*, only 70 of them are found to be used as a linking device. In example 25, the interviewee (speaker a) provides additional information about her supervisor following an utterance medial *well*:

(25)

a: …{ \ [< SO >] } { = i [START] < ER > } { = [< TO >] er } { = [WRITE] a < proPOsal > } { = and i [< TALK >] to my } { = er [< SUperviSOR >] } { = [< ER >] } { \ [< _ ***WELL*** >] } { = he's a [< LECturer >] }…

(HKCSE, B082, 913)

However, this use of *well* can also appear at the beginning of a turn when the speaker is interrupted earlier and intends to carry on talking when the opportunity arises. In example 26, the two speakers are having an informal conversation about the maid employed by the native speaker of English (speaker B). In the middle of the talk, the Hong Kong Chinese (speaker b) interrupts and shows his understanding. Notice the use of *well* in initial position by the native speaker of English as an attempt to further supply information about the work routine of the maid:

(26)

B: …{ \/ [^ I'M] not sure what the < RULES > are } { = but [SHE] gets up in the morning and GOES and < CLEANS > } { \ an [< OFfice >] } { \ it's [< PAID >] for that } { \ [THAT'S] < ilLEgal > } { \ [THAT'S] < ilLEgal > } { / as i [< underSTAND >] } { \/ [< filiPInos >] } { = [< CAN'T >] } { ? as [I] < underSTAND > fi } { / [filiPInos] CAN'T have any < WORK > } { = [< exCEPT >] } { \ [UNder] an emPLOYment < CONtract > } { = to [< BE >] a } * { \ [< doMEStic >] }

b: ** { = [< ^ OH >] } { \ i [< ^ SEE >] }

b: { \ [< doMEStic >] } { / [< YEP >] } { = [< YEP >] } { = [< YEP >] }

B: { = ***well*** [^ SHE] gets up in the morning * < GO > }

b: ** { \ [< NO >] } { \ [STRICTly] speaking it is < ilLEgal > }…

(HKCSE, C013, 1338)

As is shown by the two examples above, *well* as a cohesive device supplying explanations occur in both utterance initial and medial position, with a slight preference for it to occur initially (57.1%). In both cases, *well* indicates the intention by the same speaker to continue, connecting the extra piece of information to the prior discourse.

Interpersonal functions

4.3.4.3 Responsive signal (responsive)

The use of *well* as a responsive signal has long been highlighted in particle studies. As mentioned at the beginning of the chapter, some research only looks at the use of *well* in responses. In Schiffrin (1987:102), *well* is labelled as a “marker of response”. As the responsive use of *well* is concerned with the speaker’s reaction to the prior talk, it belongs to the interpersonal functional domain. Given that *well* in responses is considered the quintessential usage of the particle in many studies, it comes as no surprise that the number of *well* prefacing responses is the highest among all the functions in the HKCSE. In total, 708 instances of *well* are used in responses, making up 37% of the overall discourse use. An exceptionally high proportion (91.8%) of them is in utterance initial position, confirming results reported elsewhere regarding the typical position of this use (see, for example, Svartvik 1980). The use of *well* as a

responsive signal can be sub-divided into three types according to the responses which follow: dispreferred responses, direct answers and follow-up responses.

As a marker of dispreferred responses, *well* has been analysed from different perspectives in various studies (see also Section 4.1). A common label given to this function is qualifier (Svartvik 1980). In this usage, *well* precedes qualified answers, indicating that something needs to be qualified in discourse. This view is in line with Lakoff's (1973) analysis of *well* as a signal of insufficiency in answers. In the present study, it is found that *well* could preface responses which are dispreferred in a wide range of ways. In general, these responses are in some respects undesirable. Example 27 shows the type of qualified answers discussed in Lakoff (1973) and Svartvik (1980). It is a response given by the guest (speaker b2), who is a professor of architecture, to the yes-no question asked by the talk show host (speaker b1) in a current affairs interview. The use of *well*, in combination with *in many cases*, modifies the guest's answer and signals that his agreement is only partial:

(27)

b1: { \ so the [SEparate] rule you are TALking < _ aBOUT > } { \ is [< ACtually >] } { \ a [^ LOWer] < STANdard > }

b2: { \ ***well*** in [^ MAny] < CAses > } { \ [< YES >] } { = er [LET] me < SHOW > you } { = [< AN >] } { \ [< eXAMple >] } { \ [< _ HERE >] }…

(HKCSE, P102, 3749)

Apart from partial agreement, disagreement is another typical type of dispreferred response in talk (cf. Luke 1990; Pomerantz 1984 for dispreferred turn shapes). Disagreements, challenges and criticisms are dispreferred in the psychological sense that they are unwelcoming to other participants. In example 28, the Hong Kong Chinese female (speaker a) is talking to her friend, a male native speaker of English (speaker B). The male friend makes a statement that it's a mistake for him to run his own company. Notice how the Hong Kong Chinese uses *well* to preface her disagreement to his comment:

(28)

B: { = [< I >] } * { / [I] WORKED i WORKED for < mySELF > } { / for a

a: ** { = [< SO >] }

[FEW] < YEARS > } { ? [< WHICH >] } { = [< WHICH >] } { \/ was [^ PRObably] a < miSTAKE > }

B: * ((laugh))

a: ** ((laugh))

a: { \/ ***well*** it's er [< ^ ALways >] } { \ a [DIFferent] < ^ exPErience > } { \ i don't [THINK] it's a < miSTAKE > } { \ [< NO >] }

(HKCSE, C094, 3834)

In prefacing disagreements and criticisms, *well* often functions as a face mitigator as it modifies the certainty of the proposition that follows. It therefore serves to minimize the potential face threat arising from such face threatening acts. This corresponds to the observation that *well* is a device for face redress in previous studies (see, for example, Owen 1981; Watts 1986).

In addition, a response can also be dispreferred in the sense that it is structurally marked (Levinson 1983). In answers, it means that the response is not an expected answer arising from the option(s) offered by the question. For yes-no questions, *well* signals an answer which is neither affirmative nor negative. For wh-questions, *well* signals a reaction deviated from the normal expectation of the answer. There are many reasons why structurally marked answers are given. For example, speakers may have insufficient knowledge regarding the topic of the question. Or they may not be in the position to answer the question. In addition, there could be wrong assumptions in the questions asked. In example 29, the British male (speaker B) wrongly makes the assumption that the Hong Kong Chinese male (speaker b) buys from the Internet. Therefore, his friend has to correct his supposition before the conversation can proceed. Notice how the Hong Kong Chinese uses a combination of particles (*well* and *actually*) and laughter to weaken the illocutionary force of his assertion:

(29)

B: { = [< ER >] } (.) { \/ and [JUST] < CHOOsing > } { \/ the [RIGHT] < PROducts > } * { ? for the Internet } * { / [^ WHAT] sort of [< THINGS

b: ** { / [< Uhuh >] } ** { / [< Uhuh >] }

>] } { \ would you } { \ [< ^ BUY >] } (.) { \ [WHAT] sort of < THINGS > } { \ would [< YOU >] buy } { \ from the [< INternet >] }

b: { = [< ME >] }

B: { \ [< YEAH >] }

b: { / [< ***WELL*** >] } { \ [< ^ ACtually >] } { \ i [< SELL >] them } ((laugh)) { = [< ^ ERM >] } { \/ [< SOMEtimes >] } { \ i [FIND] some SEcond hand < PROducts > }…

(HKCSE, C112, 2384)

It has been suggested in studies elsewhere that the qualifying function of *well* is associated with fall-rise tone (see, for example, Leech and Svartvik 2002; Owen 1981). In the present study, however, no such correlation is found. In fact, only a mere 1.6% of instances of *well* as a qualifier carry fall-rise tone in the data. More than one-third of them are unstressed, closely followed by fall and level tone, each constituting about a quarter of those marking dispreferred sequences.

Of the three types of responses, the use of *well* to signal dispreferred responses is highly dominant. Out of the 708 instances of responsive *well* found in the HKCSE, 571 (80.6%) preface reactions which are qualified, unwelcoming or structurally marked. This provides a strong rationale for the emphasis on this function in previous studies.

Although only in small numbers in the HKCSE, *well* is also found to precede direct answers. In the data, there are 25 instances of *well* which occur in front of complete and direct answers which do not convey any dispreferred sense, refuting the claims in Jucker (1993) and Schiffrin (1987) that *well* is only used when the following utterance diverges from the prior one in terms of relevance or coherence. In the corpus, an example of *well* signalling direct answers is found in a current affairs TV interview. In example 30, the guest (speaker b2) uses *well* to preface his direct answer regarding how Hong Kong should deal with SARS if the epidemic returns in the year after:

(30)

b1: …{ = [^ WHAT] do you think < HONG > kong } { = should [LEARN] from this < ER > } { \ [< _ exPErience >] } { \ i mean there are [^ SOME] people < SAYing > } { \ that the [SARS] thing will come BACK again NEXT < YEAR > } { = [< ^ IF >] so } { \ [HOW] should hong kong < HANdle > it }

b2: { / [< ***WELL*** >] } { \ [WE] must get < prePARED > } { = we [< MUST >] } { = [< TRY >] } { = [< TO >] er } { \ [BUILD] a < HOSpital > } { \/

as [SOON] as < POSsible > } { = we [MUST] < HAVE > a } { \/ [^ GOOD] rePORting < SYStem > }…

(HKCSE, P103, 4009)

As shown by the case above and examples cited in other studies (see, for example, Schourup 2001), it is indeed possible for *well* to occur in responses where no divergence of any sort is detected. In these instances, *well* appears to indicate that the speaker accepts the question and is prepared to provide the information required by the previous speaker (cf. also the semantic meaning of acceptance in D-use *well* and adverb *well* as in Bolinger 1989 and Carlson 1984). In this sense, it shows that the speaker is cooperative in conversation. Müller's (2005) study of *well* also provides instances of *well* in direct answers from corpus data. Similar to the finding in the present study, this use of *well* only occurs with a very low frequency in her analysis.

A final type of response preceded by *well* in the corpus involves the use of *well* in the introduction of a comment or a question arising from the prior utterance. In this function, *well* signals that the speaker's contribution is coherent with and relevant to the previous speaker's message. It serves to create a "harmonious interactive basis" on which communication can carry on smoothly (Fischer and Drescher 1996:860). Aijmer and Simon-Vandenbergen (2003:1146) discuss this use of *well*, glossing it as "in the light of the information you have just given me". In the data, there are 112 instances of *well* signalling follow-up responses. In example 31 in a placement interview, the interviewer (speaker a1) was asking the student (speaker a2) how to handle a difficult situation at work prior to the text provided below. After the student has provided her answer, the interviewer expresses her positive appraisal of the student's response with *well* at the beginning of her turn:

(31)

a2: …{ = [< I >] will } { / [< ASK >] } { / the [< STAFF >] } { = [< ER >] } { / [< inVOLVED >] } { = [TO] inVEStigate what < HAD > } { = [< WHAT >] really had } { \ [< HAPpened >] } { = [< UM >] } { = [< AND >] } { = to [deCIDE] < THE > } (.) { = [< UM >] } { / [< PUnishment >] } { = [< OR >] } { \ [someTHING] like < THAT > }

a1: { \/ [< oKAY >] } { \ [< _ ***WELL*** >] } { \ [BETter] THAN i < exPECT > } ((laugh)) { \ [< _ SO >] } { = [< THAT'S >] } (.) { / very [< GOOD >] } { \ [< oKAY >] }…

(HKCSE, B070B, 1283)

When compared with dispreferred responses, there is a lower proportion of level tone with *well* in direct answers and follow-up reactions. While 26.8% of *well* in dispreferred sequences carry level tone, only 16% and 17% instances of *well* are found with level tone in direct answers and follow-up responses respectively. In addition, it appears to be slightly more common for *well* in direct answers to carry a falling tone than in other responses. A possible reason is that in giving direct and complete answers, the speaker is more likely to use a proclaiming tone to convey a sense of certainty and definiteness (cf. Schourup 1985), while doing so in dispreferred responses may falsely put across the idea that the speaker is being too domineering in his/her assertion.

From the descriptions of *well* in the three types of responses above, it is apparent that although *well* marking dispreferred responses is the most common, the particle can also occur in direct answers and follow-up responses, signalling the speaker's willingness to show cooperation and to contribute to discourse coherence in interactions.

4.3.4.4 Expressing feelings (emotive)

Apart from marking speakers' reactions to the preceding discourse, *well* can also be used to convey a range of feelings such as surprise, concession and dismissal (cf. Carlson 1984). Schourup (2001:1043) describes this function of *well* as a "mental state interjection", signalling the speaker's state of mind as an emotional flow. As with the function of *well* as a linking device, the use of *well* to express emotive feelings is not frequent in the corpus. Only 70 instances are found to achieve this interpersonal function. As a feeling carrier, *well* is most commonly found in the HKCSE to convey a sense of resignation. In example 32, the Hong Kong Chinese female (speaker a) is talking to a British male (speaker B) about her wedding night, which fell within the tournament of the World Cup in 1990. As a football fan, her husband told her that he would spend the whole wedding night watching the World Cup matches. The British male first indicates his surprise at the husband's decision with two instances of *oh dear*. This is followed by a brief pause and repeated uses of *well* to show a sense of helplessness and resignation:

(32)

B: ((laugh)) { \ oh [< DEAR >] } { \ and he [ACtually] watched the < WORLD > cup }

a: { \ he [< WATCHED >] it } ((laugh))

B: { \ oh [< ^ DEAR >] }

a: ((laugh))

B: { \ a [< TRUE >] } { \/ a [TRUE] < FOOTball > fan * there } (.) { \ but a

a: ** { \ [< ^ MM >] }

[BAD] < HUSband > really }

a: ((laugh))

B: { \ [< _ YEAH >] } { \ [< ^ OH >] } { \ [< ^ DEAR >] } (.) { = [< ***WELL*** >] ***well*** } { \ [< ^ ***WELL*** >] } { = but er i [TAKE] it < HE > er } { \ [< _ YOU >] know }{ \ it's [<GOOD >] } { \ it's good [JOB] the < WORLD > cup } { = didn't [< LAST >] er } { \ very [< LONG >] } { \ right it [COMES] around every < FOUR > years } { \ so it's [< alRIGHT >] }…

(HKCSE, C001, 2410)

When compared with other functions, *well* as a marker of feelings is more likely to occupy final and stand-alone positions. In fact, most of the instances of D-use *well* in final position are found in this function. In stand-alone positions, *well* frequently occurs with other particles as in *ah well*, *oh well* and *well anyway*. Although the affective meaning expressed by *well* can be undesirable, this use of *well* is different from its use in dispreferred responses in the sense that here *well* alone is "purely emotive" and contains "no referential content" (Quirk et al. 1985:88n) while *well* in its qualifying function has to be followed by some other proposition which expresses attitudinal meaning. This difference can also be seen from the tone pattern for the two functions. As mentioned above, there are more unstressed examples of *well* than examples with any other tone in dispreferred sequences (see also Section 4.3.4.3). On the other hand, *well* as an emotive signal displays a different pattern. The instances of *well* with falling tone make up the largest category, constituting 37.1% of the total in this function. This seems to suggest that when *well* plays an emotive role, there is a greater need to give emphasis on the word in order to convey feelings.

Interactional functions

4.3.4.5 Processing device (processing)

Apart from textual organisation and maintenance of interpersonal relationship, *well* also serves important functions on the interactional level. When speakers interact with each other, they are constantly engaged in the process of planning. Impromptu speech, in particular, exerts extra pressure on real-time processing. This use of *well* thus not only allows speakers to signal to other participants that some internal processing is going on; it also enables them to gain extra time for their turn. In the HKCSE, there are 335 instances of *well* used to facilitate processing, making up 17.5% of the total number of D-use. As a processing device, *well* marks a wide range of self repair phenomena such as word recovery problems, self-correction and rephrasing. In example 33, the speaker encounters a word recovery problem when she talks about the element to which derivational morphemes are added. Notice her use of the filler *um* and the particle *well* as place-holders while she undergoes some internal contemplation before resting on her choice of the expression *word stem* among other options:

(33)

a: …{ = [BUT] < ^ WHAT > } { = i'm [< SAYing >] } { \ [< IS >] } (.) { \ [DEriVAtional] < MORphemes > } { = they [Usually] when they ADD to < UM > } { \ [< _ ***WELL*** >] } { \ [WHEN] they ADD to a WORD < STEM > } (.) { \ they [enTAILS] a CHANGE of < MEAning > }…

(HKCSE, A028, 11049)

In self correction, *well* serves to signal that the word following *well* replaces an error made before. The mistake could be phonological or semantic, as in example 34:

(34)

a: { = [< YEAH >] } ((laugh)) { \ of [< ^ COURSE >] } (.) ((laugh)) { = [< YEAH >] } { = [< BUT >] she } { = [SHE'S] looking < AFter > my } { \ [< _ ***WELL*** >] } { ? [S##'s[34]] < ^ BROther > }

(HKCSE, C006, 361)

[34] The name of the person is replaced by two hashes here and in other examples for anonymity.

However, *well* can also signal self editing where there is no discernible error. In other words, the speaker is subconsciously repairing what is in his/her mind even when the "mistake" has not been actually spelt out. In example 35, the native speaker male (speaker B) is revising his use of the expression the *last governor* as if he is responding to a mistake made by himself:

(35)

a: { = [SOMEtimes] * we see VEry < GOOD > erm } (.) * { ? [< UM >] }

B: ** { ? and the } ** { ? ***well*** [NOT] the < LAST > governor but } { \ the [GOvernor] < beFORE > er } (.) { \ [< ^ W## >] } (.) { \ [apPArently] was VEry very good at speaking < CANtonese > }…

(HKCSE, C090, 7442)

Another kind of self-repair involves the speaker refining what is just said before in order to give a more accurate and precise description. In example 36, the speaker is explaining the term *sabbatical*. Instead of rejecting his original choice of the word *leave*, he elaborates what kind of leave a sabbatical is:

(36)

B: …{ = i'm i'm [ONE] of the FEW < PEOple > } { \ that have managed to [neGOtiate] one of these things called a < ^ sabBAtical > } (.) { = [< _ ER >] } { = which is like [< A >] } { = [< UM >] } { \ it's [NOT] actually a < LEAVE > } { ? [< IT'S >] } { \ ***well*** it [< ^ IS >] leave } { ? [< BUT >] it's } { \ but it's [< LEAVE >] } { = to do a [< speCIfic >] } { \ [< _ THING >] }…

(HKCSE, C080, 2662)

In addition, *well* is used as a delay tactic when the speaker is hesitant about what to say next. It signals that the speaker is facing some processing difficulties and needs additional time for planning before proceeding, as shown in example 37:

(37)

a: …{ = [AND] < _ ER > } { = [< ER >] } { = and [< ALso >] } { = [< _ UM >] } { = [< ER >] } (.) { \ [< ***WELL*** >] } { = [< ER >] } { = [< AND >] } { \ and [ALso] the TRAditional < MEthod > }…

(HKCSE, B082, 807)

When *well* is used for processing purposes, it is typically in utterance medial position. More than 80% of *well* in this function occur medially. *Well* in processing is also more likely to carry falling tone than unstressed, unlike the general intonational pattern of D-use *well*. More than 40% of *well* in this function are with falling tone. In terms of associations with pauses and other linguistic items, *well* as a processing device frequently occurs with filled and unfilled pauses (cf. Stenström 2006). The voiced hesitation words such as *er*, *erm* and *um* are common in the vicinity of *well*, especially immediately preceding the particle (cf. Svartvik 1980). It is also found in the close proximity of another particle, *I mean*, which is sometimes used as a repair marker (Aijmer 2002). In addition, repetition, false starts and incomplete words and structures abound in such instances, giving further indications that the particle *well* signals speakers' need for some 'time off' for mental processing and internal consultation before the interaction can proceed properly.

4.3.4.6 Turn management (turn managing)

Apart from indicating to the hearer that some sort of processing is taking place on the part of the speaker, *well* can also signal speakers' desire to take control of the conversational floor. The use of *well* in turn management can be divided into two types: floor-holding and turn-taking. In the HKCSE, there are 85 instances of *well* contributing to the organisation of turns. When participants are interrupted while holding the conversational floor, they may need to reinstate their speaker role at times in case someone else wishes to take over. In example 38, for instance, the native speaker male (speaker B) is indicating his intention to carry on speaking despite the interruption from the Hong Kong Chinese male (speaker b):

(38)

B: …{ = the [Other] one to aNOther < ENglish > } { = [LANguage] TYPE < TEAching > } (.) { \ [< JOURnal >] } * { = [< SO >] } { ? ***well*** it's it's }

b: ** { = is it } { / [ER] asSESSment and eVAluation in < HIGHer > education }

(HKCSE, C004, 1457)

The above example also shows the multi-functional nature of *well* as the native speaker is obviously in a state of mentally processing the preceding discourse while

the overlapping of speech occurs. Unsurprisingly, the use of *well* in floor-holding is normally in utterance medial position. On the other hand, participants could also use *well* when they are currently assuming the hearer role and would like to seize the right to voice their opinion from the current speaker. As a consequence, the turn-taking use of *well* is more likely to occur initially or in stand-alone position. In example 39, the Hong Kong Chinese female (speaker a) attempts to take the conversational floor from her colleague (speaker y) when he is in the middle of his talk:

(39)

y: { = [< ER >] } { \ it's between [DOUble] and THREE times * < MORE > }

a: ** { \ ***well*** you know if you [^ CHECK] with your < STANdards > } (.) { \ we could [SET] the same < PAY > }…

(HKCSE, B148, 7132)

The use of *well*, sometimes alongside other particles such as *you know* in the example above, seems to make the interruption less direct and abrupt. Compared with initiating interruptions with content-bearing words, the use of discourse particles as turn openers appears to be a milder indicator of the hearer's turn-taking desire. As a result, the interruption might come out as less aggressive and thus less face-threatening. In addition, using the particle *well* to take the turn also ensures that when the speaking right is secured, there is no loss of propositional meaning during the course of simultaneous speech, hence enhancing the overall efficiency of talk (cf. Sacks, Schegloff & Jefferson 1978).

As is apparent from the above examples, *well* typically occurs in simultaneous talk with this function. When compared with other functions, the use of *well* as a floor-holder or turn-taker has a higher proportion of nuclear tones. Of all the instances of *well* with a turn managing function, only 20.0% are unstressed while falling tone and level tone each constitute more than one-third of the total. The use of level tone in particular is the highest among all the functions identified in the corpus. As discussed above, even when participants intend to assert their right to speak, they might still wish to come across as cooperative in the interaction. Choosing the level tone appears to be fitting as it highlights the desire of the speaker to take control of the conversational floor without sounding too authoritative.

4.3.4.7 Unclassified instances

As is so often the case in studies working with authentic data, the present study of *well* inevitably contains a number of instances which cannot be classified. This is a natural consequence arising from the volatile and interactive nature of conversation, which "reflects the multifunctionality and fuzziness inherent in natural language" (Sinclair and Mauranen 2006:68). In the corpus, there are 24 tokens of D-use *well* whose functions are indeterminate. Unclassifiable examples include instances with insufficient contextual information owing to inaudible speech or interruption. The use of *well* may also be uninterpretable because the speaker has a sudden change of mind and abandons his/her original train of thought altogether after the use of the particle. In other words, immediately after using *well* the speaker leaves the idea that s/he originally intends to express unfinished and moves on to a different grammatical construction. In addition, some repeated uses of *well* only constitute a single function and thus are counted once. The other consecutive uses are then categorised as unclassified to avoid double counting. Example 40 illustrates an unclassified instance owing to the lack of contextual information:

(40)

a: { = [< UM >] } { = [< THERE >] are } { \ [EIGHT] < CHANnel > for } { = [< AND >] i } { = [< I >] can } (.) ((inaudible)) ((laugh))

b: { \ ***well*** [< acCORding >] to * ((inaudible)) }

a: ** { = [< FOR >] er } { \ for [ONE] < loCAtion > }

(HKCSE, A018, 15384)

The functional analysis above shows that as a discourse particle *well* achieves a spectrum of functions in the textual, interpersonal and interactional domains. It also suggests that the six different functions display distinctive patterns in relation to a range of linguistic variables. In the following sections, these linguistic factors as well as a number of sociolinguistic and contextual factors will be discussed with regard to their inter-relationship with the functions of *well*.

4.3.5 Linguistic factors associated with the functions of well

In this section, the three linguistic factors collocations, positions and prosodic features of *well* will be discussed in turn in relation to the six pragmatic functions described above in order to investigate the extent to which these variables are related to the functions identified.

4.3.5.1 Collocations

It is suggested in the functional analysis of *well* above that some functions have a greater likelihood to co-occur with certain linguistic elements. As pointed out by Aijmer (2002), the linguistic elements which are frequently associated with discourse particles serve as clues for their functional interpretation. This section discusses the key collocational patterns of *well* identified in the HKCSE.

As a sub-type of frame, the use of *well* in signalling direct speech normally follows the reporting verbs *say* and *think*. Other reporting verbs such as *tell* and *figure* occur much less frequently with *well*. In particular, the reporting verb *say* is found to be most frequently associated with *well* with the function to report direct speech. This confirms the findings reported in Svartvik (1980). Out of the 109 instances of *well* in reporting clauses, there are 75 occurrences (68.8%) with the lemma *say*, including its variant forms *says*, *said* and *saying*. Table 4.8 shows the number of *well* within a span of five words after the two reporting verbs *say* and *think* in direct speech quotation:

Table 4.8. The distribution of *well* within a span of five words after the two most common reporting verbs *say* and *think* in direct speech quotation

Lemma before *well*	**n**	**in percentage (%)**
say	75	68.8%
think	17	15.6%
other reporting verbs	17	15.6%
Total	109	100%

In addition, the use of *well* in reporting clauses mostly occurs immediately following the reporting verb (see example 21). Direct speech quotation with *well* after an indirect object or a reflexive pronoun, as in example 41, occurs far less frequently (cf. James 1983):

(41)

B: …{ = we can [SAY] to ourselves < ***WELL*** > } { at [^ Any] point in < TIME > }…

(HKCSE, A006, 671)

In dispreferred responses, it is not uncommon to see the pattern *yes* / *yeah well… but*, especially in partial agreement. In total, there are 348 instances of *yes well* and *yea well* in the corpus while *but* occurs 69 times within the five-word span after the particle. The occurrences of words with negative connotation are also quite frequent, as in *well no*, *well I don't know* or *well I'm not sure*. Of the 1,913 discourse uses of *well* in the HKCSE, there are 347 uses of *no*, *not* and *don't* in the proximity of the particle. In the function of *well* expressing affective meaning, the co-occurrences of *well* with other particles as in *ah well* and *oh well*, in addition to the repeated uses of *well*, are found more frequently than in other functions. This gives further support to Aijmer's (2002:129) view that *oh well* is frequently used to convey the feelings of "nonchalance, indifference and lack of involvement". This type of collocation also suggests that the clustering of *well* with other discourse particles is common for the expression of feelings. Further, linguistic items such as *anyway* and *never mind* are used in combination with *well* to express dismissal and resignation. The strong interpersonal role of *well* is also indicated by its readiness to be immediately followed by a personal pronoun. In the data, the number of *well* followed directly by a personal pronoun (*I*, *you*, *he*, *she*, *we*, and *they*) amounts to 581, making up a total of 30.3% for this position for the overall D-use. The percentage is comparable to the proportion of 33% in a corpus of New Zealand English as reported in de Klerk (2005). This high percentage of co-occurrences with personal pronouns highlights the interpersonal function of *well* to express speakers' attitudes, evaluations and feelings.

When *well* is used as a processing device, its occurrences with a variety of other planning signals are notable. In the corpus, there are 638 instances of *er*, *erm* and *um* within the ten word span of D-use *well*, making these fillers one of the most frequent combinations with the particle. Their frequent co-occurrence with *well* indicates the hesitation and indecisiveness speakers are going through in processing.

4.3.5.2 Positions

In the functional analysis presented above, it is noted that the six discourse functions of *well* have different positional preferences. Table 4.9 summarises the positional distribution of the six functions of D-use *well* in an utterance:

Table 4.9. The positional distribution of the discourse functions of *well* in an utterance in the HKCSE[35]

Position (utterance)		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Final	Count	0	1	0	12	0	1	14
	% within Position	.0%	7.1%	.0%	85.7%	.0%	7.1%	100.0%
	% within Function	.0%	.2%	.0%	17.1%	.0%	1.2%	.7%
Initial	Count	40	261	650	30	62	48	1091
	% within Position	3.7%	23.9%	59.6%	2.7%	5.7%	4.4%	100.0%
	% within Function	57.1%	42.0%	91.8%	42.9%	18.5%	56.5%	57.8%
Stand-alone	Count	0	0	15	16	4	27	62
	% within Position	.0%	.0%	24.2%	25.8%	6.5%	43.5%	100.0%
	% within Function	.0%	.0%	2.1%	22.9%	1.2%	31.8%	3.3%
Medial	Count	30	359	43	12	269	9	722
	% within Position	4.2%	49.7%	6.0%	1.7%	37.3%	1.2%	100.0%
	% within Function	42.9%	57.8%	6.1%	17.1%	80.3%	10.6%	38.2%
Total	Count	621	70	708	70	335	85	1889
	% within Position	32.9%	3.7%	37.5%	3.7%	17.7%	4.5%	100.0%
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Some interesting findings could be observed from the table. When the particle *well* is involved in textual organisation, there appears to be a similar proportion in initial and medial positions. As discussed in Section 4.3.4.1, the preference for initial or medial position for the frame function depends on which sub-type of frame *well* belongs to. The use of *well* for topic and discourse management is more likely to occur in initial position while in other frame functions the particle is more likely to be embedded in discourse. It is worth noticing that as a textual device, *well* very rarely occurs in

[35] The 24 instances of *well* whose functions are unclassified are excluded in Table 4.9.

utterance final position and as a separate turn, since the particle can barely contribute to the division or connection of talk in these positions.

On the other hand, the use of *well* with interpersonal functions shows contrastive patterns. For responsive *well*, there is an overwhelmingly high proportion (91.8%) of instances in utterance initial position. In comparison, the use of *well* in expressing feelings has a more even distribution among the four positions. In particular, the emotive function has a higher proportion of *well* in final position (17.1%) than any other functions. It also has a higher proportion of *well* in stand-alone position (22.9%) when compared with all the other functions except turn managing. In fact, of the 14 instances of utterance final *well* in the corpus, 12 tokens (85.7%) are found to be associated with the expression of feelings. In other words, the data appear to suggest that when D-use *well* is found at the end of an utterance, it has a high probability of serving the emotive function.

For the interactional domain, a large proportion (80.3%) of *well* as a processing signal are embedded in discourse. This indicates that *well* is used to signal self repair and hesitation phenomena frequently in the middle of the talk. As for turn management, the majority of *well* occur initially or as a separate turn. This indicates that using *well* in turn taking is more common, as speakers utter a single *well* as an attempt to assume the speaker role or use the particle to initiate talk in simultaneous speech.

Within a tone unit, the various pragmatic functions of *well* appear to show a more conforming pattern. The positional distribution of the discourse functions of *well* in a tone unit in the HKCSE is shown in Table 4.10:

Table 4.10. The positional distribution of the discourse functions of *well* in a tone unit in the HKCSE[36]

Position (tone unit)		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Head	Count	62	9	41	0	25	2	139
	% within Function	10.3%	13.0%	6.0%	.0%	7.8%	2.4%	7.6%
Pre-head	Count	188	20	198	12	70	11	499
	% within Function	31.1%	29.0%	29.1%	17.1%	21.9%	13.4%	27.3%
Tail	Count	27	0	9	8	3	0	47
	% within Function	4.5%	.0%	1.3%	11.4%	.9%	.0%	2.6%
Tonic syllable (in a shared tone unit)	Count	72	8	55	7	38	9	189
	% within Function	11.9%	11.6%	8.1%	10.0%	11.9%	11.0%	10.4%
Separate tone unit	Count	255	32	377	43	184	60	951
	% within Function	42.2%	46.4%	55.4%	61.4%	57.5%	73.2%	52.1%
Total	Count	604	69	680	70	320	82	1825
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

As indicated by Table 4.10, the dominant positional preference for all pragmatic functions identified in the corpus is as a separate tone unit. The proportion is especially high for turn management (73.2%), matching the observation that *well* is more likely to constitute a separate turn for this function. In addition, it is also least likely to be in the pre-head position (13.4%). As a frame, the percentage of *well* in the head or tail position is relatively higher than that of many other functions (10.3% and 11.4% respectively). This is mainly due to the distinctive positional preference of the framing sub-type introducing direct speech quotation for these two positions (see also Section 4.3.4.1). As a linking device, *well* is also more likely to be the head. In example 42, *well* occupies the head position and acts as a link to connect the explanatory remark of how many dedicated hospitals are needed with the preceding short affirmative answer:

(42)

b: { \ if we [< ^ ARE >] to have } { \ [Other] inFECtious diSEAses < _ PAtients > } { \ [SHOULD] we HAVE DEdicated < HOSpitals > } { = [<

[36] Of the 1,847 instances of D-use *well* for which prosodic information is available, 22 tokens are functionally unclassified and are excluded from Table 4.10.

FROM >] } { \ the [PREsent] HOSpital < ^ SYStem > } (.) { \ the [ANSwer] is < YES > } { = [***WELL***] we NEED a few of < THESE > er } { \ er [DEdicated] < HOSpital > }

(HKCSE, P103, 2648)

In its emotive function, on the other hand, *well* is less likely to be found in the head position. In fact, no instances of *well* occupy the head position for the expression of feelings. Similarly, there is a lack of tail position of *well* in linking and turn management.

4.3.5.3 Prosodic features

While the prosodic features of *well* give indicative clues to the differentiation between a discourse use and propositional use of the word, they do not seem to provide distinct differences among the six pragmatic functions of *well*. The distribution of the prominence pattern of the six discourse functions of *well* is shown in Table 4.11:

Table 4.11. The prominence distribution of the discourse functions of *well* in the HKCSE[37]

Stress pattern		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Unstressed	Count	236	21	212	20	76	11	576
	% within Function	39.1%	30.4%	31.2%	28.6%	23.8%	13.4%	31.6%
First	Count	40	8	36	0	22	2	108
	% within Function	6.6%	11.6%	5.3%	.0%	6.9%	2.4%	5.9%
Middle	Count	1	0	0	0	0	0	1
	% within Function	.2%	.0%	.0%	.0%	.0%	.0%	.1%
Last	Count	31	1	1	1	6	0	40
	% within Function	5.1%	1.4%	.1%	1.4%	1.9%	.0%	2.2%
Sole prominent syllable (in a shared tone unit)	Count	41	7	54	6	32	9	149
	% within Function	6.8%	10.1%	7.9%	8.6%	10.0%	11.0%	60.3%
Sole prominent syllable (in a separate tone unit)	Count	255	32	377	43	184	60	951
	% within Function	42.2%	46.4%	55.5%	61.4%	57.5%	73.2%	60.3%
Total	Count	604	69	680	70	320	82	1825
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 4.11 above shows that the most frequently occurring stress pattern for all the discourse functions of *well* is as the only prominent syllable in a separate tone unit. It constitutes the highest proportion in all functional categories, making up 42.2% to 73.2% of use. In particular, 73.2% of the instances of *well* used in turn management are the only stressed syllable constituting a separate tone unit. This corresponds to its greater likelihood to be a free-standing element in an utterance. This also accounts for the relatively low proportion of unstressed *well* in this function. As a frame, *well* has a higher proportion of being the last prominent syllable in a shared tone unit. Again, this is largely owing to the characteristic prosodic pattern of *well* in signalling direct speech quotation as reported in Section 4.3.4.1.

[37] Of the 1,847 instances of D-use *well* for which prosodic information is available, 22 tokens are functionally unclassified and are excluded from Table 4.11.

Similarly, the pragmatic functions of *well* largely seem to converge on their intonational distribution. Table 4.12 represents the pattern of nuclear tones on *well* among the different functions:

Table 4.12. The distribution of nuclear tones for the discourse functions of *well* in the HKCSE

Tone		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Fall	Count	143	16	184	25	140	30	538
	% within Function	43.7%	40.0%	42.6%	50.0%	63.1%	43.5%	47.2%
Fall-rise	Count	5	1	11	1	2	1	21
	% within Function	1.5%	2.5%	2.5%	2.0%	.9%	1.4%	1.8%
Level	Count	125	16	176	17	59	30	423
	% within Function	38.2%	40.0%	40.7%	34.0%	26.6%	43.5%	37.1%
Rise	Count	45	5	45	5	13	5	118
	% within Function	13.8%	12.5%	10.4%	10.0%	5.9%	7.2%	10.4%
Unclassified	Count	9	2	16	2	8	3	40
	% within Function	2.8%	5.0%	3.7%	4.0%	3.6%	4.3%	3.5%
Total	Count	327	40	432	50	222	69	1140
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

The table above shows that when *well* is the tonic syllable, the general intonational preference for all pragmatic functions is falling tone, followed by level tone. In its linking and turn management functions, however, the proportions of fall and level tone are exactly the same. On the other hand, *well* has a stronger inclination to carry falling tone as a processing device than any other functions, with an overall percentage of 63.1% as opposed to the 26.6% with level tone.

Contrary to some of the observations made in other studies (see Section 4.3.4), it appears that no considerable difference exists in the prosodic pattern among the six major functions identified from the large number of *well* in the corpus data, though relative variations are discussed in the functional analysis and can be seen from the figures above. In the following section, three socio-linguistic and contextual variables will be examined in relation to the functions to determine to what extent these factors affect the functional distribution of *well*.

4.3.6 Sociolinguistic and contextual factors associated with the functions of well

This section investigates how the three sociolinguistic and contextual variables language background, gender, domain and text type influence the use of *well* in the corpus. Specifically, it is found that all these factors have an effect on the rate of D-use *well* and the functional distribution of the particle in the data.

4.3.6.1 Language background

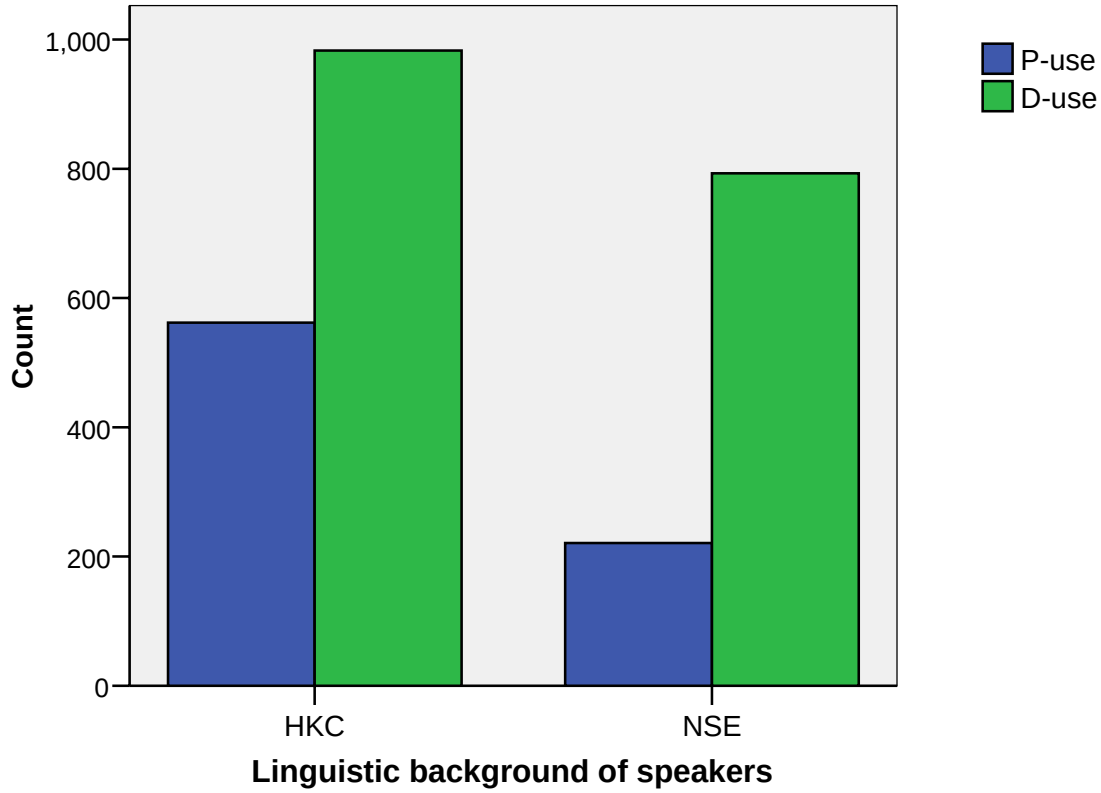
In the HKCSE, there is a higher rate of discourse use of *well* among native speakers of English than Hong Kong Chinese. Table 4.13 shows the distribution of *well* in native and non-native speakers' speech:

Table 4.13. The distribution of *well* in the HKCSE according to speakers' linguistic background (Hong Kong Chinese vs. native speakers of English)

	Hong Kong Chinese	**Native speakers of English**
Total number of words produced	669,431	240,847
Total number of *well*	1,545	1,014
Number of D-use	983	793
Number of P-use	562	221
D-use / total use (%)	63.62	78.21
D-rate (per 10,000 words)	14.68	32.93

The above table shows that although Hong Kong Chinese produce more than twice as many words as the native speakers of English in the corpus, the total number of *well* in the speech of native speakers is about two-thirds of that by Hong Kong Chinese. This suggests that *well* is more frequently occurring in the repertoire of native speakers. More importantly, when *well* is used by native speakers, it is more likely to be a discourse particle. This can be reflected by the higher discourse-function ratio in the speech of native speakers of English than in Hong Kong Chinese (78.21% vs. 63.62%). Figure 4.4 illustrates the difference between the number of D-use and P-use *well* produced by Hong Kong Chinese and native speakers of English:

Figure 4.4. The distribution of D-use and P-use *well* by Hong Kong Chinese (HKC) and native speakers of English (NSE)



The difference in the discourse and propositional uses of *well* between Hong Kong Chinese and native speakers of English is statistically very highly significant with a moderate association ($x^2 \geq 61.288$; d.f. = 1; $p < 0.0005$; V = 0.155). In addition, the discourse-rate of *well* is much higher in the speech of native speakers than in Hong Kong Chinese (see Table 4.13). In a sample of 10,000 words in the corpus, just about 15 instances of D-use *well* are produced by Hong Kong Chinese, compared to about 33 tokens by native speakers of English, making the use of *well* as a discourse particle twice as frequent in the speech of the latter speaker group. This supports findings reported in previous studies that D-use *well* is more often found in the data of native speakers of English (see, for example, Cheng and Warren 2000; Romero Trillo 2002). As far as the pragmatic functions of *well* are concerned, Hong Kong Chinese and native speakers of English also display different patterns of usage. Table 4.14 shows the functional distribution of *well* according to the linguistic background of speakers:

Table 4.14. The distribution of the discourse functions of *well* by linguistic background in the HKCSE (Hong Kong Chinese vs. native speakers of English)[38]

Speaker (Linguistic Background, LB)		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Hong Kong Chinese	Count	298	34	352	20	233	34	971
	% within Speaker (LB)	30.7%	3.5%	36.3%	2.1%	24.0%	3.5%	100.0%
Native speakers of English	Count	298	29	295	41	78	42	783
	% within Speaker (LB)	38.1%	3.7%	37.7%	5.2%	10.0%	5.4%	100.0%
Total	Count	596	63	647	61	311	76	1754
	% within Speaker (LB)	34.0%	3.6%	36.9%	3.5%	17.7%	4.3%	100.0%

One of the noteworthy features of Table 4.14 is the difference between the two groups of speakers in terms of their use of *well* for process purposes. For the Hong Kong Chinese, 24% of the instances of D-use *well* are used as a processing device. In contrast, the proportion is only 10% for the native speaker group. This percentage difference reaches a very high statistical significance ($x^2 \geq 58.527$; d.f. = 1; $p < 0.0005$; V = 0.183). This gives the implication that the Hong Kong Chinese in the corpus are more likely to use *well* to signal the processing difficulties they encounter in interactions. The greater likelihood of *well* as a strategy in repair and hesitation for non-native speakers confirms the results discussed in Cheng and Warren (2000) and Müller (2005).

On the other hand, the native speakers in the data are more likely to use *well* as a frame and as an emotive expression. As a textual device, the framing function of *well* makes up 38.1% of the instances of D-use for native speakers, as compared to 30.7% for Hong Kong Chinese. A highly significant statistical difference is achieved ($x^2 \geq 10.492$; d.f. = 1; $p < 0.01$; V = 0.077), though it is not as marked as for the processing function. Although the emotive use of *well* is a relatively infrequent function in the data of both groups, its proportion in the speech of native speakers (5.2%) is more than twice of that for Hong Kong Chinese (2.1%), again achieving a

[38] 135 instances of D-use *well* produced by speakers of other languages and the functionally unclassified tokens are excluded from Table 4.14.

very highly significant statistical difference ($x^2 \geq 13.030$; d.f. = 1; $p < 0.0005$; V = 0.0003). This might suggest that the native speaker group is more likely to use *well* to convey feelings than their non-native counterparts.

4.3.6.2 Gender

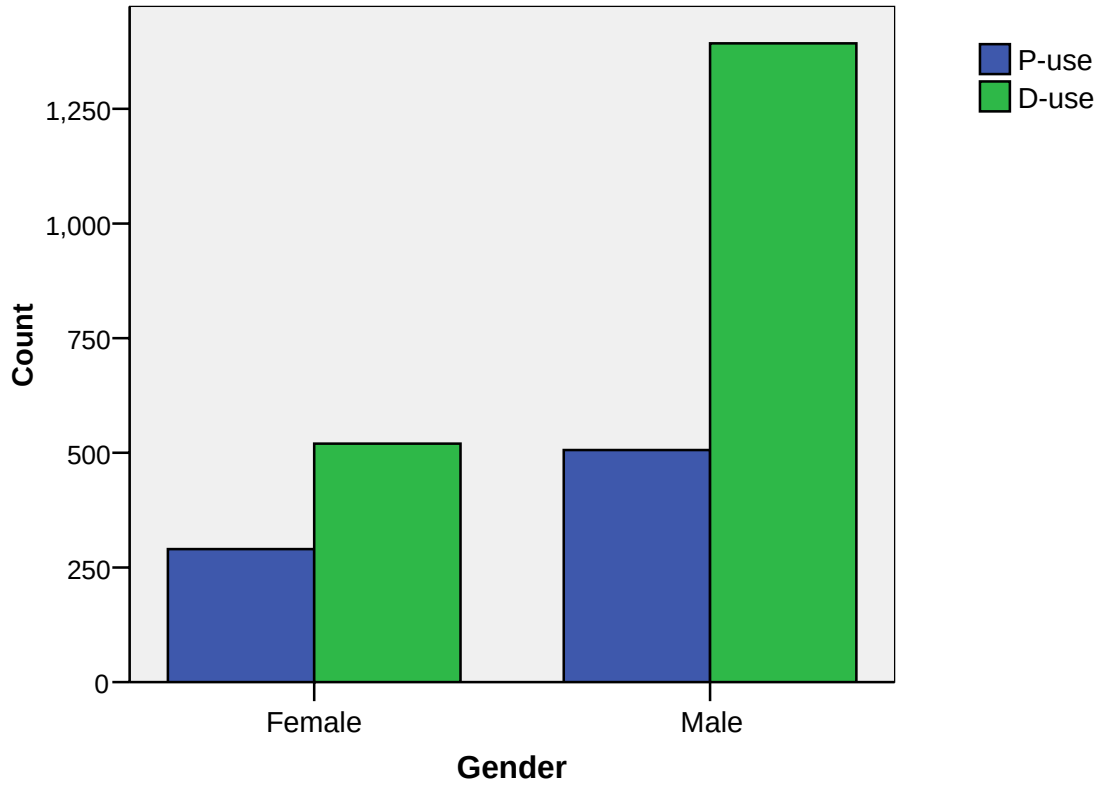
As regards the comparison of the use of *well* in the speech of male and female speakers, there also appears to be significant differences between them. Table 4.15 summarises the distribution of *well* in the speech of the two gender groups in the corpus:

Table 4.15. The distribution of *well* in the HKCSE according to speaker gender (male vs. female)

	Male	**Female**
Total number of words produced	556,815	389,659
Total number of *well*	1,899	810
Number of D-use	1,393	520
Number of P-use	506	290
D-use / total use (%)	73.35	64.20
D-rate (per 10,000 words)	25.02	13.35

As shown in Table 4.15, male speakers contribute a higher number of *well* and a greater overall number of words in the corpus. In addition, male speakers have a higher discourse-function ratio for their use of *well* than the opposite sex (73.35% vs. 64.20%). This means that *well* is more likely to be used as a discourse particle in men's speech. The difference between male and female speakers in the discourse-function ratio of *well* again is statistically very highly significant ($x^2 \geq 22.945$; d.f. = 1; $p < 0.0005$; V = 0.092), though the linguistic background of speakers appears to be a more determining variable due to its greater values of chi-squared and Cramér's V ($x^2 \geq 61.288$; d.f. = 1; $p < 0.0005$; V = 0.155, see also Section 4.3.6.1). This contrast is represented in Figure 4.5:

Figure 4.5. The distribution of D-use and P-use *well* by female and male speakers in the HKCSE



Apart from the fact that *well* is more likely to be a discourse particle when the word is used by male speakers, its D-use is also more frequently found in the speech of men. Table 4.15 reveals that the discourse rate of *well* is about 25 tokens per 10,000 words in men's talk, as opposed to about 13 tokens per 10,000 words in female speech. In other words, D-use *well* is twice as likely to be found in the data of male speakers as female speakers for the same number of words produced. This suggests that D-use *well* is more common in the speech pattern of men than women (cf. Macaulay 2005).

As regards the functional analysis of *well*, an examination of the functions of the particle among the two gender groups also reveals interesting patterns. Table 4.16 shows the functional distribution of *well* according to the gender of speakers:

Table 4.16. The distribution of the discourse functions of *well* by gender in the HKCSE (female vs. male speakers)[39]

Gender		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Female	Count	170	17	138	10	162	18	515
	% within Gender	33.0%	3.3%	26.8%	1.9%	31.5%	3.5%	100.0%
Male	Count	451	53	570	60	173	67	1374
	% within Gender	32.8%	3.9%	41.5%	4.4%	12.6%	4.9%	100.0%
Total	Count	621	70	708	70	335	85	1889
	% within Gender	32.9%	3.7%	37.5%	3.7%	17.7%	4.5%	100.0%

The table above displays some contrast in usage between the two gender groups. For women, there is a higher percentage of *well* used to signal self-repair and hesitation. About one-third (31.5%) of all the instances of D-use *well* produced by female speakers are used as a processing device while the proportion is only about one-eighth (12.6%) in men's speech. Statistically, this difference is very highly significant ($x^2 \geq 91.382$; d.f. = 1; $p < 0.0005$; V = 0.220). On the other hand, male speakers are more likely to use *well* as a responsive signal and as an emotive particle. While *well* in response is the most frequently occurring function in the data of male speakers (41.5%), it is only the third most common usage in female speech (26.8%). Again this achieves very highly significant statistical difference ($x^2 \geq 34.491$; d.f. = 1; $p < 0.0005$; V = 0.135). In particular, a large number of instances of the responsive *well* employed by men are associated with dispreferred responses, indicating that male speakers are more likely to express unwelcoming reactions with the particle. It might also suggest that men voice out negative responses more readily than women, and use *well* more often to preface those contributions. This is in accordance with results presented in gender studies which suggest that women are more cooperative in talk and men are more likely to challenge other participants (see, for example, Tannen 1990; 1994). Interestingly, the emotive *well* is also twice as frequent in men's talk (4.4%) as in the opposite sex (1.9%), indicating a greater likelihood that men will use *well* to express interpersonal feelings. It might possibly be due to men's stronger inclination to signal negative emotions such as resignation and dismissal, which are frequently associated with the use of *well*.

[39] The 24 functionally unclassified tokens are excluded from Table 4.16.

4.3.6.3 Domain and text type

At the beginning of the analysis of *well* in the HKCSE presented earlier in this chapter (see Section 4.3.1), it is shown that the D-use and P-use of *well* in the four domains achieve statistically significant difference. In fact, the four domains also differ in terms of the pragmatic functions identified. Table 4.17 represents the functional distribution of *well* in the four domains of the HKCSE:

Table 4.17. The functional distribution of D-use *well* by domain in the HKCSE[40]

Domain		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Academic	Count	114	11	38	4	37	0	204
	% within Domain	55.9%	5.4%	18.6%	2.0%	18.1%	.0%	100.0%
Business	Count	165	17	127	12	114	18	453
	% within Domain	36.4%	3.8%	28.0%	2.6%	25.2%	4.0%	100.0%
Conversational	Count	257	34	310	49	156	54	860
	% within Domain	29.9%	4.0%	36.0%	5.7%	18.1%	6.3%	100.0%
Public	Count	85	8	233	5	28	13	372
	% within Domain	22.8%	2.2%	62.6%	1.3%	7.5%	3.5%	100.0%
Total	Count	621	70	708	70	335	85	1889
	% within Domain	32.9%	3.7%	37.5%	3.7%	17.7%	4.5%	100.0%

As a frame which contributes to the textual structure of talk, *well* has a higher proportion of occurrences in academic texts than in any other settings. More than half (55.9%) of the instances of D-use *well* in the academic domain have a framing function. This highlights the importance of the particle in discourse organisation in teacher talk as pointed out by Sinclair and Coulthard (1975) and Tsui (1995). In the public domain, the use of *well* in responses is the most dominant. More than 60% of all instances of D-use *well* found in the public sub-corpus are associated with interpersonal reactions. This is largely due to the considerable number of *well* used in public TV interviews to signal dispreferred responses. In conversations, the emotive use of *well* makes up 5.7% of the total instances of *well*, contributing the highest proportion of this use when compared with the other three domains. This underlines

[40] The 24 functionally unclassified tokens are excluded from Table 4.17.

the emotive aspect of social discourse and demonstrates the significance of expressing interpersonal feelings in informal conversational talk.

Further, a look at the functional distribution of *well* across the different text types in the four domains gives additional information about the relation between functions and settings. Table 4.18 represents the functional distribution of *well* in some of the monologic and dialogic text types containing the highest numbers of the particle in the corpus:

Table 4.18. The functional distribution of D-use *well* across some major text types in the HKCSE

Text Type			Function						Total
			Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Dialogic	conversation	Count	257	34	310	49	156	54	860
		% within Text Type	29.9%	4.0%	36.0%	5.7%	18.1%	6.3%	100.0%
	interview	Count	87	10	259	8	101	13	478
		% within Text Type	18.2%	2.1%	54.2%	1.7%	21.1%	2.7%	100.0%
	meeting	Count	33	9	46	5	8	9	110
		% within Text Type	30.0%	8.2%	41.8%	4.5%	7.3%	8.2%	100.0%
Monologic	lecture	Count	62	5	3	0	6	0	76
		% within Text Type	81.6%	6.6%	3.9%	.0%	7.9%	.0%	100.0%
	speech	Count	36	3	4	0	8	0	51
		% within Text Type	70.6%	5.9%	7.8%	.0%	15.7%	.0%	100.0%
Sum		Count	475	61	622	62	279	76	1575
		% within sum	30.2%	3.9%	39.5%	3.9%	17.7%	4.8%	100.0%

As observed in the table above, monologic and dialogic text types display different preferences for the pragmatic functions of *well*. In multi-party talks, *well* is most frequently used as a response marker. This emphasises the interpersonal nature of dialogues. In particular, more than half (54.2%) of the instances of D-use *well* in interviews are used in responses. As mentioned in the previous section, this is mainly due to the overwhelming use of *well* in signalling dispreferred responses in public

current affairs interviews. In these talk show interviews, guests are specially selected to represent divergent views. In addition, talk show hosts are likely to raise sensitive and controversial questions in order to arouse interests. As a result, disagreements, indirect and insufficient responses occur more frequently, leading to a much higher proportion of *well* in dispreferred sequences (Lam 2006). Further, turn-taking is more rapid in dialogues, especially in conversations and meetings where speaker roles are not as rigidly assigned as is the case in interviews, leading to a higher proportion of *well* in achieving this function.

In monologues, on the other hand, the vast majority of the D-use instances of *well* are found to achieve the textual function of framing. More than 70% of *well* are used in lectures and public speeches as a framing device, compared to 30% or less in multi-party talks. In addition, no tokens of *well* are found in turn management in the two monologic text types. This suggests that in monologic discourse, speakers are more concerned with the organisation of text. Hence there is a higher proportion of *well* used to divide a long stretch of talk into shorter chunks for ease of comprehension. This is consistent with Sinclair and Mauranen's (2006) view that there are more text-oriented organising elements in long and complex turns. The interpersonal and interactional aspects of talk, in consequence, are sidelined by the textual dimension and are thus less emphasized.

4.3.7 Summary of findings in the HKCSE

In sum, the analysis of *well* in the HKCSE above reveals some interesting findings as regards the interrelationship between the particle and the various linguistic, sociolinguistic and contextual factors. While the position and the prosodic features of *well* appear to give strong clues to the distinction of the D-use and P-use of *well*, the prosodic features provide less guiding indications concerning the pragmatic functions of *well* when compared to its position. In addition, the linguistic co-texts of the particle offer useful hints about the interpretation of its discourse functions as different pragmatic uses tend to have distinctive patterns of lexical combinations. From the findings presented in the previous sections, it can be concluded that significant differences exist between speaker groups in relation to their gender and linguistic background. Furthermore, the setting of the discourse event constitutes an important contextual factor in determining how the particle *well* functions in talk.

In the remaining part of the chapter, the main findings reported in the HKCSE will be compared with those in the customised corpus from the BNC and the textbook database to investigate in detail the effect of the linguistic background of speakers on the use of *well* and the similarities and differences concerning how the particle is used in 'real' English and textbook English.

4.4 Comparison of findings in the HKCSE and the BNC customised corpus

In this section, the use of *well* in the customised corpus of the BNC will be examined and compared with the core findings from the HKCSE to determine the extent to which the differences identified in the Hong Kong corpus owing to the linguistic background of speakers are comparable to another independent source of data.

4.4.1 Frequency of occurrence

In the customised corpus of the BNC (BNC customised) with 948,920 running words spanning across 258 texts in four domains, there are 5,404 instances of *well*. Discourse and propositional uses of *well* constitute 4,379 and 1,021 tokens respectively, with four instances of unclassified *well* owing to unintelligible passages. Table 4.19 shows the distribution of the uses of *well* in the BNC (customised):

Table 4.19. The distribution of *well* in the BNC (customised)

	Total (N=258)	**Academic (N=25)**	**Business (N=86)**	**Conversational (N=122)**	**Public (N=25)**
Total number of words in BNC (customised)	948,920 (100.0%)	212,532 (22.4%)	260,484 (27.5%)	258,391 (27.2%)	217,513 (22.9%)
Total number of *well*	5,404 (100.0%)	875 (16.2%)	1,420 (26.3%)	2,497 (46.2%)	612 (11.3%)
Number of D-use	4,379 (100.0%)	668 (15.3%)	1,106 (25.3%)	2,184 (49.9%)	421 (9.6%)
Number of P-use	1,021 (100.0%)	207 (20.3%)	314 (30.8%)	309 (30.3%)	191 (18.7%)
Unclassified use	4 (100.0%)	0 (0.0%)	0 (0.0%)	4 (100.0%)	0 (0.0%)
D-use / total use (%)	81.03	76.34	77.89	87.46	68.79
D-rate (per 10,000 words)	46.15	31.43	42.46	84.52	19.36

The figures generated from the HKCSE (originally presented in Table 4.1) are reproduced in Table 4.20 for comparative purposes:

Table 4.20. The distribution of *well* in the HKCSE (reproduced from Table 4.1)

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Total number of words in HKCSE	949,972 (100.0%)	213,204 (22.4%)	259,484 (27.3%)	258,882 (27.3%)	218,402 (23.0%)
Total number of *well*	2,714 (100.0%)	316 (11.6%)	674 (24.8%)	1,071 (39.5%)	653 (24.1%)
Number of D-uses	1,913 (100.0%)	206 (10.8%)	454 (23.7%)	878 (45.9%)	375 (19.6%)
Number of P-uses	796 (100.0%)	110 (13.8%)	219 (27.5%)	189 (23.7%)	278 (34.9%)
Unclassified uses	5 (100.0%)	0 (0.0%)	1 (20.0%)	4 (80.0%)	0 (0.0%)
D-use / total use (%)	67.99	65.19	67.36	81.98	57.43
D-rate (per 10,000 words)	19.56	9.66	17.50	33.92	17.17

A comparison of the general quantitative findings of *well* from the BNC (customised) with those from the HKCSE shows largely comparable results. Similar to the patterns shown in the HKCSE, the conversational sub-corpus of the BNC (customised) contains the highest number of tokens of *well* in the collection (46.2% of total). It also houses the greatest number of the D-use of *well* (49.9% of total). In addition, the percentage rate of discourse-function is the highest in conversations (87.46%) and the lowest in the public domain (68.79%), displaying a parallel observation made in the HKCSE. Again, the difference in the discourse-function ratio between the four domains in the BNC (customised) is very highly significant statistically with a moderate association ($x^2 \geq 151.760$; d.f. = 3; $p < 0.0005$; V = 0.168), confirming the remark made from the findings of the HKCSE that formality of the setting influences the proportion of the different senses of the word.

In addition, an examination of the quantitative data from the BNC (customised) supports the findings reported earlier as regards the comparison of the rate of D-use *well* by native and non-native speakers in the HKCSE. From Table 4.19 and Table 4.20, it is noted that although the general patterns across the four sub-corpora remain similar in the two sources, the frequencies of *well* are uniformly higher in the BNC (customised). Considering that the two corpora are of almost equal size and similar composition, the fact that there are more tokens of *well* and a higher proportion of the lexical item used as a particle shows that the word *well* and its discourse use are more common in the speech of native speakers. The discourse-function ratio in the BNC (customised) is of comparable value to that found in the speech of native

speakers of English in the HKCSE (81.03% vs. 78.21%, see also Section 4.3.6.1). The discourse rate of *well* is even higher in the customised corpus than in the native data from the HKCSE (46 tokens vs. 33 tokens per 10,000 words). This shows an even greater contrast in the frequency of D-use *well* between Hong Kong Chinese and native speakers of English and further supports the claim that the pragmatic uses of *well* are much more commonly found in native speech.

4.4.2 Functional analysis

In terms of the pragmatic functions of *well*, an analysis of all the instances of D-use *well* in the BNC (customised) does not reveal functions which are unidentified in the HKCSE, though the proportions of some functions appear to vary slightly. Table 4.21 compares the functional distribution of the discourse use of *well* in the HKCSE and the customised corpus from the BNC:

Table 4.21. The distribution of the discourse functions of *well* in the HKCSE and the BNC (customised)

Speaker (Linguistic Background, LB)			Function						Total
			Framing	Linking	Responsive	Emotive	Processing	Turn managing	
HKCSE	Hong Kong Chinese	Count	298	34	352	20	233	34	971
		% within Speaker (LB)	30.7%	3.5%	36.3%	2.1%	24.0%	3.5%	100.0%
	Native speakers of English	Count	298	29	295	41	78	42	783
		% within Speaker (LB)	38.1%	3.7%	37.7%	5.2%	10.0%	5.4%	100.0%
BNC (customised)	British speakers of English	Count	1467	261	1557	133	533	288	4239
		% within Speaker (LB)	34.6%	6.2%	36.7%	3.1%	12.6%	6.8%	100.0%

As shown in Table 4.21, the native data from the BNC (customised) offer a generally comparable pattern with the one from the HKCSE. In particular, the uses of *well* in response and as a frame dominate in the two corpora for the three speaker groups. Although the framing function has a lower proportion in the BNC (customised) compared to the native data in the HKCSE (34.6% vs. 38.1%), the difference

between British speakers and Hong Kong Chinese is still statistically significant ($x^2 \geq 5.412$; d.f. = 1; $p < 0.05$; V = 0.032). This suggests that native speakers might use *well* as a textual device to divide chunks of talk slightly more often than Hong Kong Chinese. Another noteworthy feature is the higher proportion of *well* as a linking device and as a turn managing signal in the BNC (customised). While there is not a discernible difference in the linking use of *well* for the two speaker groups in the HKCSE, the British speakers in the customised corpus tend to use *well* more frequently to connect ideas. In example 43, the teacher gives an explanatory remark of what he meant by *well done* following *well*:

(43)

N: Right, (unclear).
Oh, you went through all the psycho history books?
Yeah, well done.
Erm, ***well*** it's a great achievement.
I think that given that you were er, you couldn't find the basic book that you were expected to read, I thought you, you gave an excellent account of it…

(BNC customised, he2, sn128, 3333)

Similarly, there appears to be a more frequent use of *well* with a turn managing function in the BNC (customised). While the difference in this function between native and non-native speakers in the HKCSE is not statistically significant, the comparison between Hong Kong Chinese and British speakers in the customised corpus achieves very high statistical significance ($x^2 \geq 14.770$; d.f. = 1; $p < 0.0005$; V = 0.053). However, since the number of instances for these two functions is relatively small and the strength of association is very weak, more data from comparable sources are needed until conclusions can be drawn.

As regards the emotive use of *well*, the observation made on the basis of the HKCSE regarding a more frequent use of this function in native speech is not supported by the customised corpus. While 5.2% of D-use instances of *well* produced by the native speakers of English in the HKCSE are associated with the expression of feelings, only 3.1% are found in the BNC (customised). The difference in expressing emotive feelings between Hong Kong Chinese and British speakers is not statistically significant, though a higher proportion is still observed in native speakers' talk. Again, since there are only relatively few instances expressing this function in the

two sources, more studies need to be carried out to examine whether the linguistic background of speakers has any influence on the emotive use of *well*.

On the other hand, the customised corpus provides strong evidence for the result reported in the HKCSE that native speakers are less prone to use *well* as a processing device. Only 12.6% of discourse tokens of *well* in the BNC (customised) are associated with self-repair and hesitation, compared with 24% found in the speech of non-native Hong Kong Chinese speakers of English. This difference is again very highly significant ($x^2 \geq 82.189$; d.f. = 1; $p < 0.0005$; V = 0.126), confirming the view that non-native speakers are more likely to use *well* when confronted with processing difficulties.

It should be noted that the customised corpus from the BNC contains 140 instances of *well* with unclassified functions, which are not shown in Table 4.21. This is a much larger number compared with the 24 tokens in the HKCSE. The majority of these examples are unclassified due to unclear speech. In the BNC, it is not uncommon to find a whole utterance transcribed as unclear. These unintelligible passages simply render the interpretation of *well* impossible as the preceding context is totally absent, as in example 44:

(44)

N1: (unclear)

N2: ***Well*** she says Michael hasn't rung so she said I assume that he's (pause) coming (pause) cos if he's (pause) going to be late, if he's anywhere (pause) reasonable

(BNC customised, extract from kb8-090601, sn5453, 3411)

Notwithstanding these unclassified instances, the above analysis shows that the customised corpus from the BNC offers further support for the statement that native speakers use *well* as a discourse particle more frequently. In addition, it provides additional evidence that native speakers of English are less likely to employ *well* for processing purposes when compared with non-native speakers in general. There are also slight indications that native speakers may use *well* as a frame more frequently. Other findings observed from the HKCSE need to be compared with more studies in order to reach conclusive results.

4.4.3 Collocations

In the collocational analysis of *well* from the HKCSE presented earlier (see Section 4.3.5.1), it is reported that the six pragmatic functions of *well* co-occur with different lexical items. An examination of the accompanying elements of *well* in the BNC confirms this finding. As a frame marking the introduction of direct speech, *well* is also more likely to follow the reporting verb *say* than any other reporting verb. Table 4.22 presents a comparison of the distribution of the most common reporting verbs before *well* in direct speech between the HKCSE and the BNC (customised):

Table 4.22. The distribution of *well* after the two most common reporting verbs *say* and *think* in direct speech in the HKCSE and the BNC (customised)

Lemma before *well*	**Number of *well* in HKCSE**	**Number of *well* in BNC (customised)**
Say	75 (68.8%)	275 (66.6%)
Think	17 (15.6%)	75 (18.2%)
other reporting verbs	17 (15.6%)	63 (15.3%)
Total	109 (100%)	413 (100%)

As shown in Table 4.22, the distributions of *well* following the reporting verbs *say* and *think* in the two corpora are remarkably similar. The slight differences in the use of these two verbs between the two sources are not statistically significant, indicating that the proportion of the different reporting verbs preceding *well* is fairly constant in both corpora. Similarly, *well* is more likely to occur immediately after the reporting verb in the customised corpus, though there appears to be a greater likelihood of *well* to follow a reflexive pronoun with *think* than with *say*, as in example 45:

(45)

N: … I thought to myself ***well*** if I don't write it down now (pause) cos that's why I got back in my car

(BNC customised, extract from kc9-014307, sn4468, 3140)

This stronger inclination of *think* to be followed by a reflexive pronoun when compared with *say* in the quotative function of *well* is not apparent in the HKCSE. This is possibly due to the relatively small number of instances of *well* with *think* in the data. In addition, there are a small number of tokens of *well* following the lemma

go and its variants in direct speech in the BNC (customised), which is not found in the HKCSE. Example 46 shows the use of *well* after *went* in direct speech quotation:

(46)

N: … You know and they probably went ***well*** yeah, what do we do?
What what exactly are we meant to do.

(BNC customised, jjl, sn648, 14189)

This probably suggests that Hong Kong Chinese are less likely to use the verb *go* together with *well* in direct speech. In fact, no instances of *go* are found in direct speech quotation at all in the speech of Hong Kong Chinese in the HKCSE.

Unlike the co-occurrences with reporting verbs, the combination of *well* with other lexical items in dispreferred responses is found to be less similar in the two corpora. It is mentioned earlier in the chapter that *well* occurs frequently in the patterns of *yes* / *yeah well but* … and *well … no, not* or *don't* in the Hong Kong corpus. In the BNC (customised), although such associations are found, they appear to be less common. Overall, there are only 136 instances of *yes well* and *yea well* in the corpus and 81 occurrences of the adversative conjunction *but* in the immediate surroundings of *well*, as opposed to 348 and 69 tokens of these elements respectively in the HKCSE. In addition, there are 463 uses of *no*, *not* and *don't* in the neighbourhood of *well* out of the 4,379 pragmatic instances of the word, a number which is only slightly higher when compared with the 347 tokens in the 1,913 instances of D-use *well* in the Hong Kong corpus. Given the fact that the number of D-use *well* in the BNC (customised) is more than twice the number in the HKCSE and the proportion of *well* used in dispreferred response is similar in the two data sources, there is a weaker association of *well* with *no, not* and *don't* in the British corpus in achieving the function. A closer look at the examples from the BNC (customised) reveals that instead of simply relying on a rigid linguistic formulation or a fixed lexical item to signal dispreferred responses, British speakers make use of a wider variety of lexical options to express the sense of 'dispreferredness' semantically. In example 47 which takes place in a Q&A session of a company annual results announcement, the speaker N2 uses *well* to signal his disagreement with the assumption made in the question asked by speaker N1, without resorting to words such as *no*, *not*, *don't* and *but* which explicitly convey negative or adversative meaning:

(47)

N1: I suppose one wonders whether the Financial Times feels un uneasy about the proposed reduction in the price of the er, of the Times, as, as no doubt the Independent, probably the, the Telegraph.

N2: ***Well***, I was discussing that with Frank last night and he actually pointed out that it might help the sales of the Financial Times because er, er the FT's very often bought as a second newspaper and if you get your first newspaper cheap, you've got more change in your pocket.

(BNC customised, hyf, sn176, 4499)

In other words, there may be reasons to suspect that although Hong Kong Chinese and native speakers of English similarly associate *well* with dispreferred sequences, these qualified responses might be presented differently by the two speaker groups. While Hong Kong Chinese are more likely to opt for word forms traditionally associated with negative and adversative meaning, the British speakers of English tend to express the sense of 'dispreferredness' more implicitly through the overall semantics of the message. In consequence, the dispreferred aspect of the response might be less overt, making such reactions less face-threatening to hearers.

As for the lexical association of *well* with personal pronouns, the results generated from the BNC (customised) give evidence for the observation from the HKCSE that *well* is frequently followed by a personal pronoun. Of the 4,379 instances of *well* which serve discourse functions, more than 40% of them (n=1,763) are immediately followed by a personal pronoun. This again emphasizes the strong interpersonal dimension related to the use of *well*.

In the previous section while discussing the processing function of *well*, it is found that a lower proportion of *well* is employed for this use in the speech of native speakers. This is verified by the smaller number of filled pauses associated with D-use *well* in the BNC customised corpus. Compared with the 638 instances of *er, erm* and *um* in the proximity of D-use *well* in the HKCSE, only 538 tokens of these elements are found in the BNC (customised), despite a much larger number of pragmatic uses of *well* in the British data. The less frequent lexical combinations of *well* with these fillers possibly indicate that British speakers of English have less processing difficulty to overcome in the interaction as they are conversing in their first language.

4.4.4 Utterance positions

It is noted in the analysis of D-use *well* in the HKCSE that although many instances of the particle occur initially, it is not uncommon to find *well* embedded in the utterance. This is again consonant with findings in the BNC (customised). A comparison of the positional distribution of the discourse uses of *well* in the two sources is presented in Table 4.23:

Table 4.23. The positional distribution of D-use *well* in an utterance in the HKCSE and the BNC (customised)[41]

Position	**D-use *well* in HKCSE**	**D-use *well* in BNC (customised)**
Final	14 (0.7%)	38 (0.9%)
Initial	1,107 (57.9%)	2,728 (64.4%)
Stand-alone	64 (3.3%)	117 (2.8%)
Medial	728 (38.1%)	1,356 (32.0%)
Total	1,913 (100%)	4,239 (100%)

In general, the overall pattern of positional preference is the same in the two corpora: a dominance of initial and medial positions accounts for 96% of all instances of D-use *well*. However, there appears to be a higher proportion of utterance initial *well* and a lower proportion of medial *well* in the BNC (customised). Statistically, the differences are very highly significant (for initial position: $x^2 \geq 23.630$; d.f. = 1; $p < 0.0005$; V = 0.062; for medial position: $x^2 \geq 21.658$; d.f. = 1; $p < 0.0005$; V = 0.059). This contrast in position is investigated in greater detail with a breakdown of the positions in relation to the pragmatic functions of the particle, which is presented in Table 4.24:

[41] The 140 instances of *well* whose functions are unclassified are excluded in Table 4.23 for the sake of consistency in comparison with Table 4.24.

Table 4.24. The positional distribution of the discourse functions of *well* in an utterance in the BNC (customised)[42]

Position (utterance)		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Final	Count	7	1	1	12	13	4	38
	% within Position	18.4%	2.6%	2.6%	31.6%	34.2%	10.5%	100.0%
	% within Function	.5%	.4%	.1%	9.0%	2.4%	1.4%	.9%
Initial	Count	560	190	1493	79	192	214	2728
	% within Position	20.5%	7.0%	54.7%	2.9%	7.0%	7.8%	100.0%
	% within Function	38.2%	72.8%	95.9%	59.4%	36.0%	74.3%	64.4%
Stand-alone	Count	0	0	8	26	21	62	117
	% within Position	.0%	.0%	6.8%	22.2%	17.9%	53.0%	100.0%
	% within Function	.0%	.0%	.5%	19.5%	3.9%	21.5%	2.8%
Medial	Count	900	70	55	16	307	8	1356
	% within Position	66.4%	5.2%	4.1%	1.2%	22.6%	.6%	100.0%
	% within Function	61.3%	26.8%	3.5%	12.0%	57.6%	2.8%	32.0%
Total	Count	1467	261	1557	133	533	288	4239
	% within Position	34.6%	6.2%	36.7%	3.1%	12.6%	6.8%	100.0%
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

From the table, it can be observed that there is a higher proportion of initial *well* for certain functions in the BNC (customised) when compared with the HKCSE (cf. Table 4.9). The differences are especially noticeable for the use of *well* as a link (72.8% vs. 57.1%), a processing device (36.0% vs. 18.5%) and a turn managing signal (74.3% vs. 56.5%). A closer examination of the tokens of *well* from the corpus suggests that the higher proportion of initial positions for these functions may arise from the 'intensive' nature of the interactions in the BNC customised corpus. There appears to be a lower tolerance for silence in the British data, resulting in more instances of initial *well* following incomplete structures. Example 48 illustrates the use of *well* as a linking device in such situations. On the first line of this example, speaker N1 is talking about the money he could have received if there had not been an educational cutback. Speaker N2, his wife, cuts in and comments on his talk

[42] The 140 instances of *well* whose functions are unclassified are excluded in Table 4.24.

before she is abruptly interrupted by her husband. He continues talking on the same topic using *well* to further elaborate on how to use the two hundred and thirty quid:

(48)

N1: I could have done with the two hundred and thirty quid though

N2: *Well* I know, but er

N1: ***Well*** that'll pay the ferry

(BNC customised, kdr-071101, sn907, 1855)

In an analogous scenario in the Hong Kong data, speakers are more likely to be given enough time to finish their turn before other interlocutors intervene with their own comments. Similarly, more instances of initial *well* are found in processing and turn managing functions in the BNC (customised) because speakers are likely to initiate talk when there are periods of silence in the interaction or when the other participant is just pausing for a breath, even though they are hesitant about what to say. The increase in the ratio of utterance initial *well* indirectly reduces the share of *well* embedded in discourse, accounting for the difference in proportion of positions observed in the two corpora. The influence of culture and context might be a driving force in this lower tolerance for silence. As a collectivist cultural group, Chinese may be more at ease with silence in talk when compared with British who are more individualistic (see, for example, Goodwin 1999; Gudykunst 1994; Hofstede 1994). In addition, since most of the interactions are intercultural in the Hong Kong corpus, participants may display a higher level of patience and allow for silence more readily, as they are aware that some speakers are not communicating in their mother tongue. In addition, the familiarity level between the participants may also be a determining factor. In the BNC (customised), more interactions are carried out between speakers whose relationship is characterised by a high level of intimacy, such as the husband and wife shown in example 48. As a result, they might be more at ease in initiating talk even if the other parties have not finished their turn, as the maintenance of face is of less concern in such circumstances.

4.4.5 Domain and text type

In the analysis of *well* from the HKCSE, it is argued that the contextual settings affect how the particle is used. In general, results from the BNC (customised) corroborate the above statement, though some variations are noticeable between the

two sources. Table 4.25 lists the functional distribution of *well* in the four domains of the BNC (customised):

Table 4.25. The functional distribution of D-use *well* by domain in the BNC (customised)

Domain		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Academic	Count	227	28	204	4	115	38	616
	% within Domain	36.9%	4.5%	33.1%	.6%	18.7%	6.2%	100.0%
Business	Count	381	65	396	44	110	78	1074
	% within Domain	35.5%	6.1%	36.9%	4.1%	10.2%	7.3%	100.0%
Conversational	Count	696	160	753	79	277	165	2130
	% within Domain	32.7%	7.5%	35.4%	3.7%	13.0%	7.7%	100.0%
Public	Count	163	8	204	6	31	7	419
	% within Domain	38.9%	1.9%	48.7%	1.4%	7.4%	1.7%	100.0%
Total	Count	1467	261	1557	133	533	288	4239
	% within Domain	34.6%	6.2%	36.7%	3.1%	12.6%	6.8%	100.0%

A number of similarities can be observed between the two corpora in terms of the functional distribution across the four domains (cf. Table 4.17). For instance, the public sub-corpus of the BNC (customised) has the highest proportion of responsive *well* (48.7%), supporting its importance as a reactive strategy in media interviews. In addition, conversations contain the highest number of *well* (n=79) used in the expression of feelings, thus reinforcing the idea that sharing an affective bond is a core part of informal casual talk. However, some differences are perceivable between the British corpus and the Hong Kong corpus. For instance, there is a much higher proportion of responsive *well* and a lower ratio of the particle as a frame in the academic domain of the BNC (customised) than the HKCSE. This is largely due to the fact that lectures and seminars are more interactive in the British data. In the teaching sessions of the customised corpus, students are more likely to raise questions and give feedback to the teacher, leading to a rise in the frequency of *well* in responses. In comparison, similar text types in the HKCSE are mostly monologic and student input is rare. Consequently, there is a higher proportion of *well* in response and a lower fraction of *well* in its framing function in the BNC (customised). Another interesting observation is related to the percentage of *well* as a frame in the public domain. While 22.8% of *well* have a framing function in the public sub-corpus

of the HKCSE, the percentage goes up to 38.9% in the same domain of the BNC (customised). A possible reason is the variation in the nature of public speeches. In the Hong Kong corpus, most of the speeches in the public domain are heavily scripted. As a result, the use of *well*, or discourse particles for that matter, is less frequent. Since speakers are virtually reading aloud a fully-prepared monologue, there is less need for textual organisation and structuring when compared with a more spontaneous text. On the other hand, there are more semi-scripted speeches in the BNC, leading to a greater likelihood of the particle being employed to divide a long stretch of talk for ease of comprehension.

Another way to look at the interplay between functions and settings is to investigate the functional distribution of *well* across various text types. As is the case in the HKCSE, this reveals the functional preferences of *well* for monologic and dialogic texts. Table 4.26 illustrates the functional distribution of *well* in some major monologic and dialogic text types in the BNC customised corpus:

Table 4.26. The functional distribution of D-use *well* across some major text types in the BNC (customised)

Text Type			Function						Total
			Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Dialogic	conversation	Count	696	160	753	79	277	165	2130
		% within Text Type	32.7%	7.5%	35.4%	3.7%	13.0%	7.7%	100.0%
	interview	Count	241	26	307	22	67	35	698
		% within Text Type	34.5%	3.7%	44.0%	3.2%	9.6%	5.0%	100.0%
	meeting	Count	61	12	74	9	20	10	186
		% within Text Type	32.8%	6.5%	39.8%	4.8%	10.8%	5.4%	100.0%
Monologic	lecture	Count	63	5	12	1	20	1	102
		% within Text Type	61.8%	4.9%	11.8%	1.0%	19.6%	1.0%	100.0%
	speech	Count	54	1	5	1	7	0	68
		% within Text Type	79.4%	1.5%	7.4%	1.5%	10.3%	.0%	100.0%
Sum		Count	1115	204	1151	112	391	211	3184
		% within Sum	35.0%	6.4%	36.1%	3.5%	12.3%	6.6%	100.0%

Similar to the observations made in the HKCSE (see Table 4.18), multi-party texts display a preference for the interpersonal functions of *well*, highlighting the importance of maintaining social relationships in these discourse events. On the other hand, monologic text types such as lectures and speeches contain an overwhelming proportion of *well* as a frame, testifying to its central role in organising and structuring discourse (cf. Stenström 1990). The negligible number of *well* for turn management in the monologues, in addition, is parallel to findings in the HKCSE. Collectively, these results from the BNC (customised) give further support that the contextual setting is an influential variable in the pragmatic functions of *well*.

4.4.6 Summary of findings and comparisons between the HKCSE and the BNC (customised)

As shown by the analysis above, the comparison of the findings from the customised corpus of the BNC with those from the HKCSE yields interesting results. On the one hand, the overall findings from the two sources are largely comparable. The higher discourse-function ratio and discourse rate of *well* in the BNC (customised) confirm the results from native speaker data in the HKCSE, indicating that D-use *well* is a more common phenomenon in the speech of native speakers. In addition, the functional distributions of D-use *well* in the two corpora are generally similar, with framing and responsive functions being the most prominent uses. In particular, the observation that Hong Kong Chinese are more likely to use *well* as a processing device than native speakers is substantiated by the British data, with a very highly significant difference shown in the comparison.

The examination of the linguistic and contextual features of *well* in the BNC (customised) also points to the fact that these variables interact with the pragmatic functions of *well*. On the whole, the positional preferences of D-use *well* are similar in the two corpora and some key collocational patterns are observed in both sources. Further, the study of the contextual settings of the BNC (customised) in relation to the discourse functions supports the findings from the HKCSE that the four domains display varying functional preferences of D-use *well*. The interactive level of text types, above all, determines whether the pragmatic functions of *well* are more inclined to the textual domain or the interpersonal and interactional domains.

On the other hand, different findings can be observed from the two corpora, indicating some minor discrepancies in the sources of data. In particular, cultural and contextual variations are suspected to give rise to the slight differences found in the use of D-use *well* between the two speaker groups, with a lower tolerance for silence and a wider range of strategies expressing dispreferredness in the British data. Finally, the differences in the composition and nature of texts between the two corpora also attribute to the small degree of variations in the functional distribution of *well* between the HKCSE and the BNC customised corpus.

4.5 Comparison of findings in authentic speech and textbooks

In this section, the discourse use of *well* in the textbook database will be examined and compared with the findings from the HKCSE and the BNC (customised). In particular, the analysis will focus on three aspects of D-use *well* to study the major similarities and differences between authentic usage and textbook descriptions, namely frequency of occurrence, positions and functions. While the pragmatic uses of *well* in the sample section which contains presentations and discussions in the textbooks will be studied both qualitatively and quantitatively, the examination of the particle in the teaching section which mostly contains short and detached examples and teaching points will be qualitative.

4.5.1 Frequency of occurrence

One of the most striking differences between authentic data and sample texts in the textbook database in relation to the discourse use of *well* is the rate at which it appears in discourse. Table 4.27 compares the use of *well* in the three sources under investigation: the textbooks, the HKCSE and the BNC (customised):

Table 4.27. The comparison of *well* in the textbooks, the HKCSE and the BNC (customised)

	Sample presentations in textbook database (N=38)	**Sample discussions in textbook database (N=11)**	**HKCSE (N=311)**	**BNC customised (N=258)**
Total number of words	10,817	13,032	949,972	948,920
Total number of *well*	9	117	2,714	5,404
Number of D-use	1	107	1,913	4,379
Number of P-use	8	10	796	1,021
Unclassified use	0	0	5	4
D-use / total use (%)	11.11	91.45	67.99	81.03
D-rate (per 10,000 words)	0.92	82.11	19.56	46.15

A number of important observations can be noted from the above table regarding the particle *well*. Firstly, there is a considerable difference in terms of the total number of *well* in the two sample text types in textbooks. While the total number of words in textbook discussions (n=13,032) is only 20% more than that in textbook presentations (n=10,817), the number of *well* in discussion texts is twelve times more than that in presentations (n=117 vs. n=9). This indicates that the word occurs much more frequently in textbook discussions. In addition, there is a huge discrepancy in the discourse-function ratio of *well* between the two sample text types. While only 11.11% instances of *well* are used as particles in presentations, the percentage rises up to 91.45% in textbook discussion texts. As shown from the number of D-use and P-use in the table, only one out of nine instances of *well* conveys discourse meaning in textbook presentations. In contrast, there are 107 tokens of D-use *well* in 117 occurrences of the word in discussion texts. Unsurprisingly, this contrast in the ratio of D-use to P-use *well* between the two text types in textbooks achieves very high statistical significance with a strong association ($x^2 \geq 44.054$; d.f. = 1; $p < 0.0005$; V = 0.591), giving a strong impression that when *well* is used it is most likely to be a particle in textbook discussions and vice versa in presentations.

Since a noticeable contrast is found in the number of D-use *well* in presentations (n=1) and discussions (n=107), the discourse rate of *well* in the two text types varies greatly. In a sample of textbook presentations of 10,000 words, the discourse rate of *well* is a mere 0.92. On the other hand, textbook discussions with the same sample size have a discourse rate of 82.11 for the particle. This signifies that D-use *well*

occurs considerably much more frequently in textbook discussions than in textbook presentations.

When compared with authentic data, the use of *well* in textbooks also shows highly contrastive patterns. As regards the discourse-function ratio, textbook discussions have the highest figure among the four data sources while the ratio in textbook presentation texts is unusually low. A similar pattern can be observed concerning the discourse rate of *well*, with the highest rate found in textbook discussion texts and the lowest in textbook presentations. These findings appear to suggest an irregularity in the discourse rate of *well* in textbooks when compared with authentic data. This is substantiated by a comparison between textbook sample texts and their corresponding text types in the two corpora. Given the rather formal and overwhelmingly monologic nature of sample presentations in textbooks, they can be compared with business presentations, lectures and speeches in the HKCSE and the BNC (customised). For sample discussions which involve a clear agenda and the negotiation of ideas, business interviews, meetings and public TV talk shows in the two corpora are chosen for contrast. Table 4.28 shows the comparison of the discourse rate of *well* in some monologic texts in the HKCSE, the BNC (customised) and the textbooks:

Table 4.28. The comparison of the discourse rate of *well* in presentations in the HKCSE, the BNC (customised) and the textbook database

Data source	**Text type**	**Discourse rate of *well* (per 10,000 words)**
HKCSE	business presentation	5.47
	lecture	12.20
	speech	5.07
BNC	business presentation	3.62
	lecture	17.19
	speech	7.63
Textbook database	sample presentation	0.92

Admittedly, there is no equivalent text type in authentic data which strictly parallels the sample texts in textbooks which are constructed for exam situations. Nevertheless, a comparison of these sample texts with naturally-occurring speech still indicates the extent to which the textbook data diverge from real-life examples. From the table

above, it is quite clear that presentations in textbooks do not match their corresponding text types in authentic talk in terms of the discourse rate of *well*. Even in heavily scripted business presentations and public speeches in the two corpora, the discourse rate of *well* is at least more than three times that in the textbook presentations. The difference is more apparent if textbook presentations are compared with lectures in the two corpora, which are more spontaneous in nature. From the analysis of *well* in monologues in the two corpora earlier in the chapter, it is shown that the particle serves important pragmatic functions, especially in the textual domain as regards discourse organisation. There is thus no reason why textbook presentations should have such a low discourse rate of *well* if they mirror authentic talk.

Interestingly, the contrast in the discourse rate of *well* between textbook discussions and the corresponding text types in the two corpora shows exactly opposite results. Table 4.29 compares the discourse rate of *well* in discussion texts in the HKCSE, the BNC (customised) and the textbooks:

Table 4.29. The comparison of the discourse rate of *well* in discussions in the HKCSE, the BNC (customised) and the textbook database

Data source	**Text type**	**Discourse rate of *well* (per 10,000 words)**
HKCSE	business interview	22.13
	business meeting	30.33
	public TV talk show	34.65
BNC	business interview	50.62
	business meeting	46.61
	public TV talk show	30.70
Textbook database	sample discussion	82.11

The above table shows that the sample discussions in textbooks have a much higher discourse rate of *well* when compared with similar text types in authentic speech. The rate is even higher than the ones in the speech of native speakers, who in general use the particle more frequently than Hong Kong Chinese. This suggests that there is an unusually high rate of D-use *well* in textbook discussions when compared with similar texts in corpus data. Again, it is not clear why the particle *well* should be used so frequently in these textbook discussions if they reflect authentic usage.

4.5.2 Positions

Apart from the differences shown in the frequency of D-use *well* between textbooks and authentic data, there are also positional variations regarding the pragmatic functions of the particle. As discussed earlier in the chapter, although utterance initial *well* constitutes more than half of the pragmatic examples in both corpora, a significant proportion of occurrences of D-use *well* are embedded in discourse. In the textbooks examined, however, this is not the case. Table 4.30 gives a comparison of the positional distribution of *well* as a discourse particle in an utterance in the three sources of data:

Table 4.30. The positional distribution of D-use *well* in an utterance in the HKCSE, the BNC (customised) and sample texts in the textbook database

Position	**D-use *well* in HKCSE**	**D-use *well* in BNC (customised)**	**D-use *well* in sample texts in textbook database**
Final	14 (0.7%)	38 (0.9%)	0 (0.0%)
Initial	1,107 (57.9%)	2,728 (64.4%)	97 (89.8%)
Stand-alone	64 (3.3%)	117 (2.8%)	0 (0.0%)
Medial	728 (38.1%)	1,356 (32.0%)	11 (10.2%)
Total	1,913 (100%)	4,239 (100%)	108 (100%)

Of the 108 instances of D-use *well* found in sample presentations and discussions in textbooks, there are 97 instances occurring at the beginning of an utterance, making up 89.8% of the total pragmatic use. Example 49 shows an utterance initial *well* in a sample discussion text in textbooks:

(49)

T1: … Anyway, my uncle has satellite TV, and we can all go over to his flat and watch it ... or I can just record it on the VCR.

T2: ***Well***, look, let's try to summarise what's been said so far. It seems that our theme should be what Asian singers and what styles of Asian pop music are internationally popular. Does everyone agree?

(textbook database, sample discussion 5.1, 1036)

In contrast, only eleven tokens (10.2%) are found medially, making them a small minority in the database. An utterance medial D-use *well* is shown in example 50:

(50)

T: O.K. What's next? An airport ... ***well***, they'll be coming by plane, so they'll see the new airport when they arrive...

(textbook database, sample discussion 3.2, 716)

In addition, no instances of D-use *well* are found in final and stand-alone positions in textbooks, although a small number can be found in the two corpora. As regards the two main positions which D-use *well* occupies, a comparison of the three columns of Table 4.30 indicates that there is a much higher proportion of initial D-use *well* and a much lower proportion of medial D-use *well* in textbooks when weighed against the two corpora. While 89.8% of discourse *well* are in initial position in sample texts in textbooks, the percentages are only 57.9% and 64.4% in the HKCSE and the BNC (customised) respectively. In medial position, merely 10.2% of D-use *well* can be found in textbooks, while the proportions increase to 38.1% and 32.0% respectively in the HKCSE and the BNC (customised). This high percentage of discourse *well* in initial position in textbooks appears to disperse the notion that initial D-use *well* is vastly dominant, which unfortunately is not validated by the findings from the two corpora.

4.5.3 Functional analysis

In this section, the pragmatic functions of *well* as presented in textbooks will be examined in detail. Firstly, a purely qualitative analysis of the description of the discourse uses of *well* in the teaching section of the textbooks will be provided, followed by a quantitative study of the functional distribution of D-use *well* in the sample section in textbooks.

In the teaching section of the textbook database, the focus of the teaching materials is often on how to communicate effectively in presentations and discussions, as these are the scenarios students have to face in the spoken component of the public examination. As these textbooks are mainly designed to achieve the goal above, the teaching materials are recurrently filled with examples of speech functions which are common in presentations and discussions. Example 51, for instance, shows how the speech function “disagreeing with a suggestion” is claimed to be expressed by the following utterance (Sutton and Duncan 1999:64):

(51)

T: ***Well***, I don't think that's possible, because...
(textbook database, teaching section, 7939)

In other words, there are no paragraphs or sections which discuss the discourse functions of *well* separately. The various uses of the particle can only be found in examples showing how different speech functions are realised in the textbooks, which are all short, detached examples of language totally void of contextual information such as example 51. An analysis of these instances of *well* shows that all speech functions associated with *well* in textbooks could be subsumed under the functional categories identified in the two corpora. No new functions which are absent in authentic talk are found in the teaching materials. Accordingly, these instances of *well* in the teaching section are examined in the textual, interpersonal and interactional functional domains according to the speech functions they are suggested to be associated with by the textbook writers. This examination serves to compare the teaching of D-use *well* in textbooks with their use in authentic speech.

In the textual domain, *well* is found in examples related to the maintenance of the overall flow of presentations and discussions. This includes initiating a topic in a presentation, starting and ending a discussion and moving on to the next point in a discussion. Example 52 illustrates the use of *well* as an initiator in a textbook discussion text:

(52)
T: ***Well***, shall I start? I suggest we divide our discussion into three parts…
(textbook database, teaching section, 16207)

These examples in the textbooks highlight the discourse use of *well* in topic and discourse stage management and largely correspond to the frame function identified in the corpora. The use of *well* as a linking device, however, is not discussed in the teaching section of the textbook database. This possibly conveys the idea that *well* can only be used to divide parts of discourse but not in connecting them, which is actually not the case in authentic data.

As regards the interpersonal aspect of *well*, the teaching section contains a large number of speech functions associated with responses. Examples illustrating how disagreements and qualified agreements should be expressed are especially abundant. In one of the textbooks where *well* receives a brief mention, it is suggested that *well*

is one of the lexical items used to "soften disagreement and doubt" (Potter 2003a:42). However, it is not apparent how the word is used to achieve this function as no examples are given in the textbook. Other kinds of dispreferred responses associated with *well* in the teaching section include agreeing but not in a very enthusiastic way (Esser 1999) and "avoiding giving an opinion" (Sutton and Duncan 1999:56). In addition, *well* is found as part of a textbook example which serves to illustrate "disagreeing strongly with someone" (ibid.), reproduced here as example 53:

(53)

T: ***Well***, I really don't agree at all!

(textbook database, teaching section, 7475)

Although *well* is frequently seen in dispreferred responses in the two corpora, examples like the one above are not found in the authentic texts. In fact, the word combination *I really don't agree at all* is not present in either the HKCSE or the BNC (customised). This casts doubts as to whether the linguistic realisations associated with disagreements in textbooks reflect those in authentic data (Cheng and Warren 2005).

In the description of how disagreements or contrastive ideas should be realised, one of the textbooks discusses the intonation of *well*, which is the only occasion where prosody is given any attention in the teaching materials. The textbook writer claims that a rise-fall-rise *well* is used "to show that the previous idea is not really true" (Potter 2003a:14). However, this statement is not supported by the prosodic information in the HKCSE. In fact, instances of fall-rise *well* are infrequent in the corpus. In the authentic texts, the particle is more likely to carry fall or level tone to achieve this function if it is prominent.

While the responsive use of *well* abounds in the teaching section, no instances of discourse *well* are found to be associated with affective meaning. In fact, no speech functions in the teaching section are concerned with the expression of feelings, possibly suggesting that textbook writers do not find emotions a necessary component in presentations and discussions.

As far as the interactional aspect of *well* is concerned, the uses of *well* as a processing device and as a turn managing signal are both found in the teaching section. It is suggested that the particle is one of the "hesitation words" along a wide range of lexical items such as *er*, *in fact* and *let me see* which students could employ

when they need time for planning what to say (Potter 2003a:25). With regard to turn management, *well* is associated with the speech function "interrupting politely" (Potter 2003b:39). This corresponds to the turn-taking sub-type of *well* identified in the two corpora. Again, examples given in the textbook teaching section are decontextualised and it is unclear how the particle assists in expressing the function of taking the conversational floor politely, as the preceding utterance is not provided. A case in point is example 54, which is given in a textbook to illustrate how to interrupt without being rude:

(54)

T: ***Well***, I think the solution is simple. We should ...

(textbook database, teaching section, 12130)

When compared with the textbook teaching section, the sample section generally provides more contextual information as the sample presentations and discussions have structures resembling authentic text types of a similar nature. As a result, the functions of the particles can be examined together with the linguistic co-texts in the samples. For the functional analysis of D-use *well* in the sample section, all instances of *well* could be categorised according to the six pragmatic functions identified in the two corpora. No new functions are found. Table 4.31 shows the functional distribution of D-use *well* in the sample section of the textbooks:

Table 4.31. The functional distribution of *well* in the sample presentation and discussion texts in the textbook database

Text type in textbooks		Function						Total
		Framing	Linking	Responsive	Emotive	Processing	Turn managing	
Sample presentation texts	Count	1	0	0	0	0	0	1
	% within Function	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Sample discussion texts	Count	26	1	74	2	3	1	107
	% within Function	24.3%	0.9%	69.2%	1.9%	2.8%	0.9%	100.0%

As discussed in the previous section regarding the frequency of *well* in the textbooks, only a single instance of D-use *well* is found in sample presentations, acting as a frame in the text. For sample discussions, the use of *well* in response constitutes 74 out of a total of 107 instances. As the most dominant pragmatic function in textbook

discussions, responsive *well* makes up more than two-thirds (69.2%) of the total. The framing use of *well* contributes another 24.3%. The remaining small proportion (6.5%) of use is shared between the other four functions.

A major difference in the use of *well* between textbooks and authentic data is the distribution of functions. While it is true that framing and responsive *well* are two key pragmatic functions found in the two corpora, the other four functions also add up to a reasonable proportion in the data. In the textbooks, however, the other functions only contribute a very small number of instances, making up less than 7% of the total. Specifically, the responsive function of *well* appears to be over-emphasized in textbooks while the frame function is slightly overlooked when compared with findings from the HKCSE and the BNC (customised). This is reflected by the proportions of framing and responsive *well* in sample discussions and the negligible number of *well* in sample presentations for textual organisation.

4.6 Pedagogical implications

Despite the fact that *well* is one of the most frequently occurring words in the spoken language and many of the 15 textbooks collected for the present study claim to focus on oral skills, none of them assigns a separate section or paragraph to the description of the particle, not to mention a discussion of discourse particles as a group. This is disappointing given the importance of *well* to express various textual, interpersonal and interactional functions in talk, as evidenced by the findings reported from the two corpora of authentic texts. From the analysis of *well* in the textbooks above, it is apparent that noticeable differences exist in terms of the frequency of occurrence, positions and functions of *well* between textbook data and naturally-occurring speech. The huge difference in the discourse rates of *well* between textbook presentations and discussions gives a strong impression that the particle is virtually non-existent in presentations while being ubiquitous in discussions. This, however, is not substantiated by the corpus evidence in the present study. While most of the instances of *well* found in textbooks occur in utterance initial position, corpus evidence suggests that the particle is also fairly common in medial position. In addition, the functions of *well* as described and realised in textbooks do not seem to be a close match to their functions in authentic talk. The pragmatic functions of linking and emotive *well* are not mentioned in textbooks and the responsive use of

well appears to be over-stressed in sample texts. This possibly leads to an overuse of *well* in responses if students follow the examples from the textbooks.

A more fundamental problem, however, lies in the fact that many of these tokens of *well* are placed in short, detached examples with minimal linguistic and contextual background information. The utterances prior to and right after the suggested examples are not provided in the teaching section and no explanations are offered regarding how the examples express a particular speech function. The linguistic realisations of some functions in textbooks also show divergence from authentic data. This makes one wonder to what extent the textbooks being examined paint an accurate picture of the discourse use of *well* for foreign language learners who have little exposure to spoken interactions in English.

Summary

This chapter carries out a detailed analysis of *well* by comparing its use in three different sources of data. It starts by reviewing some previous studies of the particle and then moves on to discuss how the discourse use and propositional use of the word are distinguished in the present study. It then presents an in-depth examination of the use of *well* in the HKCSE. The analysis demonstrates that the positions and prosodic profiles of *well* are determining factors in deciding whether the word conveys discourse or propositional meaning. Six key pragmatic functions are identified in the Hong Kong corpus, showing different linguistic, sociolinguistic and contextual variations.

In order to investigate in detail the influence of linguistic background of speakers on the use of *well*, findings from the HKCSE are then compared to those from a customised corpus from the BNC. The comparison shows largely parallel results as reported in the HKCSE. Two major differences in the use of *well* between native and non-native speakers of English in this study are related to the frequency of use of the particle and its functional distribution. The discourse use of *well* is more frequently observed in the repertoire of native speakers and they are less likely to use *well* as a processing device when compared with Hong Kong Chinese. The exploration of the British data also reveals differences in terms of the composition of the two corpora, which in turn affects the use of *well* as a particle.

Finally, the descriptions and occurrences of *well* in a textbook database are inspected to see whether teaching materials reflect authentic uses of *well* from corpus

data. Results from the comparison show that textbook descriptions at best provide a partial and decontextualised account of the particle. At worst, these teaching materials could be misleading as the sample texts show atypical discourse rates of *well* when compared with similar text types in the two corpora. This leads to the conclusion that the textbooks being examined in the present study fail to give a genuine profile of the particle *well*, which might prevent English language learners from knowing how this highly frequent word is actually used.

Chapter 5

Study of *so*

Abstract

This chapter presents the analysis of *so* in the three sources of data being examined. It first describes some previous studies of *so* which provides background for the present study (5.1). It then discusses with examples the discourse use and propositional use of *so* (5.2). This is followed by a detailed analysis of *so* in the HKCSE which specifically looks at various dimensions, including the overall frequency of occurrence of the particle, its positional distribution, its prosodic features and the pragmatic functions it serves in the corpus. The analysis will also investigate a number of linguistic, sociolinguistic and contextual factors which may have an effect on the use of *so* (5.3). Findings generated from the HKCSE will be compared with results from the BNC customised corpus to determine the extent to which the conclusions drawn regarding the use of *so* from the Hong Kong corpus are valid (5.4). Subsequently, the chapter examines the use of *so* in the textbook database and compares the descriptions in teaching materials with findings from authentic corpus data (5.5). This leads up to the last section of the chapter, where pedagogical implications arising from the study of *so* are proposed (5.6).

5.1 Previous studies of *so*

The monosyllabic word *so* is one of the most frequently occurring words in the English language. In the CANCODE spoken corpus which consists of five million words, *so* is the 19th most frequently occurring item, with a frequency of 40,071 tokens (O'Keeffe et al. 2007). In the British National Corpus, *so* ranks 33rd in the frequency list of the spoken section and 47th of the written section, suggesting that it has a comparatively higher frequency in speech than in writing (Leech et al. 2001). Despite its ubiquity, relatively few particle studies have focused on the linguistic item *so*. For instance, it is not included in the list of "interactional signals and discourse markers" in Stenström (1994:59). This probably reflects the general perception that *so* is more a peripheral member of the class of discourse particles (Müller 2005), if such a category arguably exists. This section reviews some of the major approaches to the study of *so*. In particular, it discusses how the key discourse

functions of *so* are described in previous studies from various perspectives. It also highlights some aspects of *so* which are relevant to the present study but have not been fully explored, including the prosody of the word and its use in the speech of non-native speakers of English. For the purpose of the present study, this review only focuses on earlier studies of *so* as a discourse particle. Its propositional use, which is extensively described in grammar books and dictionaries (see, for example, Carter and McCarthy 2006; Quirk et al. 1985), will not be discussed in this review section.

The study of *so* can be traced back to the influential account of discourse connectives in terms of the Gricean notion of conventional implicature. In Grice's (1989) view, *so* is one of the non-truth-conditional discourse connectives which are used to perform higher-order speech-acts, i.e. to signal how speakers comment on some more basic or lower-order speech-acts. Specifically, *so* is associated with the higher-order speech-act of "explaining" (Grice 1989:362). As this conventional meaning is linguistically encoded in *so*, it follows that the speech-act of explaining is performed every time the word is used, regardless of the context in which it occurs. This view highlights the conventional connective function associated with *so* but fails to take into account the specific meaning potential which could be indicated by the linguistic item in different contexts.

Closely related to the Gricean analysis of *so* as a connective is the study of *so* under relevance theory. Within a relevance-theoretical approach, Blakemore (1988) considers *so* a constraint on relevance. She argues that the main function of *so* is to guide the listener "to establish an inferential connection" (Blakemore 1988:193). She further discusses four uses of *so*. In conjoined utterances, *so* connects two propositions together and suggest either a causal effect or a deductive consequence between them. When *so* occurs initially without an explicit linguistic antecedent, it signals that the utterance it prefaces is relevant to the situational context at hand. When *so* is used in responses as a stand-alone utterance or as in *so what*, it implies that the speaker is "unable to see the significance of what someone has said" (Blakemore 1988:189). By imposing a constraint on relevance, *so* minimizes the processing effort involved in interpretation, and thus enhances the efficiency of the cognitive system (Sperber and Wilson 1995).

For Schiffrin (1987), *so* conveys a central meaning of result. Studying *so* together with *because* under the notion of coherence, Schiffrin gives a comprehensive account of *so* in three aspects: a complementary marker of main idea units, a marker of

resultative relations, and a turn-transition device which signals a potential speaker change. Similar to Blakemore (1988), Schiffrin identifies different sub-types of resultative relations. In her framework, *so* indicates three kinds of results: fact-based, knowledge-based and action-based. The use of *so* in fact-based and knowledge based relations largely corresponds to Blakemore's (1988) discussion of *so* in causal effect and deductive consequence. In action-based relation, *so* signals an action which has just been motivated by the preceding context. She further discusses three types of actions which are prefaced by *so*, namely requests, compliances and claims, though she only gives an example of *so* introducing a request for information. As admitted by Schiffrin (1987) in the latter part of the chapter, the distinction between the three types of consequential relations is not always clear. Multiple readings are possible. Owing to "our understandings of causality" (Schiffrin 1987:211), it is often difficult to determine whether conclusions drawn are purely based on objective facts and states of affairs or on our interpretations of those facts in accordance with our knowledge base as well as our cultural and personal preference.

In a similar fashion, Redeker (1990) also examines the role of *so* within the framework of coherence. In her study, *so* operates in two aspects of discourse structure: ideational and pragmatic. As a marker of ideational structure, *so* signals a resultative or consequential relation (cf. for example, Schiffrin 1987). As a marker of pragmatic structure, *so* is used to indicate a sequential relation between "successive elements in a chain of events" (Redeker 1990:373) or to preface a conclusion made by the speaker. From her examples, however, it is not exactly clear how the pragmatic use of *so* contributes to the expression of attitudes and intentions as she suggests. In a later work, Redeker (2006:339) describes the use of *so* among other markers as "attentional cues" and focuses on its function in marking transitions in discourse, though only turn-internal instances are examined.

In his attempt to categorize discourse markers into different groups, Fraser (1988: 31) includes *so* as an example of inferential markers "which signal that the current utterance conveys a message, which is, in some sense, consequential to some aspect of the foregoing". He later explains with a number of examples that this consequence which follows from the prior context is not confined only to the narrow sense of result, but is extended to conclusions derived from the preceding information or questions arising from the previous message (Fraser 1990). Given that *so* allows for a range of interpretations, Fraser (2006) concludes that the specific meaning of *so* is

relatively opaque and hence it does not have a core meaning. His view is contrary to Schiffrin's (1987) proposal that *so* may derive other meanings from its core resultative sense.

In terms of the prosodic profile of *so*, there is a dearth of research in this area in particle studies. In the literature, it appears that no particle studies look at the intonational patterns of *so*. The scant number of research studies related to this aspect does not seem to do justice to the high frequency of this linguistic item in discourse. Some brief descriptions of the prosodic contour of *so*, however, could be found in grammar books. In *A Grammar of Contemporary English* which is based on findings from the *Survey of English Usage*, *so* is said to "usually occur without intonation or punctuation separation from what follows" (Quirk et al. 1972:527). However, when it is used in utterance initial position to mark an inference made by the speaker based on the preceding linguistic context, *so* often constitutes an independent tone unit (Quirk et al. 1972). Nevertheless, no tonal patterns have been suggested to be associated with this specific function of *so*.

As regards the use of *so* in the speech of non-native speakers of English, the small number of related studies seems to show contradictory results concerning the frequency of use of this linguistic item by non-native speakers. He's (2002) study shows that there is an overuse of *so* in the speech and writing of Chinese EFL speakers. By using a number of sources including spoken and written corpora of British English and Chinese learner English supplemented by English textbooks and Chinese conversations, He (2002) investigates the frequency, position and collocation of *so* and compares these parameters in her data. The research findings show that the word *so* is more frequently used by Chinese EFL learners than by native speakers of English. Following this result, she gives possible reasons for the overuse and misuse of *so* by Chinese speakers, including unawareness of stylistic difference in spoken and written English, limited exposure to English, textbook influence and negative transfer from learners' first language. In particular, she points out that *so* is taught in the early stage at school and its use as a connective is prevalent in middle school English teaching materials. As far as the impact of mother tongue is concerned, He (2002) examines the Chinese equivalent of *so* in conversations. She maintains that similar to *so*, the Chinese correspondence *gum* is multi-functional and also undergoes the process of semantic bleaching while developing new functions. While her study provides useful comparative findings

especially concerning Chinese speakers' use of *so*, it only discusses "the connective use of *so* with meanings similar to *so that* ('with the result that'), *thus* ('in this way') and *therefore* ('for that reason')" (He 2002:43). This significantly restricts its comparison of the discourse uses of *so* in the two speaker groups.

Apart from Chinese speakers, the use of *so* is also examined in the speech of German speakers. In a paper comparing the use of discourse markers by American and German students based on 70 conversations collected from a movie re-telling experiment, Müller (2004) finds that the German EFL speakers use *so* much less frequently when compared with American speakers. She suggests that the small number of occurrences of *so* in beginner's textbooks may be one of the reasons for the findings. Interestingly, in her study, the influence of first language does not lead to a negative transfer as in He (2002). Instead, German students underuse *so* because of its phonological and semantic similarity to the German word *also* in order to avoid the German-sounding association when speaking English. In Müller (2005), the difference in the discourse functions of *so* between the two groups of speakers is described in more detail. Specifically, the sequential, resultative and the summarising functions of *so* are found to show statistically significant differences in usage between American speakers and German speakers. In every case, it is the American speakers who use the function significantly more than the German speakers. The most significant difference is found in the use of *so* to signal a sequential relationship. Compared with the Germans, the Americans use this function eight times more often. Müller (2005) thus concludes that the difference in the use of *so* between the two speaker groups is greatest on the textual level.

In sum, the review of previous studies above shows that although the word *so* has been approached from different angles, most studies concur that its association with inference or consequence constantly forms an integral part of the interpretation of its meaning. With regards the prosodic pattern of the word and its non-native usage, however, the limited number of research studies prohibits any definitive conclusions from being drawn. In particular, the prosody of *so* is severely understudied. In the following sections, the findings from the present study will be compared with those reported earlier whenever possible, while at the same time highlighting the prosodic pattern of *so* and the use of *so* by non-native speakers, which are the two areas that remain to be thoroughly investigated.

5.2 Distinguishing the discourse use and propositional use of *so*

This section discusses some examples of the propositional use and discourse use of *so* in the data. When *so* acts as a modifying adverb, a substitute form, a conjunction of purpose or as part of a fixed phrase, it contributes to the propositional content of the utterance. As an adverb, *so* modifies a following adjective or another adverb. It is sometimes referred to as "adverb of degree or manner" (Müller 2005:68). In this role, *so* serves an intensifying function and gives extra emphasis to the constituent it modifies. In example 1, *so* modifies the adjective *upset* and underlines the emotional aspect of the utterance:

(1)

A: { \ and i was [***SO***] < upSET > } { = at [< THAT >] time }

(HKCSE, C011, 4455)

In example 2, *so* modifies the adverb *very*, which in turn functions as a modifier of the adjective *expensive* itself:

(2)

B: …{ \ of course the [PROblem] is it's ***SO*** very very < exPENsive > }…

(HKCSE, A005, 6640)

Its most common emphatic use in the corpora, however, is its occurrence in front of the words *many* and *much* to refer to a large quantity, amount or degree, as in the following example:

(3)

b: …{ = [< UM >] } { = [THANK] you < ***SO*** > much } { = for [THOSE] VEry < KIND > words }…

(HKCSE, P003, 11)

As a substitute form or pro-form, the word *so* can replace an adjective phrase or a noun phrase. At the same time, it can also substitute a *that*-clause which is often a complete sentence on its own (Quirk et al. 1985). This use of *so* is frequently preceded by verbs of belief or assumption, such as *hope*, *believe* and *think*, as in example 4:

(4)

A: { \ i [< THINK >] ***so*** } { \ [< GREAT >] } { \ [< ^ GOOD >] }

(HKCSE, C119, 4751)

The third type of propositional use of *so* concerns its function of expressing purpose as a conjunction. In these instances, *so* or *so that* links the two clauses together and presents the following component as the desired outcome arising from the preceding component. They are less formal alternatives for *in order that* (Leech and Svartvik 2002). In example 5, speaker a uses *so that* to introduce the purpose of giving the hearer the phone number of a third person, i.e. for them to contact each other:

(5)

a: { / [< WHAT >] } { = what [aBOUT] if < I > } { = [GAVE] < HER > } { \ er [GIVE] you her phone < NUMbers > } { ? ***so*** [THAT] < YOU > could } { = [< YOU >] could } { = [< ER >] } { \ [< CONtact >] her } { ? [OR] or < SHE > could } { \ [< RING >] you up }

(HKCSE, C005, 444)

In many fixed phrases, *so* contributes to the propositional meaning of the construction and thus is syntactically mandatory. Examples include *so called*, *so far*, *so to speak*, *and so on* (*and so forth*) and *or so*. Interestingly, *so called* is more common in the academic data than in other contexts. It is frequently used by teachers to indicate that the expression followed is the name generally given to the item being discussed, as in example 6:

(6)

b: …{ \ [THIS] is ***SO*** called the < HYSTEresis > } { = [< _ oKAY >] }…

(HKCSE, A009, 3307)

On the other hand, the discourse use of *so* is more loosely attached to the utterance, both syntactically and semantically. Unlike its propositional use, *so* as a discourse particle can be omitted without changing the grammaticality or intelligibility of the component it constitutes. Hence it is not unusual to find examples of *so* as a discourse particle after a pause, signalling a syntactic or semantic separation, as in this example:

(7)

b: …{ = [< WE >] } { \ we [MAY] have SOME intermediate < STATE > } { / [< oKAY >] } (.) { = [< ***SO*** >] } { = [^ HOW] to < eLIminate > } { = this [< KIND >] of } { = [< INT >] } { = [interMEdiate] < STATE > }…

(HKCSE, A010, 3875)

As discussed earlier in the review of previous studies of *so,* the resultative or consequential aspect of the word is often considered in its discussion as a discourse particle (see also Section 5.1). Following Müller (2005) and Schiffrin (1987), the present study classifies the use of *so* in marking result or consequence as a discourse function. This is because the use of *so* in such instances only guides or indexes the hearer to select the inferential relation from the range of meaning potential available through the content of talk in the interpretation but it does not create meaning on its own (Schiffrin 1987). In other words, *so* makes explicit a relation that is already given by its neighbouring context, thus it does not contribute to the propositional content of the utterance. This view is supported by the fact that the resultative or consequential reading of these instances can be arrived at without the presence of *so,* implying that *so* in this function is syntactically optional. In example 8 below, the consequence that the speaker never wants to see doctors again is based on the painful experience mentioned in the preceding context. It can be inferred from our perception of and reaction to pain, even if *so* is absent:

(8)

b: …{ \ it is [VEry] < PAINful > } { = [< ^ ***SO*** >] } { = [< I >] } { = [NEver] want to < SEE > } { \ [< DOCtors >] }…

(HKCSE, A009, 4571)

While the remaining part of the chapter will chiefly discuss the functions of *so* as a discourse particle, the propositional use of the word will also be studied quantitatively to compare with its discourse use. This serves to demonstrate that there are considerable differences between the two main uses of *so* in terms of various linguistic, sociolinguistic and contextual variables.

5.3 *So* in the HKCSE

This section gives a detailed account of the use of *so* in the HKCSE. It first provides the frequency statistics of *so* in the corpus for a general quantitative distribution of

the word as regards its use as a discourse particle across the four sub-corpora. This is followed by an examination of the positional distribution of *so* and its prosodic pattern in the data. A largely qualitative discussion of the pragmatic functions of *so* found in the Hong Kong corpus is then presented. Finally, this section investigates the relationship between the various linguistic, sociolinguistic and contextual factors and the discourse functions of *so*.

5.3.1 Frequency of occurrence

There are altogether 8,296 instances of *so* in the HKCSE which consists of approximately 950,000 words. With its large number of occurrences, it is hardly surprising that *so* ranks high in the wordlist by being the 17th most frequently occurring lexical item. These occurrences of *so* are found in 278 out of 311 texts in the corpus, showing the pervasiveness of the word in the spoken language. Of these instances of *so*, 6,721 are discourse uses (D-use) while 1,525 are propositional uses (P-use). Owing to insufficient contextual information, 50 examples of *so* are unclassified as regards whether they are used as a discourse particle. These however constitute a mere 0.6% of the total instances of the word. The distribution of the uses of *so* in the Hong Kong corpus is presented in the following table:

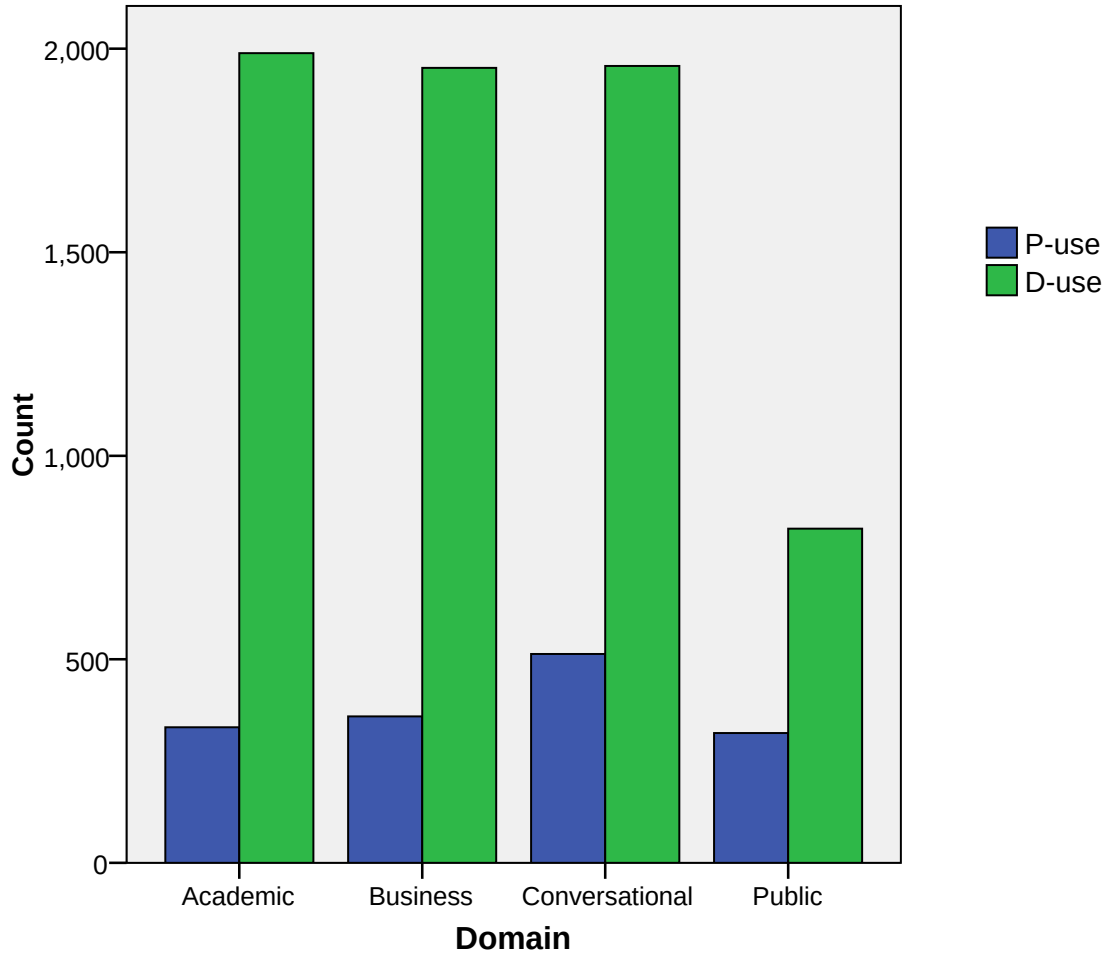
Table 5.1. The distribution of *so* in the HKCSE

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Total number of words in HKCSE	949,972 (100.0%)	213,204 (22.4%)	259,484 (27.3%)	258,882 (27.3%)	218,402 (23.0%)
Total number of *so*	8,296 (100.0%)	2,333 (28.1%)	2,329 (28.1%)	2,490 (30.0%)	1,144 (13.8%)
Number of D-use	6,721 (100.0%)	1,989 (29.6%)	1,953 (29.1%)	1,958 (29.1%)	821 (12.2%)
Number of P-use	1,525 (100.0%)	333 (21.8%)	360 (23.6%)	513 (33.6%)	319 (20.9%)
Unclassified use	50 (100.0%)	11 (22.0%)	16 (32.0%)	19 (38.0%)	4 (8.0%)
D-use / total use (%)	81.01	85.26	83.86	78.63	71.77
D-rate (per 10,000 words)	70.75	93.29	75.26	75.63	37.59

As indicated in Table 5.1, the occurrences of *so* are fairly evenly spread across the sub-corpora, except for the public domain. In the public data, only 1,144 instances of *so* are found. This is less than half of any of the other sub-corpora. Similarly, the

public sub-corpus houses the smallest number of discourse use of *so* (n=821) while no considerable difference is found for the other three sub-corpora. This might be ascribed to the fact that the public sub-corpus contains the highest number of prepared and scripted texts. It is interesting to note that conversations contain the highest number of P-use *so*, amounting to about one-third (33.6%) of the total number found in the corpus. On the whole, the overall discourse-function ratio of *so* in the HKCSE is 81.01%, indicating that roughly four out of five instances of *so* found in the corpus serve discourse functions. This shows that *so* is primarily a discourse particle, despite its marginal status in many particle studies. The variation in the discourse-function ratio of *so* across the four domains is represented graphically in Figure 5.1:

Figure 5.1. The distribution of *so* over the four domains in the HKCSE – D-use versus P-use



The above figure shows that despite the fact that the academic, business and conversational sub-corpora contain roughly the same number of D-use *so*, conversations have a considerably higher number of *so* with propositional meaning. Consequently, the highest number of P-use *so* found in conversations reduces its discourse-function ratio in this sub-corpus. Yet the lowest discourse-function ratio, at 71.77% as indicated in Table 5.1, again can be found in the public data. This can be observed by comparing the relative lengths of the bars representing D-use and P-use for the public domain in Figure 5.1. The differences in the discourse-function ratio between the four domains achieve very high statistical significance with a moderate association ($x^2 \geq 116.252$; d.f. = 3; $p < 0.0005$; $V = 0.119$). This indicates that the setting of the speech event could have an effect on how likely the word is going to be used as a discourse particle. From Table 5.1, it appears that *so* is most likely to achieve discourse functions if used in the academic domain (85.26%) and least likely in the public domain (71.77%).

As regards the discourse rate (D-rate) of *so*, Table 5.1 shows that the four sub-corpora vary quite substantially in this respect. While the highest discourse rate of *so* is observed in the academic sub-corpus with a rate of about 93 instances of D-use in a 10,000-word sample, the public sub-corpus only has about 38 instances of D-use in a sample text of the same size. In business texts and conversations, the D-rates are fairly similar. In every 10,000 words, there are roughly 75 instances of D-use *so*. The overall discourse rate in the HKCSE is slightly lower, at approximately 71 instances per 10,000 words. Whether the formality of public setting leads to a lower use of *so* as a discourse particle remains to be seen, yet the fact that conversations do not have particularly high discourse-function ratio or discourse rate when compared with other types of discourse seems to suggest that spontaneity may not be a crucial factor in determining whether *so* serves discourse or propositional function.

5.3.2 Position (utterance and tone unit)

5.3.2.1 Position in utterance

In the HKCSE, the word *so* displays a preference for certain positions in an utterance. The positional distribution of *so* in an utterance in the HKCSE is shown in Table 5.2:

Table 5.2. The positional distribution of *so* in an utterance in the HKCSE[43]

Position	**Total number of *so***	**D-use *so***	**P-use *so***
Final	166 (2.0%)	82 (1.2%)	84 (5.5%)
Initial	2,649 (32.1%)	2,560 (38.1%)	89 (5.8%)
Stand-alone	110 (1.3%)	110 (1.6%)	0 (0.0%)
Medial	5,321 (64.5%)	3,969 (59.1%)	1,352 (88.7%)
Total	8,246 (100%)	6,721 (100%)	1,525 (100%)

As seen from Table 5.2, most instances of *so* occur in utterance medial position (64.5%) followed by utterance initial position (32.1%). Altogether they constitute more than 96% of the total use. Utterance final (2.0%) and stand-alone (1.3%) instances of *so* are only in a small minority. However, there are positional variations regarding the discourse use and propositional use of the word. While the percentages of initial and medial D-use *so* remain similar to those for the overall use of the word, *so* occurs overwhelmingly (88.7%) in medial position when it conveys propositional meaning. As a discourse particle, *so* is most commonly found in utterance medial position (59.1%), as in the example below:

(9)

a: …{ = [< DIFferent >] } { = cultural [< GROUPS >] } { = [THINK] in different < WAYS > } { \/ and [< ***SO*** >] } { = [< UM >] } { = [< ER >] } { \ for [< ^ eXAMple >] } { = [< ER >] } { = [WEStern] < CULtures > } { \ versus [EAStern] < CULtures > }…

(HKCSE, A017, 177)

Discourse use of *so* also frequently occurs in utterance initial position. More than one-third (38.1%) of the instances of *so* are found at the onset of an utterance, as in example 10:

(10)

B: { \ ***so*** [WHAT] HAppens to THAT < reCORding > }

(HKCSE, B076, 1618)

[43] The 50 unclassified uses of *so* are not included in the quantitative analysis from this point onwards.

In addition, all free-standing instances of *so* in the corpus are discourse uses. None of them contributes to the propositional content of the one-word utterance containing them. In other words, these utterances only serve pragmatic functions. In the following example, speaker a, towards the end of the extract, produces a single *so* to prompt speaker B. This is echoed by speaker B's utterance initial use of *so*:

(11)

a: { = [< ER >] } { = a [FEW] days aGO < I > } { = [HEARD] < FROM > } { = [C##] < THAT > } { / you [< WANT >] } { = [< TO >] } { / [< BUY >] } { = the [FLAT] < IN > } { = [disCOVEry] < BAY > } { / [< IF >] }

B: { / [< IF >] } (.) { \ [WE] have the < MOney > }

a: { / [< ^ ***SO*** >] } ((laugh))

B: ((laugh)) { \/ [< *SO* >] } { \ it's i [THINK] it'll be < DIFficult > }

(HKCSE, C046, 403)

As for the P-use of *so*, the vast majority of instances of *so* containing propositional meaning are found in mid-position. They make up 88.7% of the total P-use. Example 12 illustrates a propositional use of *so* in utterance medial position:

(12)

B: …{ \ in [^ REAL] terms it WON'T be ***so*** over < VAlued > } { \ because [comPARED] to other < CURrencies > }…

(HKCSE, C090, 1665)

In contrast, only 5.8% of P-use tokens of *so* are found in initial position, compared with 38.1% of D-use. The following example shows the propositional use of *so* at the beginning of an utterance when the second speaker (speaker a) says *so far so good*:

(13)

b: { \ [< I >] see } { = but so [< FAR >] } { \ so [< _ GOOD >] }

a: { \ [***SO***] far so < GOOD > } { \ [< _ YES >] }

(HKCSE, B066, 1173)

There is also a higher proportion of P-use *so* in utterance final position (5.5%), compared with its D-use (1.2%). Utterance final instances of *so* conveying propositional meaning are mostly pro-form uses of the word. As discussed earlier,

they normally occur immediately after verbs of belief or assumption, as in the following example:

(14)

a: ((laugh)) { \ i [< HOPE >] ***so*** }

B: { = [< oKAY >] } { = [< BYE >] bye }

(HKCSE, B037, 438)

Not surprisingly, D-use *so* in end position does not share this lexical pattern. Example 15 shows a discourse use of *so* in utterance final position:

(15)

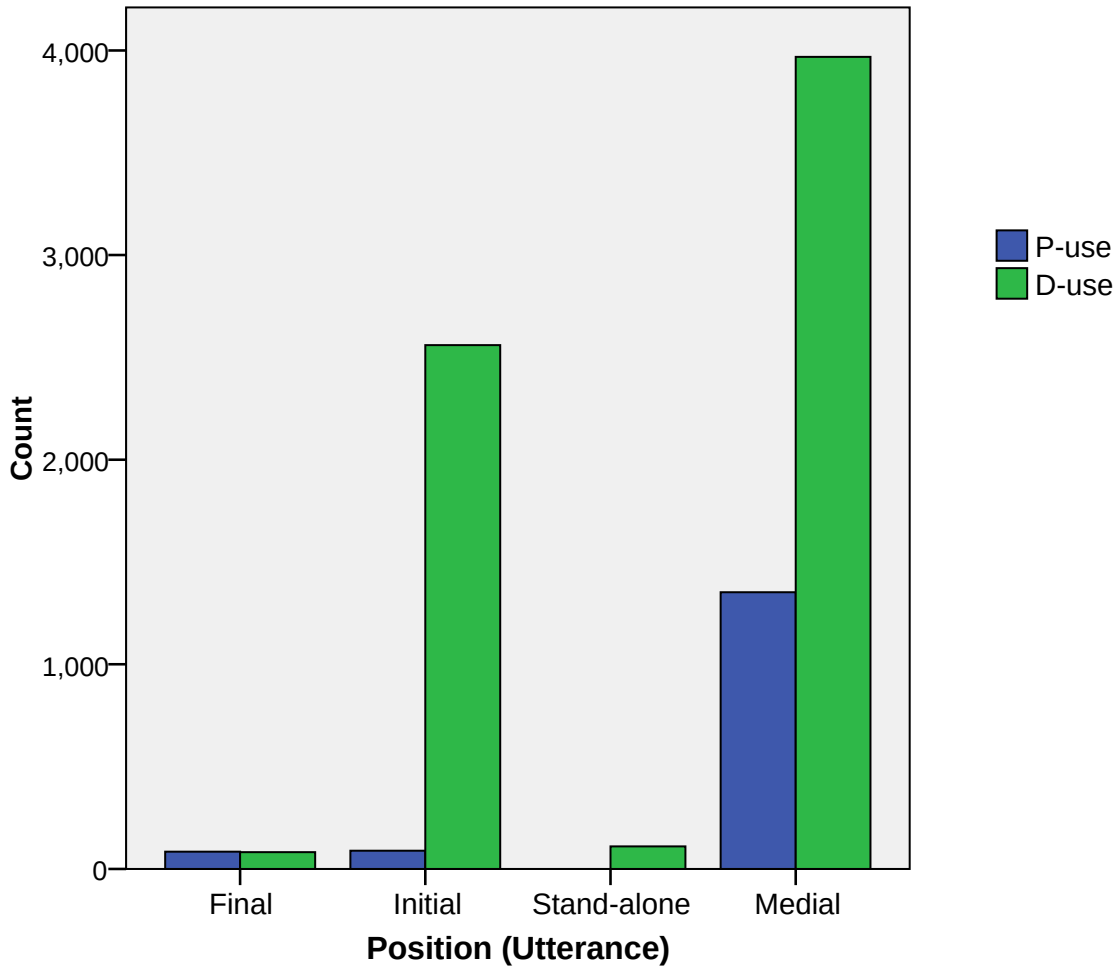
b2: { = and that [CALLS] for DIFferent < KIND > of er } { \ [< ^ Model >] } { \ business [< MOdel >] } { = [< ***SO*** >] }

b1: { \ [< _ YEA >] } { = [< NOW >] } { \ er [YI] deng the < ^ CHAMber's > } { \ been the [< FOREfront >] }…

(HKCSE, P127, 1508)

Table 5.2 and the examples above illustrate that D-use and P-use *so* differ in terms of their positional preference. Statistically, the difference in the positional distribution between D-use and P-use *so* is very highly significant with a moderate association ($x^2 \geq 709.797$; d.f. = 3; $p < 0.0005$; $V = 0.293$). This suggests that the position *so* occupies in an utterance can be considered a fairly strong indicator as regards whether it is used as a discourse particle. Figure 5.2 shows a graphical representation of the contrast:

Figure 5.2. The positional distribution of *so* in an utterance in the HKCSE – D-use versus P-use



5.3.2.2 Position in tone unit

Apart from its location in an utterance, the position of *so* in a tone unit is also analysed in the HKCSE. Table 5.3 shows the positional distribution of *so* in a tone unit in the corpus:

Table 5.3. The positional distribution of *so* in a tone unit in the HKCSE[44]

Position	**Total number of *so***	**D-use *so***	**P-use *so***
Head	1,319 (16.3%)	757 (11.5%)	562 (37.8%)
Pre-head	2,785 (34.5%)	2,481 (37.7%)	304 (20.4%)
Tail	345 (4.3%)	96 (1.5%)	249 (16.7%)
Tonic syllable (in a shared tone unit)	1,152 (14.3%)	826 (12.5%)	326 (21.9%)
Separate tone unit	2,469 (30.6%)	2,423 (36.8%)	46 (3.1%)
Total	8,070 (100%)	6,583 (100%)	1,487 (100%)

In total, 8,070 instances of *so* with prosodic information are studied. The largest number of *so* is found in the pre-head position, making up about one-third (34.5%) of all occurrences. This is followed by *so* forming a separate tone unit (30.6%). Altogether they constitute two-thirds of the total number of *so*. The large number of *so* occupying a tone unit of its own refutes Müller's (2005:61) view that *so* "does not form a separate tone group". For the remaining one-third of all instances, most of them are either the head (16.3%) or the tonic syllable in a shared tone unit (14.3%). The smallest number of *so* is found as the tail (4.3%).

When the discourse and propositional uses of *so* are compared, it is observed that they differ in terms of their positions in a tone unit. As a discourse particle, *so* is most commonly found as a pre-head (37.7%) or in a tone unit on its own (36.8%). In fact, most of the instances of *so* found in these two positions in the corpus serve discourse functions. Of the total 2,785 instances of pre-head *so*, there are 2,481 tokens used as a discourse particle. Example 16 illustrates a discourse use of *so* forming the pre-head with *if we* in front of the head *had the objective*:

(16)

a: { \ [< RIGHT >] } { \ ***so*** if we [HAD] the objective < THERE > } { \ to [< MATCH >] } { = [THEN] it < MIGHT > er } { = [< proJECT >] a } { = [< BETter >] } { \ [< PICture >] }

(HKCSE, A016, 762)

[44] 24 instances of *so* are excluded in the prosodic analysis altogether due to the lack of prosodic information. 152 instances of *so* are excluded in Table 5.3 as the prosodic patterns of the tone units containing them are indeterminable. Therefore, their positions in the tone unit are unclassifiable.

Similarly, an overwhelming number of *so* in a separate tone unit are found to be discourse use. Of the total 2,469 instances of *so* which occur on their own in a tone unit, there are 2,423 tokens serving discourse functions. The following example shows a D-use *so* as a separate tone unit:

(17)

a: …{ / i want to [DO] some < reSEARCH > } { \ [< ***SO*** >] } { \ i've [NOT] yet < deCIded > }…

(HKCSE, B071, 2089)

The fact that the majority of instances of D-use *so* either occupy the pre-head position or occur on their own in a tone unit provides further support to the syntactic detachedness of *so* as a discourse particle. As shown in Table 5.3, only a small number of *so* are found as the head (11.5%), tonic syllable in a shared tone unit (12.5%) or the tail (1.5%) when it is a discourse particle. They make up only about a quarter of the discourse use of *so*.

On the other hand, the propositional use of *so* displays a different pattern regarding its position in a tone unit. When *so* carries propositional content, it has a much higher tendency to be integrated in the tone unit with other linguistic elements. More than three quarters of the total number of propositional uses are prosodically integrated. In particular, *so* conveying propositional meaning is most likely to occupy the head position in a tone unit. More than one-third (37.8%) of all the P-use *so* are found in this position in the corpus. Example 18 shows an instance of *so* as part of the head in a tone unit:

(18)

A: …{ \ because [< NOW >] } { \ the [WORLD] is ***so*** < comPEtitive > } { \ for [YOUNG] < CHILdren > }…

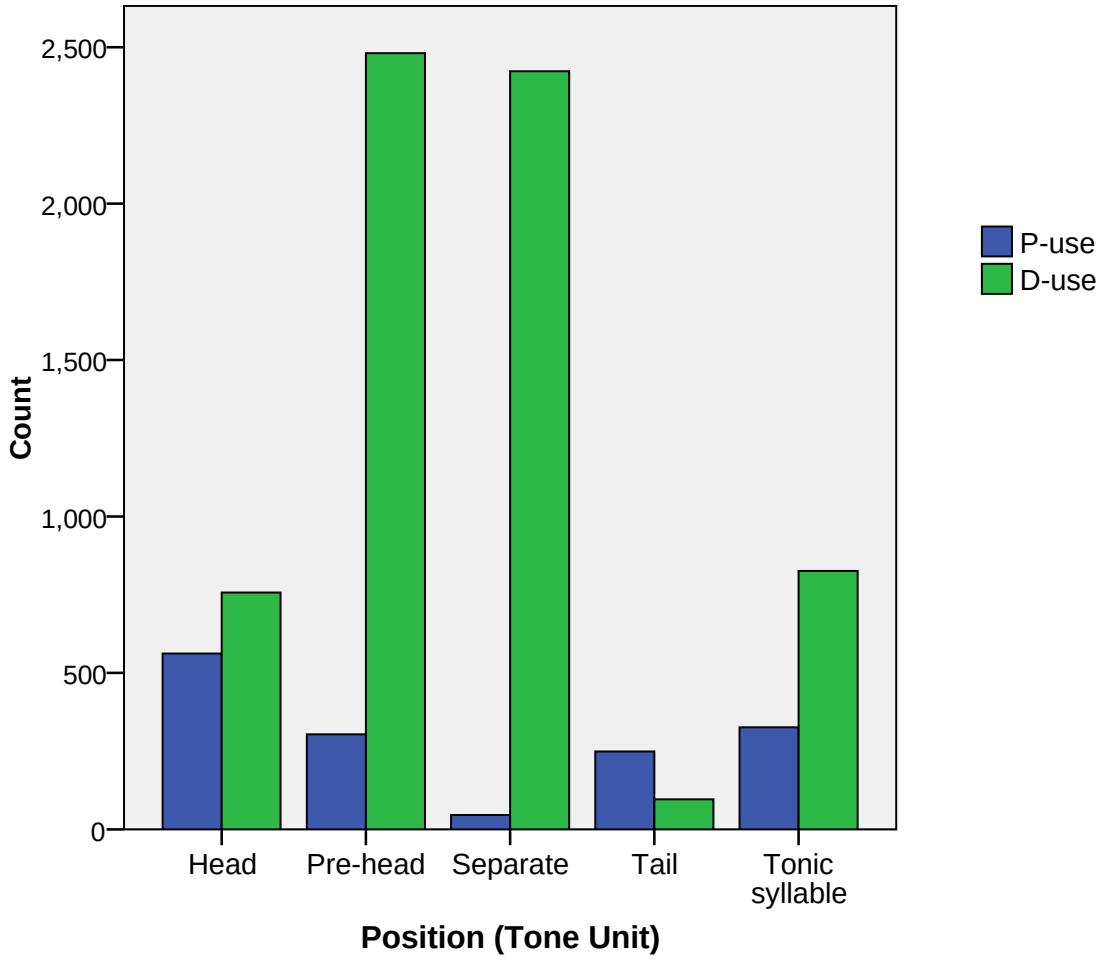
(HKCSE, C005, 1670)

Comparatively, the propositional use of *so* contains a higher proportion of instances as the tonic syllable in a shared tone unit than for its discourse use (21.9% vs. 12.5%), while the proportion of P-use *so* as a pre-head is lower than that of D-use (20.4% vs. 37.7%). The most remarkable difference, however, is found in the tail position and in a separate tone unit. While 16.7% of P-use *so* occur as the tail in a tone unit, the

percentage is only 1.5% for D-use *so*. Conversely, only 3.1% of *so* with propositional meaning constitute a separate tone unit, compared with 36.8% for discourse use.

Table 5.3 and the discussion above show that the D-use and P-use of *so* exhibit varied positional patterns in a tone unit. The difference again is statistically very highly significant, with a strong association ($x^2 \geq 1806.038$; d.f. = 4; $p < 0.0005$; V = 0.473). This provides statistical evidence that, similar to its position in an utterance, the placement of *so* in a tone unit is also highly indicative of whether it is used as a discourse particle. Although *so* is not frequently considered an archetypal discourse particle in previous studies, the findings above demonstrate that its discourse use is highly prosodically independent, thus fulfilling one of the typical properties of discourse particles to have high separability (Aijmer 2002), which is absent in its propositional use. Figure 5.3 represents the variations of D-use and P-use *so* in the distribution of positions in a tone unit graphically:

Figure 5.3. The positional distribution of *so* in a tone unit in the HKCSE – D-use versus P-use



5.3.3 The prosody of *so*

As regards the prosodic profile of *so*, the present study investigates three major aspects: its prominence pattern, its tonal distribution and the distribution of pauses in its neighbourhood. This section reports on findings in these areas in detail. Table 5.4 shows the prominence pattern of *so* in the Hong Kong corpus:

Table 5.4. The prominence pattern of *so* in the HKCSE

Prominence pattern	**Total number of *so***	**D-use *so***	**P-use *so***
Unstressed	3,398 (42.1%)	2,612 (39.7%)	786 (52.9%)
First	987 (12.2%)	710 (10.8%)	277 (18.6%)
Middle	64 (0.8%)	12 (0.2%)	52 (3.5%)
Last	154 (1.9%)	54 (0.8%)	100 (6.7%)
Sole prominent syllable (in a shared tone unit)	998 (12.4%)	772 (11.7%)	226 (15.2%)
Sole prominent syllable (in a separate tone unit)	2,469 (30.6%)	2,423 (36.8%)	46 (3.1%)
Total	8,070 (100%)	6,583 (100%)	1,487 (100%)

Of the 8,070 instances of *so* for which the prosodic pattern is studied, unstressed examples amount to 3,398 tokens. They constitute 42.1% of the total. Slightly less than one-third (30.6%) are in a separate tone unit. Hence they are the only prominent syllable in a monosyllabic tone unit. Collectively, the other four prominent patterns constitute about a quarter of the total number.

When *so* is used as a discourse particle, its stress pattern depends upon whether it is integrated in the tone unit. As a separate tone unit on its own, it obviously is the only prominent syllable in the group. More than one-third (36.8%) of D-use *so* are prosodically independent. However, when incorporated into the tone unit, it is most likely to be unstressed (39.7%), as in the following example:

(19)

b: …{ = ***so*** i'm going to [< TALK >] to the media }…

(HKCSE, P101, 797)

While more than one-third of instances of D-use *so* have their own tone unit, more than 95% of P-use instances share a tone unit with other elements. Although again

being unstressed is the most common prominence pattern for P-use (52.9%), the proportion of P-use *so* being stressed in a shared tone unit is considerably higher than for D-use. 44% of P-use *so* are prominent when it is in a tone unit with other elements, compared with 23.5% for D-use. In particular, more instances of *so* are found to be prominent in the middle of the tone unit or as the final prominent syllable (i.e. the tonic syllable) for P-use than for D-use. Example 20 shows an instance of P-use *so* being a prominent syllable in a mid-position in the tone unit, with *that* and *gauge* as the first and last prominent syllables respectively:

(20)

b: …{ = [THAT] is ***SO*** called the < GAUGE > } { \ [< RANGE >] }…

(HKCSE, A009, 3052)

Example 21 shows the propositional use of *so* as the last prominent syllable in the tone unit. In this example, *so* is a modifying adverb which gives an extra emotive emphasis to the feeling expressed. This is consistent with Leech and Svartvik's (2002) remark that *so* serving this function is stressed and may receive nuclear stress, which corresponds to *so* being the tonic syllable in Brazil's (1997) discourse intonation framework:

(21)

a: … { \ [< ^ OH >] } { \ the [Other] day is < ***SO*** > funny } { \ i [HAVE] to TELL you < _ SOMEthing > }…

(HKCSE, C079, 1005)

In sum, the discourse and propositional uses of *so* vary in terms of their prominence pattern. In particular, the findings above seem to suggest that the prominence pattern of D-use *so* depends upon whether it is prosodically integrated or independent. Hence drawing the conclusion that *so* as a discourse particle is more likely to be stressed than unstressed, without taking into account of its prosodic phrasing, is too simplistic.

For the study of nuclear tones, 3,621 instances of *so* bearing the nucleus in the HKCSE are examined. Table 5.5 details the tonal distribution of *so* in the corpus:

Table 5.5. The distribution of nuclear tones on *so* in the HKCSE

Type of tone	**Total number of *so***	**D-use *so***	**P-use *so***
Fall	760 (21.0%)	636 (19.6%)	124 (33.3%)
Fall-rise	27 (0.7%)	16 (0.5%)	11 (3.0%)
Level	2,520 (69.6%)	2,321 (71.4%)	199 (53.5%)
Rise	132 (3.6%)	109 (3.4%)	23 (6.2%)
Rise-fall	0 (0.0%)	0 (0.0%)	0 (0.0%)
Unclassified	182 (5.0%)	167 (5.1%)	15 (4.0%)
Total	3,621 (100%)	3,249 (100%)	372 (100%)

Findings from Table 5.5 show that when *so* is the tonic syllable it is most frequently associated with level tone. Of the 3,621 instances of *so* being the nucleus, there are 2,520 tokens of level *so*, making up 69.6% of the total. Falling tone is the second most frequently occurring, which lags far behind level tone and only constitutes 21.0% of the whole. Rise and fall-rise tones with *so* are even less common, each amounts to only 3.6% and 0.7% respectively.

Unlike the positions of *so* and its prominence pattern, the tone choice of *so* does not seem to give strong clues as regards whether *so* is used as a discourse particle. From Table 5.5, it is observed that D-use and P-use *so* share a similar tonal distribution, with level tone being the most frequent followed by falling tone. However, there are slight differences in tone choice between the discourse and propositional use of *so*. Although level tone is predominantly used to serve both discourse and propositional functions, its proportion for discourse use is higher. As a discourse particle, 71.4% of *so* carries level tone. The percentage drops to 53.5% when it conveys propositional meaning. Example 22 illustrates a discourse use of *so* in a tone unit on its own with level tone:

(22)

b4: { = [< ***SO*** >] } { = there's a [< CHANGE >] of } { \ from the [L] to R so it becomes < FLIGHT > }

(HKCSE, A029, 5044)

On the other hand, there is a higher proportion of P-use *so* carrying falling tone, compared with D-use. One-third (33.3%) of the instances of *so* with propositional content carry falling tone, while there are only less than one-fifth (19.6%) falling *so* for discourse use. This observation that fall tone with *so* occurs less frequently with

D-use when compared with P-use can be an outcome of the higher frequency of level tone with D-use than with P-use. The following example shows a propositional use of *so* with fall tone:

(23)

b: …{ = er [ONE] three < A > } { = [ONE] three < B > } { = [ONE] three < C > } { \ and [SO] on and < ***SO*** > forth }…

(HKCSE, A022, 5077)

Apart from the increase in the ratio of fall tone, the propositional use of *so* also has a higher proportion of rise and fall-rise tones than the discourse use. For discourse use, the percentages of rise and fall-rise tones with *so* are only 3.4% and 0.5% respectively. When *so* contains propositional value, however, the percentages go up to 6.2% and 3.0% correspondingly.

Although no relevant results on the tonal distribution of *so* are available from other studies, the comparison of the tonal pattern of *so* with other discourse particles such as *oh* (see Aijmer 2002) indicates that the dominance of level tone is only a characteristic of *so* but not of discourse particles in general. In fact, it is common for discourse particles to have different tonal preferences. Table 5.6 compares how the nuclear tones on *so* and *oh* are distributed:

Table 5.6. The comparison of tonal distribution on D-use *so* in the HKCSE and D-use *oh* in the London-Lund Corpus (Aijmer 2002:109)

Type of tone	**D-use *so* in HKCSE**	**D-use *oh* (Aijmer 2002)**
Fall	636 (19.6%)	437 (78.0%)
Fall-rise	16 (0.5%)	8 (1.4%)
Level	2,321 (71.4%)	12 (2.1%)
Rise	109 (3.4%)	25 (4.5%)
Rise-fall	0 (0.0%)	77 (13.8%)
Unclassified	167 (5.1%)	0 (0.0%)
Total	3,249 (100%)	559 (99.8%)

While *so* is typical with level tone, *oh* is most commonly associated with falling tone. Level tone *oh*, in comparison, is much less frequent (2.1%). It is also noteworthy that 13.8% of D-use *oh* is pronounced with rise-fall tone whereas this choice of tone is rather uncommon not only with *so* but in the whole HKCSE (Cheng et al.

forthcoming). More comparisons need to be made with other prosodic studies in order to examine whether this is a unique feature of Hong Kong English or a characteristic of the discourse intonation framework applied to the corpus, as the data from which Brazil's (1997) model is developed also contain very few examples of rise-fall tone.

In terms of the study of pauses in the neighbourhood of *so*, findings from the HKCSE again show some contrastive patterns between the discourse and propositional use of the word. Table 5.7 presents the distribution of pauses in the close proximity of *so*:

Table 5.7. The distribution of pauses before and after *so* in the HKCSE

Type of tone	**D-use *so***		**P-use *so***	
	Before *so*	After *so*	Before *so*	After *so*
Brief pause	765	205	40	24
Unit pause	160	16	2	7
Total	925	221	42	31

The most striking difference between D-use and P-use *so* in terms of pauses is the fact that a much higher number of pauses are associated with the discourse use of *so*. Of the 6,721 instances of D-use *so*, there are 1,146 tokens of pauses in their immediate surroundings. In other words, on average in every six instances of D-use *so* there is one token either preceded or followed by a pause (17.1%). On the contrary, of the 1,525 propositional tokens of *so*, only 73 pauses are found. This means that the occurrence of pauses before and after P-use *so* is less than one in 20 (4.7%). In addition, a considerably higher number of pauses are found before *so* than after it when it is used as a discourse particle. The number of pauses preceding *so* is more than four times that of pauses following it (n=925 vs. n=221). Example 24 shows an instance of D-use *so* preceded by a brief pause:

(24)

a: …{ / you [< KNOW >] } { \ what i [< MEAN >] } (.) { = [***SO***] you can < SAY > } { \ mainly [< SHEEP >] }…

(HKCSE, A001, 2332)

This pattern, however, is not observed in P-use *so*. The number of pauses before and after the propositional use of *so* appears to show minimal difference (n=42 vs. n=31). The following example illustrates a P-use *so* immediately followed by a brief pause:

(25)

b: { = i don't [< THINK >] ***so*** } (.) { = we [< JUST >] } { = [< ^ MENtioned >] it } { = [< Over >] } { = [< COFfee >] time }…

(HKCSE, C024, 2289)

The contrast in the association with pauses between discourse and propositional use of *so* suggests that pauses offer important indications concerning whether *so* is used as a discourse particle. The fact that the discourse use of *so* is frequently found to co-occur with pauses also provides further evidence for the structural detachedness of *so* serving discourse functions.

By examining the prominence pattern of *so*, its tonal distribution and the occurrences of pauses in its surroundings, the present study illustrates that prosody not only provides an additional dimension for the study of *so*, but it also gives important clues as regards whether the word functions discoursally. As shown by the prosodic analysis above, the discourse and propositional use of *so* display different patterns in terms of their prosodic profiles. The prosodic features of *so* will be dealt with again in Section 5.3.4 when the various discourse functions of *so* in the HKCSE are discussed.

5.3.4 Discourse functions of so *identified in the HKCSE*

The analysis of *so* in the Hong Kong corpus classifies the 6,721 tokens of discourse use into six major functional categories spanning the textual, interpersonal and interactional functional domains. Within the framework of Linear Unit Grammar (Sinclair and Mauranen 2006), *so* achieving textual functions in the present study largely corresponds to the text-oriented organisational unit (OT) whereas *so* serving interpersonal or interactional functions largely corresponds to the interactive-oriented organisational unit (OI). Because of the inherent dynamics of naturally-occurring speech, in some occasions *so* may work on more than one functional domain but a principal function is often identifiable based on the linguistic and contextual information at hand. The following discussion describes in detail the six major discourse functions of *so* identified in the HKCSE with illustrating examples.

Textual functions

5.3.4.1 Frame (framing)

Of the 6,721 tokens of discourse use of *so* found in the HKCSE, there are 2,451 instances serving the frame function. Proportionally, more than one-third (36.5%) of D-use *so* act as a frame, making this use of *so* the most frequently occurring of all the discourse functions of the word. As a frame, *so* signals textual transitions of various types. On a global scale in the speech event, it marks easily identifiable macro moves between disparate topics and discourse stages. One of the most commonly found transitions marked by *so* is the change of topics. In example 26, two friends are first having a conversation on learning languages. At the beginning of the excerpt, the native speaker of English (speaker B) is asking the Hong Kong Chinese (speaker b) whether he is learning Mandarin. As the talk evolves, the Hong Kong Chinese appears to be reluctant to continue on the topic, which is evidenced by his repeatedly short and negative responses. Notice how speaker b changes the topic from learning a language to the job of speaker B's wife by using *so* after a short pause in the conversation:

(26)

B: { ? [< _ STUpid >] } (.) { = [< WHAT >] about } { = [< YOU >] } { / are you [LEARning] < MANdarin > } { = [< OR >] }

b: { \ [< NO >] }

B: { / not [< BOthered >] }

b: { \ [NOT] < _ BOthered > }

B: { \ [< _ YEAH >] } { \/ [PRACtise] your english < FIRST > right }

b: { = [< _ NO >] } ((laugh))

((pause))

b: { \ ***so*** your [< WIFE >] } { = got a [< JOB >] }

B: { \ [< ^ YEAH >] }

(HKCSE, C109, 1401)

So is also used when the speaker wants to go back to a previously discussed topic. The digression can be initiated by a sudden change of thought of the same speaker and is only loosely related to the prior talk, or it can be triggered by an external stimulus and is irrelevant to the topic just being discussed. In example 27, two

colleagues are in a university staff canteen talking about the revision of a paper that they write together. While the Hong Kong Chinese (speaker b) is responding to the native speaker's (speaker B) comment about the hermite functions they describe in the paper, he suddenly changes the topic of the talk after a brief pause by asking speaker B whether he would like a coffee. This rather abrupt switch in topic might arise from some external sources such as the sight of drinks nearby, which prompt the ordering of coffee. After this brief digression of speech-in-action, the native speaker of English initiates his talk with *so* to go back to the original topic:

(27)

B: { = [< WELL >] } { = where the [< HERM >] } { \ [< _ YEAH >] } { \ [MAYbe] we should say < THAT > } { = where the [HERmite] < FUNCtions > are }

b: { ? the [HERmite] FUNCtions < ARE > d } (.) { / [WONDER] whether you want a k < OFfee > }

B: { \ [< SURE >] } * { = are you [GOing] to < GET > some }

b: ** { = [< YEAH >] yeah } { ? [< YEAH >] } { = [< YEAH >] yeah } { \ [< I'M >] } { = [< ALright >] }

((pause))

B: { \ [< OH >] } { = [THANK] you < ^ SIR > }

b: { \ [YOU'RE] < _ WELcome > }

B: { \ ***so*** i [^ MADE] these < CHANges > } { \ [< TOO >] } { = and i think [THEY'RE] < PROBE > er } { = ((inaudible)) cell } { = [L] s < L > }

(HKCSE, C033, 164)

Apart from the management of discrete discourse topics and stages, *so* is also used to signal moves of a more abstract type, namely the micro transitions between different levels of textual structure in talk. Compared with changes in topic and discourse stage, these transitions often operate more locally in discourse and involve a more subtle shift in focus which is not as easy to identify. Example types of transitions are inclusive of but not restricted to the following: different aspects of the same topic, change of perspective, varying degrees of generality and specificity, and meta-discourse and the actual subject matter. In example 28, a teacher (speaker a) is explaining the difference between verbs and adjectives. Notice her use of *so* in the

middle of the talk to change the focus from her explanatory remarks on adjectives in general to a specific example of the adjective *beautiful*, in order to illustrate her point:

(28)

a: …{ = you [< CAN >] } { = [< NOT >] } { = [PUT] < AN > } { \ [< ADjective >] } { / [< HERE >] } { \ and plus [I] n g AND make it INto a < NOUN > } { \ it's [NOT] < _ POSsible > } { / [< RIGHT >] } { = [***SO***] for example if you < HAVE > the } { = [< WORD >] } { \ [< ^ BEAUtiful >] } (.) { = [< BAsically >] } { = [< AND >] } { \ i mean an [< ADjective >] } { = you [< canNOT >] } { \ [< ADD >] } { = [I] n G < TO > it }…

(HKCSE, A013, 739)

The example above shows the shift between different aspects of the same topic as well as the transition between generality and specificity. While the property of adjectives remains the main discourse topic of the talk, there is a varying degree of specificity: from a general account which applies to all adjectives to a specific instance which serves to support her assertion. This use of *so* to introduce examples for a previously discussed point is consistent with Schiffrin's (1987) view that *so* introduces grounds to claims made earlier by instantiation.

On a micro level, *so* is also typically used as a summarising or concluding device. In many cases, the conclusion is the final phase of a three-stage pattern which involves the stating of a proposition, the provision of reasons or examples to support the proposition, and the reiteration of the proposition. Linguistically, this pattern often involves a *because*-clause in the second-stage and a *so*-clause in the third-stage and could be schematized as *A because B so A'* (Passot 2007). At the beginning of example 29, a teacher (speaker b1) is making a statement about the importance of being accurate in the communication with suppliers. He then moves on to give reasons why accuracy is necessary with the *because*-clause. Finally, he uses *so* to reinstate the significance of accuracy and precision again as a conclusion to his statement:

(29)

b1: { = [ONE] < THING > is } { \ you [^ HAVE] to be < ACcurate > } { \ [< _ aBOUT >] it } { = [beCAUSE] if you < DON'T > } { \ if [THERE] is the < misunderSTANding > } (.) { \ you [< ^ WILL >] } { = [reCEIVE]

the < WRONG > } { \ [< PROduct >] } (.) { \ ***so*** you [HAVE] to be < ACcurate > } { \ and [< preCISE >] }

(HKCSE, A004, 4676)

In her chapter of *because* and *so* as markers of cause and result, Schiffrin (1987:191) also describes this conclusive use of *so*, calling it the "marker of main idea units". Her examples seem to suggest that as a frame *so* is only used to mark the return to a main point previously discussed. However, evidence from the present data shows that as a boundary marker *so* is also employed when a new topic is introduced, as example 26 demonstrates. In addition, example 28 illustrates that *so* can also mark a transition from a generic statement to a specific example, which is difficult to fit within "the main idea unit" concept. Although the category of boundary marker is found in Müller (2005), her examples are confined to the transition between instructions and narratives owing to the nature of her data, which are retrieved from a movie-retelling experiment. Instances of *so* marking boundaries between other descriptive levels such as stating a scenario, explaining, instantiating, which are abundant in the Hong Kong corpus, are not reported in her study. The absence of these examples in earlier studies seems to imply that *so* has a wider scope of application as a frame than previously assumed.

In terms of positional preference, the use of *so* as a frame tends to occur in utterance medial position, except for *so* in the macro management of topics. While introducing a new topic or returning from a digression to a previously discussed topic, *so* is more likely to occur at the beginning of the utterance. For other sub-types of frame which are concerned with transitions of a micro level, more instances of *so* are found in the middle of the utterance.

As a frame, *so* is frequently associated with pauses. In addition, it is more common for *so* to occur immediately after a pause than before it, as in example 29 above. The occurrences of pauses before the use of *so* as a frame further cement the boundary marking role *so* plays in the division of talk. While the presence of a pause signals a break in the current conversational flow, *so* reorients participants to a direction which is not presently in focus in discourse, be it an entirely different topic or discourse stage, the resumption from a digression, or a micro level of structure in talk such as the conclusion of a claim after instantiation.

5.3.4.2 Link (linking)

Apart from dividing chunks of talk in discourse, *so* is also used in the textual domain to connect parts of text together for the purpose of coherence. In the HKCSE, there are 830 tokens of *so* used as a linking device, constituting 12.3% of the total discourse use. There are two ways in which *so* serves as a link in talk, depending on the kind of relationship between the adjoining segments. Firstly, *so* marks a sequential relationship between discourse units, introducing the following segment of talk as the next phase which occurs after the preceding segment. There is usually a temporal or step-wise element in these instances and the use of *so* can be paraphrased as *then* (cf. Müller 2005; Redeker 1990). In example 30, the Hong Kong Chinese (speaker a) is describing her son's reaction to school. At the end of the excerpt, she uses *so* to signal the transition between two events in sequence, i.e. her son cried for a few months and then he stopped crying:

(30)

a: …{ = [< ER >] } { \ [LAST] < YEAR > } { \ he [WENT] to < SCHOOL > er } { = [< DUring >] } { = [< ER >] } { = [< aROUND >] mid } { \ [MID] < AUgust > } { \ he [THEN] he CRY for few < MONTHS > } { \ [***SO***] he stop < CRYing > }…

(HKCSE, C024, 4554)

More often though, there is not a clear sequential relationship between discourse units when *so* is used as a connector. In such cases, *so* loosely introduces additional information to the preceding segment, which is roughly equivalent to the appending function of *and*. In example 31, the segment following *so* simply provides further details of chapter three, i.e. there are twenty three pages in the chapter:

(31)

b: …{ / and [< ^ THEN >] } { = chapter [< ^ THREE >] } { = [< TELLS >] you } { = the [< TYPE >] of } { = [TRANSFUsions] and THEIR < appliCAtions > } { / [< oKAY >] } { = [< ***SO*** >] } { = these [< ARE >] } { = [TWENty] < THREE > } { \ [< PAges >] } { / [< oKAY >] } { = [CHAPter] < ^ THREE > }…

(HKCSE, A008, 110)

When *so* introduces additional information to the preceding discourse, it is not uncommon to find ideas which appear at the end of the prior segment being recycled at the beginning of the following segment. In the language of functional grammar, the rheme of the prior segment becomes the theme of the following segment, as in example 32, where *the entry permit* is part of the rheme of the unit *but we will er ask the company to come and collect the er entry permit* and it turns into part of the theme of the unit *so the entry permit will be in the form of a label*:

(32)

a: …{ = [BUT] < WE > will } { = er [ASK] the < COMpany > } { = to [COME] and COLlect < THE > er } { \ entry [< PERmit >] } { \ ***so*** the entry [PERmit] will be in the FORM of a < LAbel > }…

(HKCSE, P089, 3401)

From the literature, it seems that the use of *so* as a linking device, especially its appending function, is not thoroughly discussed. Examples of a similar nature are not found in most of the previous studies reviewed (see, for example, Blakemore 1988; Fraser 1999; Schiffrin 1987). While Müller (2005) provides examples showing the sequential aspect of *so*, she admits that her data may not cover the range of possible sequential relations. This could be the reason why *so* as a marker adding extra information to the preceding discourse is not found in her study. Nevertheless, the sequential and appending function of *so* is discussed in the *Collins Cobuild English Language Dictionary* (2002). According to the dictionary, *so* is used "in stories and accounts to introduce the next event in a series of events or to suggest a connection between two events". This description matches the examples serving linking functions found in the present study.

5.3.4.3 Marker of result or consequence (consequential)

As mentioned earlier in the review of previous studies, the use of *so* marking resultative or consequential relationship has been discussed extensively and is often considered its quintessential function. In Biber et al. (1999:877), for example, *so* is referred to as a "typical resultive linking adverbial" in conversation. Schiffrin (1987) also calls *so* a marker of result and argues that other functions are an extension of this core meaning. In the present study, however, only instances of *so* which clearly mark a resultative relation between adjoining units are included in this category.

Occurrences of *so* which join together two discourse units in a relationship of sequence or addition without an explicit consequential association are treated as a linking device. This classification criterion may account for the rather small number of instances in this category in the data. In the HKCSE, there are 653 tokens of *so* marking a resultative or consequential relationship, making up 9.7% of discourse use. In the academic domain, for instance, *so* is frequently used to introduce the result of logical inference, as in example 33:

(33)

b: …{ = [< THIS >] is } { \/ [< FIVE >] } { = [< AND >] } { \/ we [KNOW] that THIS angle is < GAMma > } { \ ***so*** [THAT] angle would be GAMma minus < FIVE > }…

(HKCSE, A002, 3293)

This inferential relation can be replaced with a causal relation by reversing the preceding and following discourse segments with *because* instead of *so*, as in the constructed example below:

(34)

b: that angle would be gamma minus five because this is five and we know that this angle is gamma

(constructed example from example 33)

In some cases, the inferential relation is strengthened by the use of other linguistic items right after *so*. Example 35 shows an instance where *so* and *that's why* are used together to emphasize that younger drivers in the bus company are a result of the higher pay:

(35)

B: { ? [< DRI >] } { \ [BETter] < DRIvers > } (.) { = [BUses] are < CLEAN > }

b: { \ [< _ WELL >] } { \ i [supPOSE] they PAY < ^ MORE > } { = ***so*** that's [WHY] they can GET < THE > er } { \ [YOUNger] < ^ DRIvers > }…

(HKCSE, C013, 10467)

The reversibility of segments with *because* appears to be a unique property of *so* marking result or consequence. Only instances of *so* which clearly mark a resultative

or consequential relationship can be replaced by *because* with a reverse of segments while maintaining the meaning of the utterance. Tokens of *so* signalling the textual relations of framing and linking, on the other hand, do not permit this kind of transformation.

Interpersonal function

5.3.4.4 Responsive signal (responsive)

As a responsive signal, *so* marks the forth-coming response as initiated by the prior discourse. Similar to *so* as a marker of result and consequence, the use of *so* in response also contains an element of inference. However, the consequential relationship signalled by *so* in the interpersonal domain is between the discourse and the speaker, but not between two adjacent segments in the discourse. In this function, *so* introduces the speaker's reaction towards some preceding information, which could be the speaker's own prior talk, another speaker's contribution or even the extra-linguistic context. In the HKCSE, 1,871 tokens of *so* are found to signal interpersonal responses. This function accounts for 27.8% of total discourse use, making it the second most frequently occurring discourse function of *so* in the corpus.

The most common interpersonal response prefaced by *so* is in the form of questions. About half of the instances of *so* (n=913) in this category are followed by a question, which are mostly found in utterance initial position. In example 36, a teacher (speaker x) is having a supervision session with a student (speaker a) on writing skills. Notice the use of *so* by the teacher to introduce a follow-up question for clarification purposes as motivated by the student's talk:

(36)

a: { = [< ER >] } (.) { = i want [< TO >] } (.) { = [< ER >] } { = [< I >] want } { \ [I] want for your < _ HELP > } (.) { = on [CHECking] < MY > } { \ [APplication] < LETter > }

x: { \ ***so*** [^ WHAT] kind of < CHECking > do you mean } { \ [WHAT] do you mean by < CHECking > }

(HKCSE, A036c, 28)

Apart from genuine questions which seek information from the hearer, a small number of occurrences of *so* are also used to preface rhetorical questions which serve to raising attention. In example 37, the teacher uses *so* to introduce a rhetorical

question as motivated by his prior talk. In other words, the speaker is reacting to his own talk and the question is presented as his personal response to the preceding explanation:

(37)

b: { = so [^ ALL] the < soLUtions > of } { = [< deSIGning >] } { = [< ER >] } { = [CONtact] of a < CORrect > } { = [< CIRcuit >] } { = is [< ACtually >] } { \ [TRY] to < ^ Imitate > } { \ the [LInear] reSIStor < _ LIKE > this } { \ [< Okay >] } (.) { \ [< ***SO*** >] } { \/ what is the [< PROblem >] then } (.) { = [< ^ unFORtunately >] } { = [< unFORtunately >] } { \ [CONverter] is ^ NOT < reSIStant > } { \ it's [NOT] < reSIStant > }…

(HKCSE, A022, 414)

In the example above, the use of *so* seems to release the speaker from the 'monologic descriptive state' to interact with other participants in the discourse event. Not surprisingly, the use of *so* in introducing rhetorical questions as responses is mostly found in largely monologic talk. In particular, it is especially prevalent in lectures. This shows that teachers frequently use questions not only to elicit response but also to get attention from students.

Another type of response frequently found with *so* is in the form of comments. In the data, it is found that *so* often precedes an upcoming personal evaluation based on the information provided by prior talk. In example 38, speaker a1 and speaker a3 are in a job interview. Speaker a1 is the interviewer while speaker a3 is a candidate applying for the job. At the beginning of the excerpt, the candidate is asking the interviewer how many applications are received for the post. After hearing the interviewer's answer, speaker a3 uses *so* to introduce her evaluation of the situation, suggesting that her opinion is motivated by the information provided by the interviewer:

(38)

a3: { = so [< ^ HOW >] many } { \ [CANdidates] so < _ FAR > } (.) { \ [< I >] just want to know }

a1: { = we [HAVE] < GOT > erm } { = [< MM >] } { \ [< MAny >] } { / [< RIGHT >] } * { \ [MAny] < appliCAtions > } { = [< THAT >] }

a3: ** { ? [< AH >] } (.) { / [MORE] than < TEN > } ((laugh))

a1: { \ [< ^ YEAH >] yeah yeah } { \ [< DEfinitely >] } { = [< _ YEAH >] }
a3: { \ [Uhuh] < Uhuh > } { \ [< ***SO*** >] } { \ [VEry] < comPEtitive > }
(HKCSE, B082, 3738)

Since the responsive use of *so* still retains the resultative meaning of the word and implies that the following response is motivated by some preceding information, it is a useful device for achieving cooperativeness. This is because the use of *so* in such responses suggests that the current speaker is actively taking into consideration what the previous speaker said. Not only does it show the current speaker's attentiveness to the other interlocutor in the speech event but it also shows the current speaker's willingness to draw inference based on the preceding discourse of the last speaker. This helps to create a supportive atmosphere and establish solidarity among participants.

Alternatively, speakers can also provide personal evaluation of their own prior talk. This is usually found in cases when speakers have presented something as objective facts or have narrated a story and then move on to give their opinions about those facts and stories. In these examples, *so* is used to introduce speakers' attitudes as being triggered by the information just given. In example 39, speaker B is expressing his view on teachers going to horse racing events in Hong Kong. Notice his reference to the popularity of gambling in Hong Kong as a stated fact with the referring fall-rise tone in the tone unit *everybody gambles*, which directly leads to his opinion that a teacher should not be embarrassed when gambling:

(39)
B: …{ = because [GAMbling] is a < VEry > } { = [POpular] PAStime in hong < _ KONG > } (.) { \ [aCROSS] < SO > many } { \ [DIFferent] social < GROUPS > } { / you [< KNOW >] } { \/ [EVErybody] < GAMbles > } { \ ***so*** i don't know [< WHY >] } { \/ a [< ^ TEAcher >] } { \/ would be [< emBARrassed >] } { ? [< TO >] } { \ at the [< RACE >] track }
(HKCSE, C115, 6290)

The above example shows that using *so* to introduce speakers' personal viewpoints enables speakers to present their opinions as initiated by some previous information which provides supporting grounds for their own evaluation of the situation. This

provision of the link between reasons or justifications which are implied to be given in some preceding context and the comments made legitimises the speakers' point of view and minimizes the imposition (Luke 1990). For this reason, it is not uncommon to find the use of *so* prefacing face-threatening acts, such as criticisms, suggestions and requests. The presence of *so* in such cases allows speakers to present the face damage as motivated by some external agent instead of the speakers themselves, thus deflecting the responsibility of imposing on others and making the action less face-threatening. Therefore, *so* is a useful resource for the realisation of negative politeness strategies. In example 40, speaker a1, who is a staff member of a hotel, is asking her colleagues to submit a nomination form. By using *so* to create a link between the deadline of the nomination and her request for the nomination form, speaker a1 provides justification for her request. The potential damage of the hearer's negative face by the request is further reduced by her use of the politeness marker *please* immediately following *so*:

(40)

a1: …{ ? [toDAY] is the < DEADline > for the } { = [< TRAIning >] } { \ [< nomiNAtion >] } { = ***so*** please [FORward] your nomination < FORM > to } { \ our [OFfice] by the end of < toDAY > }

(HKCSE, B016, 513)

In the HKCSE, the use of *so* to preface requests is especially prevalent in service encounters. Example 41 shows the interaction between an airline officer (speaker b) and a passenger (speaker B). The officer's use of *so* before his request for the passenger's passport presents his request as motivated but not self-initiated, hence giving justification to his face-threatening act:

(41)

b: { = [< THE >] } { \ ((inaudible)) } { \ [toDAY] it's go to < toRONto > } { = and [ALL] the way < GO > to } (.) { \ [< MONtreal >] }

B: { \ [< MONtreal >] }

b: { \ [< YES >] } { = ***so*** [MAY] i have your passport < ^ PLEASE > } (.) { = [ONE] hundred dollar for the < AIRport > tax please }

(HKCSE, B051, 21)

In business transactions, it is important for service providers to offer prompt yet well-mannered service in order to create and maintain good customer relations. This highlights the need to attend to the interpersonal aspect of the interaction even in a seemingly purely transactional service encounter (O'Keeffe et al. 2007). The careful placement of *so* right before a request has a legitimizing effect on the pending action and makes it sound less imposing to the clients. Hence it serves to attend to the hearers' desire to be unimpeded and helps to create an overall harmonious atmosphere for the interaction.

Interactional functions

5.3.4.5 Processing device (processing)

So as a processing device indicates that the speaker is engaged in planning. Most of the instances of *so* in this category are used as a delay strategy and signal that the speaker is undergoing some processing problem and requires extra time. Under the pressure of real-time processing, speakers have to manage various aspects of the interaction simultaneously, including the control of timing and the change of direction if necessary (Sinclair and Mauranen 2006). Not surprisingly, this use of *so* is frequently accompanied by pauses, fillers and false starts. Example 42 shows the use of *so* in marking hesitation, emphasized by the co-occurrence of the fillers, *mm* and *um*, and another discourse particle, *well*:

(42)

a: …{ = [I] went < TO > } { = [< UM >] } { = [TO] the < U > k } { = [< FOR >] } { \ [< aBOUT >] } { = [THREE] < WEEKS > just } { \ to [VIsit] my < FRIENDS > there } { \ [< ***SO*** >] } { = [< MM >] } { = [< I >] think } { = [< _ UM >] } { \ [< WELL >] } { = [< BUT >] if } { = i [GOT] the < CHANCE > } { = i think [< I >] will } { \ [GO] to < SCOTland > }

(HKCSE, C121, 2736)

So is also found to mark self-editing in talk. While the use of *so* prefacing reformulation is quite common, *so* is seldom associated with word recovery problems and error replacements. When *so* is used in paraphrase, it indicates the speaker's attempt to promote the understanding of the interaction. This could be because the speaker is not satisfied with the way an idea is expressed earlier or s/he

feels that an alternative representation is required for the comprehension of the hearer. Example 43 shows the use of *so* in rephrasing by a teacher when she tries to explain the difference between active and passive voice to a student:

(43)

x: …{ \ [< ^ I >] am sharpening } (.) { \ [I] am sharpening the < _ PENcil > } (.) { \ [< S## >] is } { \ [< _ SHARpening >] } (.){ \ [< _ ***SO*** >] } (.) { \ the [PENcil] IS BEing < _ SHARpened > }…

(HKCSE, A036b, 955)

When compared with other functions, the proportion of *so* as a processing device in the HKCSE is relatively low. In the corpus, there are 483 tokens of *so* serving this function, making up 7.2% of the total discourse use.

5.3.4.6 Turn management (turn managing)

In terms of turn management, *so* indicates the current speaker's desire to keep the turn when interrupted, to give the conversational floor to other participants, or the hearer's attempt to take the turn from the current speaker. As a turn managing device, *so* is frequently found in turn transitional locations. In example 44, speaker A2's processing difficulties are evidenced by her false starts with *or* and the pauses. Despite her hesitation, speaker A2 still makes an effort in keeping her turn. This is realised through her repeated use of *so* while being interrupted by speaker A1:

(44)

A2: { \ oh that's [< RIGHT >] } { ? [< OR >] } (.) { = [< OR >] } (.) { ? [< ***SO*** >] } * { = [***SO***] < THAT'S > } { \ when they're [< ^ DOing >] it } …

A1: ** { ? [< NOW >] ((inaudible)) } { = [< YEAH >] }

(HKCSE, C020, 1886)

As a turn-taking or floor-holding device, *so* is commonly found in overlapping talk, as shown by the example above. When *so* is used to yield the turn, on the other hand, it is more likely to be found at the end of an utterance (cf. Jefferson 1983). Interestingly, the use of *so* to give away the turn is often echoed by the turn-taker. In other words, the previous speaker uses *so* for turn-yielding and the current speaker uses *so* for turn initiation. This phenomenon, which concerns the matching of discourse particles used by the speaker and the hearer in achieving a particular

interactional function, is referred to as “pairing” (Aijmer 2002:51; Bazzanella 1990). Example 45 illustrates the pairing of *so* in turn transition:

(45)

a: …{ = and [< THEN >] er } { ? it [CAME] back at < ONE > o' } { \ [ONE] o' < CLOCK > } { \ er in the [< afterNOON >] } { \ [BACK] and < FORTH > } { \ [< ONE >] a day } { \ [< ***SO*** >] }

B: { \ ***so*** just [ONE] a < _ DAY > }

(HKCSE, C039, 1446)

Speaker B’s choice of *so* to pick up the turn echoes the turn-yielding lexical selection of speaker a, which indicates his willingness to align with the previous speaker. This echoing of speakers helps to foster the relationship between participants, hence creating a friendly and collaborative ambience in discourse. It also contributes to the overall coherence of talk by means of lexical cohesion created by the repetition of *so* (Halliday and Hasan 1976).

When the particle *so* serves as a turn-yielding signal, it often conveys the meaning that something is left unsaid and can be inferred by the hearer from the information given by the speaker. Schiffrin (1987:223) notices *so* occurs when “a speaker has reached a point in the presentation of his/her ideas at which a hearer can infer what would come next even if it is not explicitly stated”. In other words, *so* invites the hearer to reconstruct the unstated message which has already been expressed in the preceding text, or which is otherwise self-evident. In example 46, speaker a3 uses *so* at the end of her talk to offer speaker a1 an extra cue for taking the turn. At the same time, she indicates that the unstated message following *so* can be easily recovered from her response earlier, i.e. she has no questions for speaker a1:

(46)

a1: { = [< MM >] } (.) { = have you got [QUEStions] for < US > then }

a3: { = [< ERM >] } { \ basically [< ^ NOT >] } * { = you've [< exPLAINED

a1: ** { / [< NO >] } { = [< MM >] mm }

>] } { \ the [TASK] very < WELL > ***so*** }

a1: { = [< oKAY >] } { = [< oKAY >] }

(HKCSE, B084, 2612)

The use of *so* in turn management is the least frequently occurring function in the Hong Kong corpus. Of the 6,721 discourse tokens of *so* found in the HKCSE, only 286 tokens are involved in turn managing activities. They constitute a mere 4.3% of all discourse use in the data, which suggests this function of *so* is probably auxiliary.

5.3.4.7 Unclassified instances

Owing to the dynamic and variable nature of language, and of spoken discourse in particular, it is unavoidable that the discourse functions of some tokens of *so* in naturally-occurring data remain undetermined. In the HKCSE, there are 147 unclassified instances of *so*. Compared to the total number of this linguistic item in the corpus, it is a relatively small proportion (1.8%). The ratio is also lower when compared with Müller's (2005) corpus-based study of *so*. In her data, approximately one-tenth of all instances of *so* are unclassified.

As discussed in Biber et al. (1999), in some cases the role of *so* is not always clear and its function is only interpretable from the situational context that is not available to analysts of a transcription or even a recording. Even if the situational context is sufficiently provided through the text, functions of *so* could still be unclassified due to inaudible speech or interruption. Some tokens are unclassified because there is a sudden change in the construction of the message. Speakers may abandon the segment containing *so* altogether and start with a completely different structure, which is not uncommon in unprepared speech given that parasyntactic structures are "not planned by the speaker from the outset, but rather incrementally in the course of speech production and on the basis of local decisions" (Mukherjee 2001:104). In such cases, it is unknown to the analyst what might have followed *so* originally if the idea had been allowed to develop further and thus it is not possible to be sure of an accurate interpretation. Example 47 shows the use of *so* in such a truncated structure, with the immediate occurrence of the concessive *but* marking the beginning of a new construction:

(47)

b1: …{ = [Anybody] who's been to the < PUtong > airport } { = [< ER >] } { \ [THEY] can < TELL > you that } { \ okay [< ^ ***SO*** >] } { = but [< ^ THESE >] } { \ are the [DIFferent] < iNItiatives > }…

(HKCSE, B094, 2862)

The majority of unclassified instances of *so*, however, are found in the repeated use of the word (see also example 44). The rather high frequency of successive occurrences of *so* in the corpus means that some consecutive uses are treated as unclassified if they seem to be serving a single function jointly with an adjoining occurrence and do not have an obvious discourse function of their own.

5.3.5 Linguistic factors associated with the functions of so

This section discusses the relationship between different linguistic features and the discourse functions of *so*. Three linguistic factors, namely, the collocations of *so*, its positional variation and its prosodic features will be described in turn in association with the six discourse functions identified above in order to determine how these variables are related to the functions identified.

5.3.5.1 Collocations

In the examination of the word associations with *so* in the corpus, it is observed that some discourse functions have higher tendency to co-occur with certain linguistic elements. When *so* is used a frame for the division of speech segments, it frequently co-occur with acknowledgement tokens *alright*, *okay* and *right*. In particular, they are more likely to appear immediately before *so* than after it. In the HKCSE, there are altogether 645 instances of *alright*, *okay* and *right* preceding *so*. The collocational pattern *okay so* is especially prevalent, amounting to 451 occurrences in the corpus. Since *okay* has the key meaning of acceptance and approval (Aijmer 2001), its placement right before *so* serves to indicate the finish of an idea unit and prepares for the transition which is brought about by *so*. In example 48 below, the lecturer first takes care of the operational aspect of the class by setting out some rules for the students. His use of *okay* in the middle of the excerpt indicates the completion of the house-keeping stage. He then uses *so* immediately after *okay* to introduce a meta-comment which signals the move to the stage where the subject matter of the lesson is discussed:

(48)

B: …{ = and [< I >] ask } { \ [^ ONE] thing from < YOU > } { / in [< reTURN >] } { = that [< YOU >] } { \ [< ^ TRY >] and } { = keep as [QUIET] as is < possiBLE > } { \/ and [TRY] and PAY as MUCH < ^ atTENtion > } { \/ to what i [< ^ SAY >] } { \ [AS] is < POSsible > } { /

[< oKAY >] } { = [< ***SO*** >] } { \/ the [< ^ FIRST >] topic } { = we're going to [START] < WITH > } { = is [< CONcern >] } { \ concerned [WITH] < ^ INventory > }…

(HKCSE, A005, 2996)

On the other hand, the use of *so* in its responsive function shows a different collocational pattern. When *so* is used to introduce a response initiated by the talk of another speaker, it is more likely to be preceded by *yes*, *yea* and its variants. In total, there are 209 instances of these occurrences found in the corpus. Other linguistic items which signal receipt of information and understanding such as *I see*, *mm*, *mhmm*, *uhuh* are also found before responsive *so*. In addition, a large number of interrogative words are found right after *so*, highlighting the responsive use of *so* before a question arising from the preceding information. There are altogether 421 instances of interrogative words following *so* in the corpus, including *how*, *when*, *where*, *which*, *who*, *why* and the most frequently occurring *what*. The responsive function of *so* in marking speakers' interpersonal attitude and evaluation is also emphasized by its readiness to be followed by personal pronouns. The number of D-use *so* directly followed by a personal pronoun (*I*, *you*, *he*, *she*, *we*, and *they*) is 2,097. This accounts for 31.2% of this position for all discourse use. In other words, on about one-third of the occasions when *so* is used as a discourse particle, it is followed by a personal pronoun.

In the interactional domain, the use of *so* is frequently accompanied by filled pauses to signal the planning process which speakers are undergoing. Within the ten word span of all instances of D-use *so*, there are 2,019 tokens of hesitation words *er*, *erm* and *um*. In combination with false starts, truncated words, incomplete structures and pauses, they underline the difficulties speakers have in real-time processing. The use of *so*, in turn, indicates to the hearer the speaker's desire to gain extra time or to reformulate a speech segment for the ease of comprehension by the hearer.

5.3.5.2 Positions

The analysis of the positions of D-use *so* in an utterance reveals some differences between the six discourse functions in the HKCSE. Table 5.8 shows how the discourse functions of *so* are distributed across an utterance:

Table 5.8. The positional distribution of the discourse functions of *so* in an utterance in the HKCSE[45]

Position (utterance)		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Final	Count	1	4	1	1	2	70	79
	% within Position	1.3%	5.1%	1.3%	1.3%	2.5%	88.6%	100.0%
	% within Function	.0%	.5%	.2%	.1%	.4%	24.5%	1.2%
Initial	Count	834	291	128	1025	146	81	2505
	% within Position	33.3%	11.6%	5.1%	40.9%	5.8%	3.2%	100.0%
	% within Function	34.0%	35.1%	19.6%	54.8%	30.2%	28.3%	38.1%
Stand-alone	Count	7	1	0	16	17	67	108
	% within Position	6.5%	.9%	.0%	14.8%	15.7%	62.0%	100.0%
	% within Function	.3%	.1%	.0%	.9%	3.5%	23.4%	1.6%
Medial	Count	1609	534	524	829	318	68	3882
	% within Position	41.4%	13.8%	13.5%	21.4%	8.2%	1.8%	100.0%
	% within Function	65.6%	64.3%	80.2%	44.3%	65.8%	23.8%	59.1%
Total	Count	2451	830	653	1871	483	286	6574
	% within Position	37.3%	12.6%	9.9%	28.5%	7.3%	4.4%	100.0%
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

So is most likely to occur in utterance medial position when it signals consequential relationship. The percentages of medial and initial *so* within this function are 80.2% and 19.6% respectively. In other words, the ratio of *so* in medial position to initial position is four to one in this function. This is hardly surprising as resultative *so* signals the relation between two adjacent discourse segments. The dominance of *so* in medial position is also found in other textual functions. The proportions of medial *so* in framing and linking are 65.6% and 64.3% correspondingly, meaning each makes up about two-thirds of the total within their functions. Again this indicates

[45] The 147 instances of *so* whose functions are unclassified are excluded in Table 5.8.

that the division or the connection of chunks of talk is more likely to occur in an extended stretch of text, where textual organisation is more necessary.

When *so* is used to preface response, on the other hand, the proportion of utterance initial *so* increases. More than half (54.8%) of the instances of *so* in this function are found at the beginning of the utterance. This reflects the use of *so* in marking reactions initiated by the previous speaker. The fact that only about one-fifth of responsive *so* are found in medial position suggests that *so* marking the evaluation of one's own prior talk is not as common as *so* marking a question or comment arising from the speech of another speaker.

In the interactional domain, about two-thirds of *so* as a processing device occur in utterance medial position. This signifies that *so* is more likely to mark hesitation and reformulation which are embedded in talk. While for most discourse functions *so* has a higher tendency to occur at the beginning or in the middle of an utterance, it is not the case for *so* in turn management. Of all the discourse functions, *so* as a turn managing signal has the most evenly spread distribution of positions. No single position accounts for more than 30% of the total within this function. When compared with other functions, it contains the greatest number of *so* in utterance final position (n=70), which are associated with turn-yielding. It also contains the greatest number of free-standing *so* (n=67), which are mostly associated with hearers' attempts to take the turn in simultaneous talk. Comparatively, the turn-keeping utterance medial *so* has the lowest proportion (23.8%) among all the discourse functions.

In terms of the positions of *so* in a tone unit, the six discourse functions show a less divergent pattern. Most of the instances of D-use *so* are found as a pre-head in the tone unit or in a tone unit on its own. However, slight variations can still be observed. Table 5.9 outlines the distribution of D-use *so* across various positions in a tone unit in the Hong Kong corpus:

Table 5.9. The positional distribution of the discourse functions of *so* in a tone unit in the HKCSE[46]

Position (tone unit)		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Head	Count	292	100	66	236	50	7	751
	% within Function	12.1%	12.2%	10.3%	12.9%	10.7%	2.5%	11.7%
Pre-head	Count	932	313	268	780	128	21	2442
	% within Function	38.7%	38.2%	41.9%	42.5%	27.4%	7.6%	37.9%
Tail	Count	27	13	7	14	10	18	89
	% within Function	1.1%	1.6%	1.1%	.8%	2.1%	6.5%	1.4%
Tonic syllable (in a shared tone unit)	Count	255	121	94	221	85	32	808
	% within Function	10.6%	14.8%	14.7%	12.1%	18.2%	11.6%	12.5%
Separate tone unit	Count	902	272	205	583	194	199	2355
	% within Function	37.5%	33.2%	32.0%	31.8%	41.5%	71.8%	36.5%
Total	Count	2408	819	640	1834	467	277	6445
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

One of the prominent features of Table 5.9 is the positional distribution of turn managing *so* in a tone unit. When compared with other discourse functions, *so* in turn management has the highest proportion of *so* constituting a separate tone unit, making up 71.8% of all instances serving this function. The prosodic independence of *so* in turn management function may be related to the fact that it contains the highest ratio of free-standing *so* of all the functions. In addition, its proportions as a head and a pre-head are only 2.5% and 7.6% respectively. These percentages are conspicuously low when compared with other uses. On the other hand, its proportion as a tail, which is 6.5% of the total within this function, is the highest. The majority of occurrences of *so* as a tail are found when speakers use *so* to instruct hearers to infer an unstated message and yield the turn to the hearer (see also example 46).

[46] Of the 6,583 instances of D-use *so* for which prosodic information is available, 138 tokens are functionally unclassified and are excluded from Table 5.9.

From the data available, the prosodic preference of *so* as a tail is probably characteristic of this use.

When compared with *so* working in the interactional domain, the use of *so* in achieving textual and interpersonal functions are more likely to be in a shared tone unit. This probably indicates that instances of *so* serving interactional functions are more prosodically detached than their counterparts in the textual and interpersonal domains.

5.3.5.3 Prosodic features

In Section 5.3.3, it is observed that the discourse and propositional use of *so* exhibit distinctive prosodic profiles. The comparison of the six discourse functions of *so* as regards their prosodic features, however, does not reveal such distinguishing differences. In general, the discourse functions of *so* display a largely consistent pattern in terms of their prominence distribution and their preference for nuclear tone type. Table 5.10 gives a summary of the distribution of prosodic prominence of the six discourse functions of *so* in the corpus:

Table 5.10. The prominence distribution of the discourse functions of *so* in the HKCSE[47]

Stress pattern		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Unstressed	Count	972	332	279	800	141	40	2564
	% within Function	40.4%	40.5%	43.6%	43.6%	30.2%	14.4%	39.8%
First	Count	272	94	61	227	46	6	706
	% within Function	11.3%	11.5%	9.5%	12.4%	9.9%	2.2%	11.0%
Middle	Count	7	0	1	3	1	0	12
	% within Function	.3%	.0%	.2%	.2%	.2%	.0%	.2%
Last	Count	23	12	5	7	5	1	53
	% within Function	1.0%	1.5%	.8%	.4%	1.1%	.4%	.8%
Sole prominent syllable (in a shared tone unit)	Count	232	109	89	214	80	31	755
	% within Function	9.6%	13.3%	13.9%	11.7%	17.1%	11.2%	11.7%
Sole prominent syllable (in a separate tone unit)	Count	902	272	205	583	194	199	2355
	% within Function	37.5%	33.2%	32.0%	31.8%	41.5%	71.8%	36.5%
Total	Count	2408	819	640	1834	467	277	6445
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

As shown in Table 5.10, the two most common prominence patterns for all discourse functions are unstressed and being the sole prominent syllable in a tone unit on its own. These two choices account for more than 70% of the total for all the discourse functions in the corpus. Most functions contain a similar ratio of unstressed and prosodically independent instances, except for the turn managing function, which has a noticeably higher proportion of *so* in a separate tone unit, which is discussed earlier in the previous section. All functions have a visibly low proportion of *so* as the last

[47] Of the 6,583 instances of D-use *so* for which prosodic information is available, 138 tokens are functionally unclassified and are excluded from Table 5.10.

prominent syllable or as a prominent syllable in the middle of the tone unit. They amount to less than 2% of the total for all uses. Some variations are found when *so* is the first prominent syllable or the only prominent syllable in a shared tone unit. Proportionally, far fewer instances of *so* are found as the first prominent syllable in turn management when compared with any other functions. Only 2.2% of turn managing tokens of *so* are the first prominent syllable in a tone unit. When used as a signal of processing, *so* has a relatively higher percentage of being the only prominent syllable in a shared tone unit when compared with other functions. This category makes up 17.1% of the total within this function.

Regarding the choice of nuclear tone on *so*, the six discourse functions also show a rather conforming pattern. Table 5.11 shows the distribution of tones for D-use *so* when it is the tonic syllable in the corpus:

Table 5.11. The distribution of nuclear tones for the discourse functions of *so* in the HKCSE

Tone		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Fall	Count	218	77	54	163	56	54	622
	% within Function	18.8%	19.6%	18.1%	20.3%	20.1%	23.4%	19.7%
Fall-rise	Count	4	3	2	5	0	2	16
	% within Function	.3%	.8%	.7%	.6%	.0%	.9%	.5%
Level	Count	866	289	220	548	192	142	2257
	% within Function	74.8%	73.5%	73.6%	68.2%	68.8%	61.5%	71.4%
Rise	Count	36	9	15	31	8	10	109
	% within Function	3.1%	2.3%	5.0%	3.9%	2.9%	4.3%	3.4%
Unclassified	Count	33	15	8	57	23	23	159
	% within Function	2.9%	3.8%	2.7%	7.1%	8.2%	10.0%	5.0%
Total	Count	1157	393	299	804	279	231	3163
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Level tone is most frequently chosen for all discourse functions of *so*, making up a proportion ranging from about 60% to 75%. Falling tone is the second most common choice, with a percentage of approximately 20% for all uses. Again the small degree

of difference is found in the turn managing function of *so*. When compared with the average, the turn managing function has a slightly higher ratio of tokens with falling tone and a slightly lower ratio of tokens with level tone.

In sum, it appears that the six discourse functions identified in the corpus have a converging prosodic profile, though slight differences could be found which may serve as suggestive cues for the interpretation of the discourse use of *so*.

5.3.6 Sociolinguistic and contextual factors associated with the functions of so

This section investigates the relations between sociolinguistic and contextual variables and the use of *so* in the corpus. In particular, it discusses whether language background, gender, domain and text type have an effect on the rate of D-use *so* and the functional distribution of the particle in the corpus.

5.3.6.1 Language background

In the HKCSE, it is found that there is a slightly higher rate of *so* as a discourse particle in the speech of native speakers of English when compared with Hong Kong Chinese. Table 5.12 details the distribution of *so* in the speech of Hong Kong Chinese and native speakers of English:

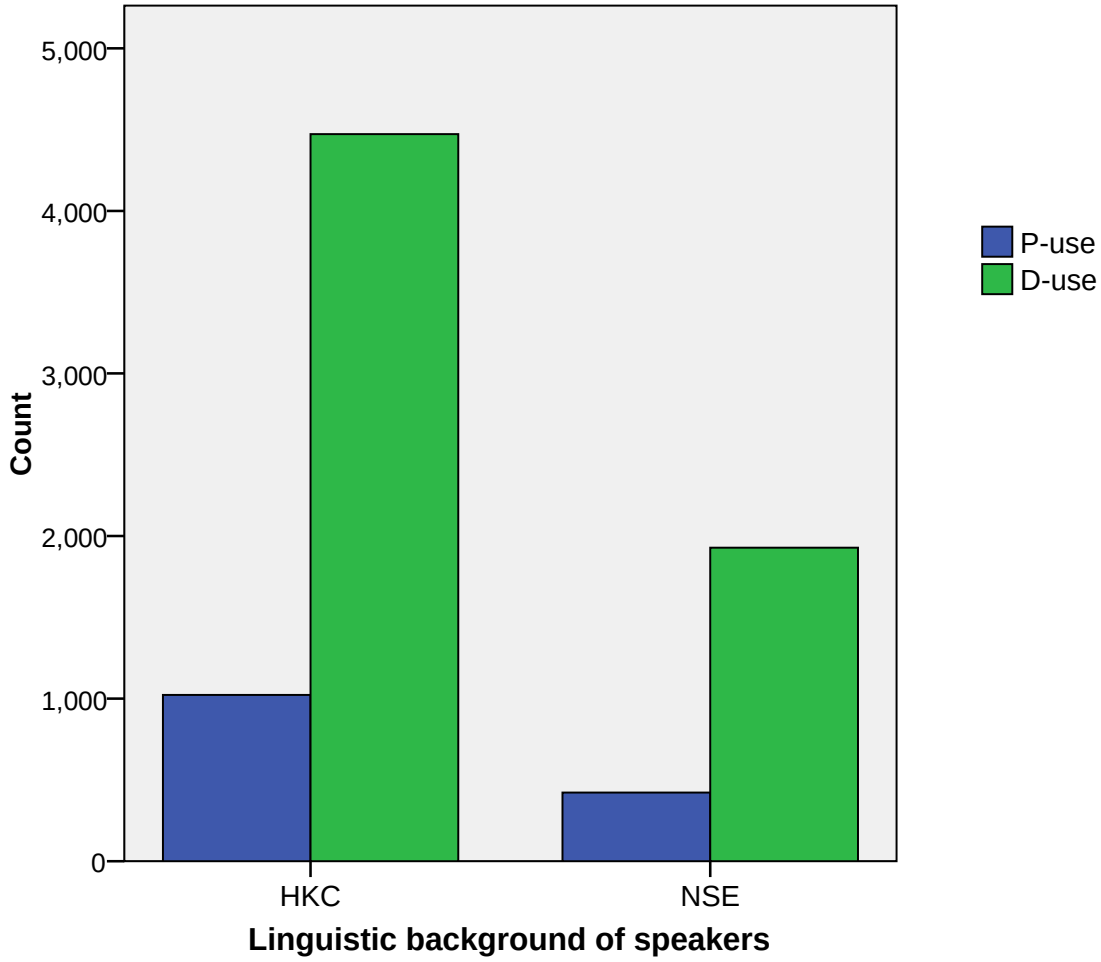
Table 5.12. The distribution of *so* in the HKCSE according to speakers' linguistic background (Hong Kong Chinese vs. native speakers of English)

	Hong Kong Chinese	**Native speakers of English**
Total number of words produced	669,431	240,847
Total number of so	5,526	2,368
Number of D-use	4,472	1,929
Number of P-use	1,023	422
D-use / total use (%)	80.93	81.46
D-rate (per 10,000 words)	66.80	80.09

From Table 5.12, it is observed that the total number of words in the speech of Hong Kong Chinese is more than two times that in the speech of native speakers of English in the corpus. Similarly, the total number of *so* produced by Hong Kong Chinese is approximately more than twice compared with native speakers of English. In other words, it seems that *so* is equally common in the speech of both speaker groups. In addition, the discourse-function ratio of *so* is analogous for Hong Kong Chinese and

native speakers of English (80.93% vs. 81.46%). This is supported by the fact that the difference in the discourse-function ratio between the two speaker groups is not statistically significant ($x^2 \geq 0.487$; d.f. = 1; p = 0.485; V = 0.008). In both speaker groups, roughly four out of five tokens of *so* are used as a discourse particle. Figure 5.4 shows a graphical representation of the distribution of D-use and P-use *so* by the two speaker groups:

Figure 5.4. The distribution of D-use and P-use *so* by Hong Kong Chinese (HKC) and native speakers of English (NSE)



While it is suggested elsewhere that there is an overuse of *so* as a discourse particle by Chinese EFL learners (He 2002), findings from the quantitative analysis of the HKCSE do not support this view concerning advanced Chinese speakers of English. From the intercultural corpus, it seems that the discourse rate of *so* is slightly higher in the speech of native speakers when compared with that in the speech of Hong Kong Chinese. Approximately 80 tokens of D-use *so* are found per 10,000 words in

the speech of native speakers of English. On the other hand, there are only about 67 tokens of D-use *so* per 10,000 words in the speech of Hong Kong Chinese.

When only the discourse functions of *so* are considered, the two speaker groups also seem to conform to a similar distribution. Table 5.13 shows the functional distribution of *so* based on the linguistic background of speakers:

Table 5.13. The distribution of the discourse functions of *so* by linguistic background in the HKCSE (Hong Kong Chinese vs. native speakers of English)[48]

Speaker (Linguistic Background, LB)		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Hong Kong Chinese	Count	1706	565	442	1182	336	152	4383
	% within Speaker (LB)	38.9%	12.9%	10.1%	27.0%	7.7%	3.5%	100.0%
Native speakers of English	Count	640	227	187	596	121	109	1880
	% within Speaker (LB)	34.0%	12.1%	9.9%	31.7%	6.4%	5.8%	100.0%
Total	Count	2346	792	629	1778	457	261	6263
	% within Speaker (LB)	37.5%	12.6%	10.0%	28.4%	7.3%	4.2%	100.0%

An examination of Table 5.13 indicates that no considerable differences are found between the two speaker groups in terms of how they use *so* as a discourse particle. Both Hong Kong Chinese and native speakers of English use *so* most frequently as a frame, followed by the use of *so* in prefacing responses. A slightly lower percentage of framing *so* and a marginally higher percentage of responsive *so* are found in the speech of native speakers of English. In other functions, the difference between Hong Kong Chinese and native speakers appears to be negligible. On the whole, Table 5.13 shows that no noticeable differences are found between the two speaker groups in terms of their usage of *so* as a discourse particle. The observation that non-native speakers use *so* less frequently on the textual level when compared with native speakers of English as reported in Müller (2005) is not noted in the present study.

[48] 135 instances of D-use *so* produced by speakers of other languages and the functionally unclassified tokens are excluded from Table 5.13.

5.3.6.2 Gender

Interestingly, the analysis of *so* based on speaker gender also generates results similar to those reported in the previous section as regards the discourse-function ratio of the word. No significant differences are found in terms of the ratio of *so* as a discourse particle per total use by speaker gender in the corpus. Table 5.14 summarises the distribution of *so* in the HKCSE with regard to the gender of the speakers:

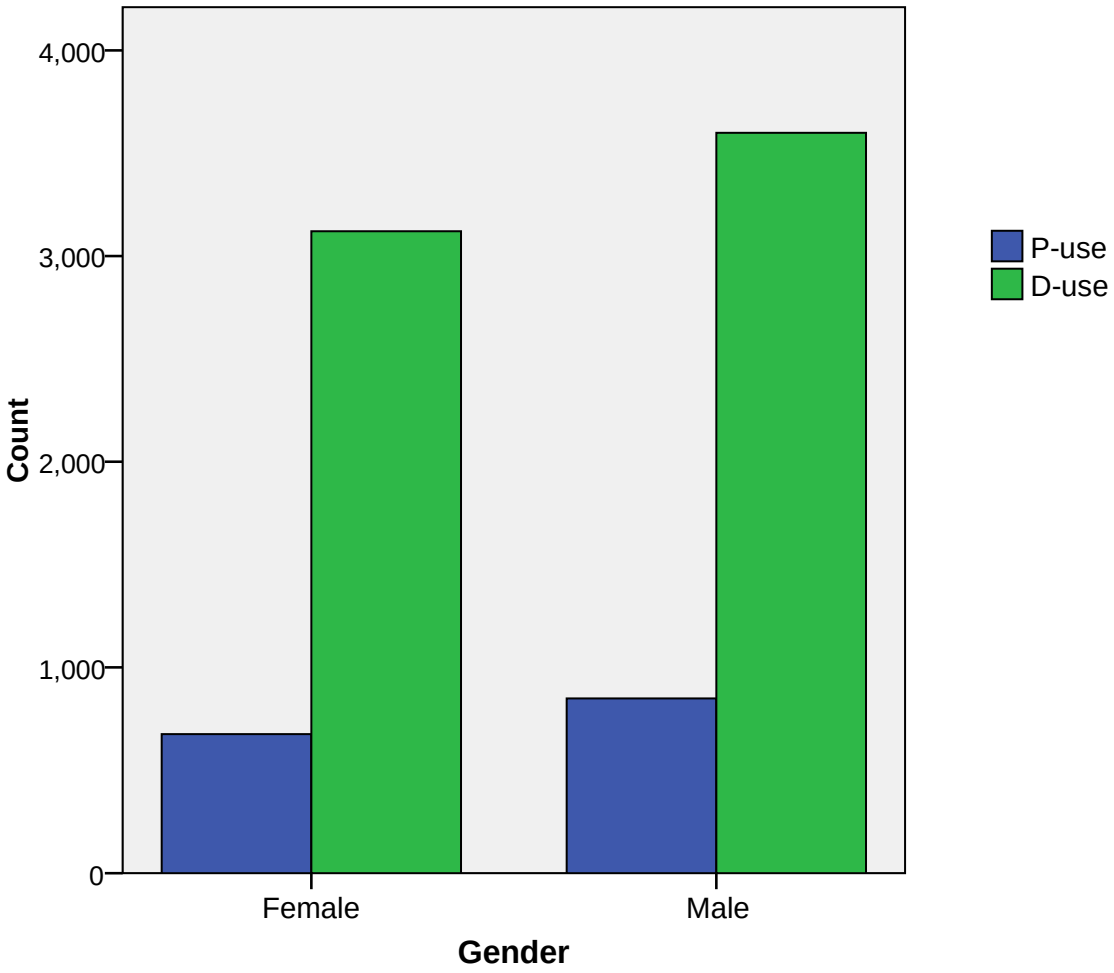
Table 5.14. The distribution of *so* in the HKCSE according to speaker gender (male vs. female)[49]

	Male	**Female**
Total number of words produced	556,815	389,659
Total number of so	4,478	3,817
Number of D-use	3,599	3,121
Number of P-use	849	676
D-use / total use (%)	80.37	81.77
D-rate (per 10,000 words)	64.64	80.10

As shown in Table 5.14, although male speakers contribute a higher number of *so* and a higher overall number of words in the corpus when compared with female speakers, there is little difference in the proportion of D-use and P-use *so* in male speech and female speech. This is reflected by the comparable discourse-function ratio of *so* in the two sexes (80.37% vs. 81.77%). The slight difference in the discourse-function ratio between men and women does not reach statistical significance ($x^2 \geq 2.238$; d.f. = 1; $p = 0.135$; $V = 0.016$). Figure 5.5 represents this comparison graphically:

[49] One token of D-use *so* produced by a speaker whose gender cannot be identified is excluded from Table 5.14.

Figure 5.5. The distribution of D-use and P-use *so* by female and male speakers in the HKCSE



From Table 5.14, women seem to have a slightly higher discourse rate of *so*. The discourse rate of *so* in women's speech is about 80 tokens per 10,000 words, as opposed to about 65 tokens in men's speech. This might suggest that there is a higher tendency for women to use *so* as a discourse particle, though it appears that no other gender studies have been done previously in relation to the discourse rate of *so* thus far to substantiate this claim.

As regards the discourse functions of *so*, the statistical analysis shows that gender has little effect on the functional distribution of *so* as a discourse particle. In fact, there is a high level of similarity between the two sexes as regards the spread of the six discourse functions. Table 5.15 illustrates in detail the resemblance of distributional pattern between male and female speakers:

Table 5.15. The distribution of the discourse functions of *so* by gender in the HKCSE (female vs. male speakers)[50]

Gender		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Female	Count	1172	399	304	818	229	132	3054
	% within Gender	38.4%	13.1%	10.0%	26.8%	7.5%	4.3%	100.0%
Male	Count	1279	431	349	1053	254	153	3519
	% within Gender	36.3%	12.2%	9.9%	29.9%	7.2%	4.3%	100.0%
Total	Count	2451	830	653	1871	483	285	6573
	% within Gender	37.3%	12.6%	9.9%	28.5%	7.3%	4.3%	100.0%

For both gender groups, *so* is used most frequently as a frame in the division of talk. More than one-third of all instances of *so* are used in framing for both sexes. The use of *so* in the initiation of response is also quite common. It accounts for 26.8% and 29.9% in female and male speech respectively. This is also the only category in which the difference between the two gender groups is statistically significant ($x^2 \geq 7.910$; d.f. = 1; $p < 0.01$; V = 0.035). As evidenced by the figures in Table 5.15, there is minimal difference in the proportion in which all the other functions are distributed for male and female speakers. For the consequential and turn managing functions of *so*, the percentages of use by gender are almost the same for both sexes. This distributional pattern from the HKCSE seems to suggest that there are more similarities than differences in the way in which men and women use *so* as a discourse particle.

5.3.6.3 Domain and text type

It is often suggested that the use of language is determined by the situational context in which speakers are engaged in. Given the fact that discourse particles serve important pragmatic functions, it is hardly surprising that their use is influenced by the setting in which they are produced. A quantitative analysis of the distribution of D-use *so* in the HKCSE gives support to the above statement. On the whole, the occurrences of *so* in the four contextual domains differ in their preferences for the six discourse functions. Table 5.16 illustrates the functional distribution of *so* as a discourse particle in the four sub-corpora of the HKCSE:

[50] The 147 functionally unclassified tokens and one token for which the speaker's gender cannot be identified are excluded from Table 5.15.

Table 5.16. The functional distribution of D-use *so* by domain in the HKCSE[51]

Domain		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Academic	Count	952	246	186	413	148	25	1970
	% within Domain	48.3%	12.5%	9.4%	21.0%	7.5%	1.3%	100.0%
Business	Count	737	272	189	499	133	72	1902
	% within Domain	38.7%	14.3%	9.9%	26.2%	7.0%	3.8%	100.0%
Conversational	Count	478	225	205	681	136	166	1891
	% within Domain	25.3%	11.9%	10.8%	36.0%	7.2%	8.8%	100.0%
Public	Count	284	87	73	278	66	23	811
	% within Domain	35.0%	10.7%	9.0%	34.3%	8.1%	2.8%	100.0%
Total	Count	2451	830	653	1871	483	286	6574
	% within Domain	37.3%	12.6%	9.9%	28.5%	7.3%	4.4%	100.0%

From Table 5.16, it is observed that framing is the most dominant function in the academic and business domains. In particular, almost half (48.3%) of the instances of *so* in the academic sub-corpus are used in marking boundaries in discourse. It should also be added that more than one-third of all instances of framing *so* are found in the academic domain. Of the 2,451 tokens of *so* acting as a frame, 952 instances come from the academic domain, making up 38.8% of the total. These findings provide strong evidence that the use of *so* serving the function of textual division is especially prevalent in academic discourse in the corpus. They also offer further support that particles are indispensable in teacher talk for discourse organisation (Sinclair and Coulthard 1975).

In the business domain, a slightly higher proportion of responsive *so* is found when compared with the academic domain (26.2% vs. 21.0%). This is probably due to the use of *so* in text types such as business meetings and job interviews in the business setting, which are more interactive than academic texts such as lectures and seminars and thus they allow for more occurrences of *so* in prefacing responses.

In the conversational sub-corpus, more than one-third (36.0%) of the tokens of *so* as a discourse particle are associated with responses. The highest proportion of

[51] The 147 functionally unclassified tokens are excluded from Table 5.16.

responsive *so* in conversations among all contextual domains highlights the interactive nature of multi-party talk. It is also worth noticing that conversations have the highest percentage of turn managing *so* (8.8%). Since participants enjoy perceived equal rights and status in conversations (Warren 2006), it follows that turn transitions are more flexible and variable when compared with some other text types such as seminars and service encounters, where allocations of turn are pre-defined by speaker roles and thus are more rigid.

In the public domain, the framing and the responsive categories are both popular, making up a fairly similar proportion of the total D-use (35.0% vs. 34.3%). This might be a result of the composition of the sub-corpus, which consists of about 50% of monologic texts such as speeches and press briefing and 50% of dialogues such as current affairs interviews.

In order to investigate the effect of different settings on the discourse functions of *so* in more detail, an examination of the functional distribution of *so* in different text types in the four domains is carried out. Table 5.17 represents the functional distribution of *so* in some of the monologues and dialogues which contain the highest numbers of the particle in the corpus:

Table 5.17. The functional distribution of D-use *so* across some major text types in the HKCSE

Text Type			Function						Total
			Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Dialogic	conversation	Count	478	225	205	681	136	166	1891
		% within Text Type	25.3%	11.9%	10.8%	36.0%	7.2%	8.8%	100.0%
	interview	Count	342	165	124	380	80	48	1139
		% within Text Type	30.0%	14.5%	10.9%	33.4%	7.0%	4.2%	100.0%
	meeting	Count	88	39	22	114	34	21	318
		% within Text Type	27.7%	12.3%	6.9%	35.8%	10.7%	6.6%	100.0%
Monologic	lecture	Count	403	115	47	127	60	1	753
		% within Text Type	53.5%	15.3%	6.2%	16.9%	8.0%	.1%	100.0%
	student presentation and Q&A	Count	223	72	88	131	44	16	574
		% within Text Type	38.9%	12.5%	15.3%	22.8%	7.7%	2.8%	100.0%
Sum		Count	1534	616	486	1433	354	252	4675
		% within sum	32.8%	13.2%	10.4%	30.7%	7.6%	5.4%	100.0%

As shown in Table 5.17, the three dialogic text types have a considerably higher proportion of responsive *so* when compared with the two monologic text types. Since the use of *so* in introducing reactions could convey a sense of cooperativeness (see also Section 5.3.4.4), it is especially useful for participants in the maintenance of personal relationship and the sharing of background knowledge, which is the main aim of conversations. In addition, as the use of *so* implies that the following response is motivated, it helps to shift the burden of some negative aspect of the response from the speaker, be it a criticism, request or a claim not fully justified. Therefore, *so* serves as a shield which protects speakers from an impeding face-threatening act by providing a signal to some justifying grounds. It is therefore also frequently found in meetings and interviews, when speakers often have to make well-substantiated assertions in the negotiation of ideas or express opposite views on sensitive issues.

The popularity of *so* as a frame in the academic domain is discussed earlier. It therefore comes as no surprise that lectures and student presentations and Q&A

sessions contain the highest numbers of *so* in monologic texts. As pointed out by Tsui (1995), teachers are involved in many aspects of classroom interaction. Apart from the teaching of the subject content of the lesson by means of verbal explanations, paraphrasing and the citing of examples and anecdotes, teachers also have to take care of the operational or practical aspect of the lesson. This includes the monitoring of the progress of the lesson and the whole course, the checking of students' comprehension, the engagement of students' participation as well as the exercising of disciplinary measures. Given the complexity of teacher talk, teachers have to employ a range of linguistic resources to ensure the smooth transition between different aspects of the ongoing discourse. The versatile nature of *so* in marking boundaries of various sorts facilitates the move between the various dimensions of classroom interactions, thus helping teachers in their overall organisation of talk.

5.3.7 Summary of findings in the HKCSE

Similar to findings from other corpus-related studies, *so* is one of the most popular linguistic items in the HKCSE evidenced by its high frequency. Despite the fact that it has a large number of difference senses and some uses of the word contribute to the propositional content of the utterance, most tokens of *so* found in the corpus are used as a discourse particle. By comparing the positional and prosodic distribution between the discourse use and propositional use of *so* in detail, the present study shows that *so* as a discourse particle has significantly different positional preferences and prosodic profiles when compared with its use which contains propositional value. These findings suggest that position and prosody are useful clues when determining whether *so* serves as a discourse particle.

Six major discourse functions of *so* spanning three functional domains are identified in the functional analysis of the data. The quantitative study of the discourse use of *so* indicates that more than half of the instances are associated with functions in the textual domain, followed by *so* working in the interpersonal domain, which make up less than one-third of the total. In comparison, the use of *so* achieving interactional functions is least common and only accounts for about one-tenth of the overall D-use in the corpus. The large number of *so* in textual functions indicates that *so* is chiefly a textual device in the data which plays an important role in discourse organisation, including the framing, linking and marking of result and consequence

between discourse segments. Less significantly in proportion, *so* introduces an interpersonal response which is initiated by some preceding information. The use of *so* before responses such as questions and comments could show the current speaker's attentiveness to some previous speaker. In prefacing face-threatening reactions such as criticisms and requests, *so* signals to the hearer that the impending face threat is motivated. This helps to justify the introduction of the act and thus preserve the negative face of the hearer. On the whole, *so* acting in the interpersonal domain contributes to the maintenance of personal relationship and facilitates the construction of a collaborative conversation. Finally, the small number of *so* working in the interactional domain in the corpus suggests that the discourse functions of signalling processing and managing turns are only supplementary.

Various linguistic, sociolinguistic and contextual factors are examined in relation to the discourse functions of *so* in order to investigate the extent to which these variables influence the distribution of discourse functions. Findings from the statistical analysis show that the six discourse functions converge in terms of their positional distribution as well as their prosodic profiles. While the two sociological factors, namely gender and linguistic background, seem to have little effect on the discourse functions of *so*, the contextual variable, i.e. domain and text type, gives important indications as regards the functional interpretation of *so* as a discourse particle. The fact that some discourse uses of *so* from authentic examples in the corpus are not discussed in great detail in previous work underlines the importance of looking at discourse particles in a diverse range of text types.

5.4 Comparison of findings in the HKCSE and the BNC customised corpus

In an attempt to investigate the effect of linguistic background of speakers on the use of *so* more thoroughly and to verify whether conclusions drawn from the core findings of the HKCSE are applicable to another data source, this section discusses the findings of *so* in the customised corpus of the BNC and compares them with the results generated from the Hong Kong corpus.

5.4.1 Frequency of occurrence

The customised corpus of the BNC (BNC customised) contains 948,920 running words which are evenly distributed across four contextual domains. In this tailor-made collection of 258 texts from the BNC, the word *so* is the 22nd most frequently

occurring item. Of the total 6,258 tokens of *so* found in the corpus, the discourse and propositional uses of the word amount to 4,652 and 1,487 occurrences respectively. The remaining 119 instances are unclassifiable owing to unintelligible passages. Table 5.18 shows the frequency distribution of the uses of *so* in the BNC (customised):

Table 5.18. The distribution of *so* in the BNC (customised)

	Total (N=258)	**Academic (N=25)**	**Business (N=86)**	**Conversational (N=122)**	**Public (N=25)**
Total number of words in BNC (customised)	948,920 (100.0%)	212,532 (22.4%)	260,484 (27.5%)	258,391 (27.2%)	217,513 (22.9%)
Total number of *so*	6,258 (100.0%)	1,828 (29.2%)	1,767 (28.2%)	1,801 (28.8%)	862 (13.8%)
Number of D-use	4,652 (100.0%)	1,507 (32.4%)	1,400 (30.1%)	1,251 (26.9%)	494 (10.6%)
Number of P-use	1,487 (100.0%)	289 (28.3%)	341 (33.4%)	524 (51.3%)	333 (32.6%)
Unclassified use	119 (100.0%)	32 (26.9%)	26 (21.8%)	26 (21.8%)	35 (29.4%)
D-use / total use (%)	74.34	82.44	79.23	69.46	57.31
D-rate (per 10,000 words)	49.02	70.91	53.75	48.41	22.71

For comparison, Table 5.19 shows the statistics of *so* generated from the HKCSE (originally presented in Table 5.1):

Table 5.19. The distribution of *so* in the HKCSE (reproduced from Table 5.1)

	Total (N=311)	**Academic (N=29)**	**Business (N=112)**	**Conversational (N=71)**	**Public (N=99)**
Total number of words in HKCSE	949,972 (100.0%)	213,204 (22.4%)	259,484 (27.3%)	258,882 (27.3%)	218,402 (23.0%)
Total number of *so*	8,296 (100.0%)	2,333 (28.1%)	2,329 (28.1%)	2,490 (30.0%)	1,144 (13.8%)
Number of D-use	6,721 (100.0%)	1,989 (29.6%)	1,953 (29.1%)	1,958 (29.1%)	821 (12.2%)
Number of P-use	1,525 (100.0%)	333 (21.8%)	360 (23.6%)	513 (33.6%)	319 (20.9%)
Unclassified use	50 (100.0%)	11 (22.0%)	16 (32.0%)	19 (38.0%)	4 (8.0%)
D-use / total use (%)	81.01	85.26	83.86	78.63	71.77
D-rate (per 10,000 words)	70.75	93.29	75.26	75.63	37.59

A comparison of the two tables above shows that the two corpora generally share the same distribution of *so* across the four contextual domains. While conversations,

academic and business discourse contain roughly the same number of *so*, the public sub-corpus has a considerably lower frequency of the word in both collections of texts. The same is true for the discourse use of the word. The public texts contain the lowest number of *so* as a discourse particle among the four sub-corpora in the HKCSE and the BNC (customised). For the propositional use of the word, the greatest number of occurrences is found in conversations for both the Hong Kong corpus and the British corpus. Another similarity between the two corpora is observed in the pattern of discourse-function ratio. For both sets of texts, the discourse-function ratio gradually decreases from left to right with regards the context of situation. The academic discourse has the highest ratio of *so* as a discourse particle per total use whereas the lowest is found in the public setting. Statistically, the difference in the discourse-function ratio between the four domains in the BNC (customised) is very highly significant with a moderate association ($x^2 \geq 228.201$; d.f. = 3; $p < 0.0005$; $V = 0.193$). This confirms the statement made earlier based on the findings from the HKCSE that the contextual setting affects the proportion of the D-use and P-use of *so*. Specifically, it is shown from the two corpora that *so* is most likely to be used as a discourse particle in the academic discourse.

In terms of the discourse rate of *so*, the two corpora also display a parallel trend among the four contextual domains. In the HKCSE as well as the BNC (customised), the academic setting has the highest discourse use of *so* per 10,000 words. This further emphasizes the popularity of *so* as a discourse particle in academic texts. While business texts and conversations have similar discourse rates of *so*, the public setting again has the lowest rate. In other words, the smallest number of D-use *so* is expected to be found in a 10,000 word sample of public text when compared with texts of the same size in other contexts. As regards the linguistic background of speakers, however, the BNC (customised) does not support findings from the HKCSE that native speakers have a slightly higher discourse rate of *so*. In the Hong Kong corpus, the discourse rates of *so* for Hong Kong Chinese and native speakers of English are about 67 and 80 tokens per 10,000 words respectively (see also Section 5.3.6.1). In the BNC (customised), the rate is roughly 49 tokens per 10,000 words only, which is comparatively lower than the figures for both speaker groups in the Hong Kong corpus. Therefore, the issue of whether Hong Kong Chinese or native speakers of English have a higher frequency of D-use *so* in their speech remains indeterminate and awaits further investigation.

5.4.2 Functional analysis

Although the BNC (customised) contains a lower discourse rate of *so* when compared with both the Hong Kong Chinese and native speakers of English in the HKCSE, its functional distribution of D-use *so* is analogous to the one found in the Hong Kong corpus. The comparison of the functional distribution of the discourse use of *so* in the HKCSE and the customised corpus from the BNC is shown in Table 5.20:

Table 5.20. The distribution of the discourse functions of *so* in the HKCSE and the BNC (customised)

Speaker (Linguistic Background, LB)			Function						Total
			Framing	Linking	Consequential	Responsive	Processing	Turn managing	
HKCSE	Hong Kong Chinese	Count	1706	565	442	1182	336	152	4383
		% within Speaker (LB)	38.9%	12.9%	10.1%	27.0%	7.7%	3.5%	100.0%
	Native speakers of English	Count	640	227	187	596	121	109	1880
		% within Speaker (LB)	34.0%	12.1%	9.9%	31.7%	6.4%	5.8%	100.0%
BNC (customised)	British speakers of English	Count	1645	453	493	1313	274	222	4400
		% within Speaker (LB)	37.4%	10.3%	11.2%	29.8%	6.2%	5.0%	100.0%

Except for 252 tokens of D-use *so* which are functionally unclassified owing to insufficient contextual information and inaudible speech, all other instances (n=4,400) of D-use *so* in the BNC (customised) are categorised according to the discourse functions they serve in talk. No other functions are discovered in the BNC (customised) in addition to the six discourse functions already identified in the HKCSE.

Table 5.20 indicates that all the discourse functions of *so* identified in the HKCSE are found in the BNC (customised). More importantly, there appears to be

little difference in the functional distribution of D-use *so* between the British speakers in the BNC (customised), the Hong Kong Chinese and the native speakers of English in the HKCSE. For all speaker groups in the two corpora, the largest number of *so* is found in serving textual functions. This give further evidence that *so* is chiefly a textual device. Specifically, about one-third of all instances of *so* are used as a frame regardless of linguistic background, making this the most common discourse function of the word. In the HKCSE, it is found that *so* marks a wide range of transitions in discourse on the macro and micro level. This is again substantiated by corpus evidence from the BNC (customised). Example 49 shows the use of *so* on the macro level in signalling a topic change. At the beginning of the talk, the speaker is explaining the drink making routine of her husband. Notice her placement of *so* after the pause to indicate the introduction of a new topic, i.e. the whereabouts of Andrea:

(49)

N: It's the one coffee I get made per day.
I get one in the morning and then at night.
Well he makes tea at night. He makes a cup of tea at night and that's his effort.
(pause) ***So*** er Sheila (gap)[52] was saying that erm (pause) Andrea's in Lon er Bath.

(BNC customised, extract from kb8-070004, sn8576, 1788)

On the other hand, *so* is also found in marking boundaries on the micro level in the BNC (customised). Example 50 illustrates the use of *so* in signalling the move from generality to specificity. In this case, the teacher shifts from a general statement to a specific example:

(50)

N: … Erm (pause) They dropped the final consonant in a co If there was like a consonant cluster at the end of the wor word.
So instead of saying kept they'd just say kep.

(BNC customised, g4v, sn194, 3545)

[52] The last name of the person is omitted for anonymity.

Similar observations could be made for the other discourse functions as regards the resemblance in proportion and parallel examples. The fact that the two corpora are highly comparable in terms of both qualitative and quantitative functional analysis of *so* gives further support to the claim made earlier that there appears to be no significant differences between Hong Kong Chinese and native speakers of English in terms of how they use *so* as a discourse particle.

5.4.3 Collocations

The collocational analysis of *so* presented earlier based on the findings of the HKCSE shows that the six discourse functions of *so* tend to co-occur with different lexical items (see Section 5.3.5.1). This observation is generally confirmed by the results from the BNC (customised), though a certain degree of variation regarding lexical associations is found between the two corpora.

In the textual domain, for example, it is observed that there is a tendency for *so* as a frame to co-occur with acknowledgement tokens in both corpora. When *so* is used to signal transitions in discourse, it is frequently found to be preceded by acknowledgement tokens *alright, okay* and *right* in the HKCSE. Similarly, in the customised corpus from the BNC, the lexical associations of framing *so* with these linguistic items are commonly found. In total, there are 253 instances of *alright, okay* and *right* immediately followed by *so*, compared to 645 instances in the Hong Kong corpus. While the collocational pattern *okay so* is particularly prevalent in the HKCSE, speakers in the BNC (customised) do not seem to have a strong preference for this combination. In fact, the associations *okay so* and *right so* are fairly similar in frequency in the British corpus, with *right so* being slightly more common (n=118 vs. n=105). Example 51 illustrates the use of *so* immediately after *right* as a frame to introduce a meta-comment which sums up what has just been discussed:

(51)

N: … and you simply may not be able to switch into milk production or horticulture, because you don't have the technological know-how, or the soil may be inappropriate, the climate may be inappropriate, right, ***so*** that's another factor, influencing, well making er, supply quite rigid, quite inflexible in, in the short run.

(BNC customised, hym, sn114, 3720)

The fact that the collocation *okay so* is not as common in the BNC (customised) when compared with the HKCSE might be related to the respective frequency of *okay* in the two corpora. In the HKCSE, there are altogether 4,909 tokens of *okay*. In the BNC (customised), in contrast, only 1,498 tokens of the word are found. On the other hand, *right* is a more frequently occurring lexical item in the British corpus. While there are 4,126 instances of *right* in the BNC (customised), the HKCSE only contains 3,252 instances of the word. The relative frequencies of the words *okay* and *right* in the two corpora may account for the difference observed above regarding the lexical combinations *okay so* and *right so*.

In the interpersonal domain, it is reported from the findings based on the HKCSE that responsive *so* is more likely to be associated with lexical items such as *I see, mm, mhmm, uhuh, yes* and *yea*. This trend, however, is not apparent in the BNC (customised). In the British corpus, it appears that a wide range of receipt signals including markers of surprise *ah, oh* and *really* are used in association with responsive *so* but none of them has an overriding presence. In addition, the acknowledgement tokens *okay* and *right*, which are common in framing, are also frequently employed in prefacing responses. The use of *so* in marking a follow-up response which is preceded by the reception marker *oh* is illustrated by example 52:

(52)

N1: Well when are they coming to take it back?

N2: Friday.

N1: Oh ***so*** it's a week thing?

N2: Mm.

(BNC customised, extract from kbw-022507, sn10357, 91)

Similar to the findings from the HKCSE, a large number of interrogative words are found following *so* in the BNC (customised). Altogether 274 instances of *how, what, when, where, which, who* and *why* are found, underlining the use of *so* in marking questions induced by preceding information. Further, the readiness of *so* to be followed by personal pronouns is reported in both corpora, which highlights the interpersonal functions of the particle. In the customised corpus from the BNC, there are 1,593 tokens of *so* directly followed by a personal pronoun (*I, you, he, she, we,* and *they*). This accounts for 34.2% of this position for all discourse use of *so* found in the corpus, which is comparable to the figure (31.2%) from the HKCSE.

When *so* works as a processing signal, its frequent co-occurrence with filled pauses is noted in the HKCSE. Although hesitation words such as *er, erm* and *um* are found in the neighbourhood of *so* when speakers are undergoing planning difficulties in the BNC (customised), they are generally less common when compared with the HKCSE. In the British corpus, only 683 tokens of these elements are found within the ten word span of D-use *so*, as opposed to 2,019 tokens in the HKCSE. In fact, the overall frequency of these hesitation words is much lower in the customised corpus. Altogether 32,124 tokens of *er, erm* and *um* are found in the HKCSE, as opposed to only 17,484 tokens in the BNC (customised). This is probably because native speakers of English are more likely to employ a variety of strategies such as repeated use of words, incomplete structures and the use of other discourse particles while they hesitate. Example 53 shows how a speaker in the BNC uses a wide range of tactics including the filled pause *erm*, the repeated uses of *so* and the incomplete structure *this was the* to signal the planning process he is going through:

(53)

N: … Nice work if you can get it.
Erm, and so on, ***so***, ***so***, ***so*** this was the, this was the, er this is what I was trying to do…

(BNC customised, he2, sn154, 3864-6)

On the whole, although there are slight variations as regards the frequencies of certain linguistic associations with *so*, similar lexical patterns are found for the different discourse functions of *so* in the HKCSE and the BNC (customised).

5.4.4 Utterance positions

In the HKCSE, it is noted that the particle *so* is mostly found in utterance medial position, followed by initial position. Utterance final and free-standing occurrences of *so*, in contrast, are much less frequent. This pattern is generally consistent with the one in the BNC (customised). Table 5.21 compares the positional distribution of *so* as a discourse particle in an utterance in the two data sources:

Table 5.21. The positional distribution of D-use *so* in an utterance in the HKCSE and the BNC (customised)[53]

Position	**D-use *so* in HKCSE**	**D-use *so* in BNC (customised)**
Final	82 (1.2%)	143 (3.3%)
Initial	2,560 (38.1%)	1,292 (29.4%)
Stand-alone	110 (1.6%)	41 (0.9%)
Medial	3,969 (59.1%)	2,924 (66.5%)
Total	6,721 (100%)	4,400 (100%)

In the two corpora, an overwhelming proportion of *so* is found either at the onset or in the middle of an utterance, making up more than 95% of the total number of D-use *so*. Despite the consistency in the general distributional pattern, the HKCSE and the BNC (customised) show a certain degree of disparity in the proportion of *so* in different positions. The most notable difference could be found in the ratio of initial *so* to the overall number. In the HKCSE, the percentage of *so* which occurs initially is 38.1%. In the BNC (customised), however, the percentage drops to 29.4%. On the other hand, there is a greater proportion of medial *so* in the customised corpus when compared with the HKCSE (66.5% vs. 59.1%). The same is true for the occurrences of *so* at the end of an utterance (3.3% vs. 1.2%). From a statistical point of view, all these differences are very highly significant (for initial position: $x^2 \geq 89.432$; d.f. = 1; $p < 0.0005$; V = 0.090; for medial position: $x^2 \geq 61.809$; d.f. = 1; $p < 0.0005$; V = 0.075; for final position: $x^2 \geq 55.278$; d.f. = 1; $p < 0.0005$; V = 0.071). It is thus worth further examining the spread of positions of D-use *so* across the six discourse functions in the British corpus for contrastive purposes with the HKCSE. The breakdown of the positional distribution is shown in Table 5.22:

[53] The 252 instances of *so* whose functions are unclassified are excluded in Table 5.21 for the sake of consistency in comparison with Table 5.22.

Table 5.22. The positional distribution of the discourse functions of *so* in an utterance in the BNC (customised)[54]

Position (utterance)		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Final	Count	8	3	6	9	2	115	143
	% within Position	5.6%	2.1%	4.2%	6.3%	1.4%	80.4%	100.0%
	% within Function	.5%	.7%	1.2%	.7%	.7%	51.8%	3.3%
Initial	Count	373	131	76	601	74	37	1292
	% within Position	28.9%	10.1%	5.9%	46.5%	5.7%	2.9%	100.0%
	% within Function	22.7%	28.9%	15.4%	45.8%	27.0%	16.7%	29.4%
Stand-alone	Count	0	0	0	4	0	37	41
	% within Position	.0%	.0%	.0%	9.8%	.0%	90.2%	100.0%
	% within Function	.0%	.0%	.0%	.3%	.0%	16.7%	.9%
Medial	Count	1264	319	411	699	198	33	2924
	% within Position	43.2%	10.9%	14.1%	23.9%	6.8%	1.1%	100.0%
	% within Function	76.8%	70.4%	83.4%	53.2%	72.3%	14.9%	66.5%
Total	Count	1645	453	493	1313	274	222	4400
	% within Position	37.4%	10.3%	11.2%	29.8%	6.2%	5.0%	100.0%
	% within Function	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 5.22 shows some interesting comparisons between the two corpora in terms of how the different discourse functions of *so* are distributed in the four positions (cf. Table 5.8). There is a generally lower proportion of initial *so* for all discourse functions in the BNC (customised) when compared with the HKCSE. The reverse is true for utterance medial *so*. Its proportion in the British corpus is higher for almost all discourse functions when compared with the Hong Kong corpus. These differences in proportion are most discernible for the framing and responsive functions of *so*. A detailed examination of the examples reveals that these differences in the proportion of positions are related to the respective ratio of sub-types in the functions. In the BNC (customised), a larger proportion of *so* is used in monologues to mark transitions which operate on a local level. Accordingly, most of them occupy utterance medial position. In the following example, the speaker first gives a general overview of the main topic that he is going to present, i.e. whether men and women

[54] The 252 instances of *so* whose functions are unclassified are excluded in Table 5.22.

speak the same language. After that he moves on to discuss the first sub-topic which is concerned with gender variation in sounds and pronunciation. This smooth transition is achieved through the sequencer *first of all* and the use of *so* in prefacing a meta-comment which introduces the next topic:

(54)

N: Right erm these notes are sort of based on a lecture that I went to when I was in sixth form erm and the theme for it is, Men and Women, Do We Speak The Same Language?

Erm ***so*** first of all sounds and pronunciation. Er men tend to have a lower pitch on the whole…

(BNC customised, g4w, sn36, 950)

Comparatively, fewer instances of *so* are associated with topic change in the BNC (customised). As discussed before, *so* in topic management is more likely to be in front position in talk. This results in a higher proportion of utterance medial *so* as a frame in the British corpus when compared with the Hong Kong corpus (76.8% vs. 65.6%). Similarly, in prefacing responses, there is a larger proportion of *so* marking speakers' personal evaluation to their own talk in the customised corpus. Example 55 illustrates the use of *so* in the middle of the speaker's talk in introducing his evaluation of the situation based on the information provided in his own prior discourse:

(55)

N: … I've got thirty four remaining ones to mark and I've got er all day tomorrow to do that and some on Thursday and possibly some time tonight.

So I'm I'm hopeful that I will comple well I'm certain that I will complete, barring accidents, the marking of final year assignments.

(BNC customised, jt0, sn42, 1075)

The larger number of tokens of *so* in prefacing self evaluation leads to a greater proportion of utterance medial *so* in responses in the BNC (customised) when compared with the HKCSE (53.2% vs. 44.3%). In combination, the higher ratio of medial *so* in framing and responses in the BNC (customised) indirectly reduces the proportion of utterance initial *so*. As a result, the percentage of *so* in initial position is comparatively lower in the British data. Since the framing and responsive uses of *so*

are two major functional categories in both corpora, the different proportions in sub-types of these two functions account for the positional variation observed above.

5.4.5 Domain and text type

It is proposed from the findings of the HKCSE that contextual settings serve as useful clues in the interpretation of the discourse functions of *so*. A comparison with the results from the British corpus substantiates this statement in general. Table 5.23 shows the distribution of the discourse functions of *so* across the four contextual domains in the BNC (customised):

Table 5.23. The functional distribution of D-use *so* by domain in the BNC (customised)

Domain		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Academic	Count	625	105	118	403	142	32	1425
	% within Domain	43.9%	7.4%	8.3%	28.3%	10.0%	2.2%	100.0%
Business	Count	493	129	180	405	58	68	1333
	% within Domain	37.0%	9.7%	13.5%	30.4%	4.4%	5.1%	100.0%
Conversational	Count	333	193	132	328	59	116	1161
	% within Domain	28.7%	16.6%	11.4%	28.3%	5.1%	10.0%	100.0%
Public	Count	194	26	63	177	15	6	481
	% within Domain	40.3%	5.4%	13.1%	36.8%	3.1%	1.2%	100.0%
Total	Count	1645	453	493	1313	274	222	4400
	% within Domain	37.4%	10.3%	11.2%	29.8%	6.2%	5.0%	100.0%

Similar to the figures from the HKCSE (see Table 5.16), a large proportion of *so* is used as a frame in the academic domain in the BNC (customised). Compared to other sub-corpora, the percentage of *so* serving the framing function (43.9%) is the highest in the academic texts. This supports the claim made earlier from the HKCSE that *so* plays a key role in marking boundary in the educational setting. In conversations, one-tenth (10.0%) of D-use *so* are associated with turn management, a proportion which is considerably higher than those in the other domains. This again is consistent with findings from the HKCSE and highlights the flexibility and variability of turn transitions in conversational talk. On the whole, the largest number of *so* is found in

serving textual functions across all contextual domains in the British data. This provides additional evidence for *so* being chiefly a textual device in discourse organisation.

When the discourse functions of *so* in different text types in the BNC (customised) are compared with results from the HKCSE (see Table 5.17), the difference in the use of *so* between monologues and dialogues as observed in the HKCSE is further corroborated. Table 5.24 represents the functional distribution of *so* across some of the monologic and dialogic texts which contain the greatest numbers of the particle in the British corpus:

Table 5.24. The functional distribution of D-use *so* across some major text types in the BNC (customised)

Text Type			Function						Total
			Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Dialogic	conversation	Count	333	193	132	328	59	116	1161
		% within Text Type	28.7%	16.6%	11.4%	28.3%	5.1%	10.0%	100.0%
	Interview	Count	195	48	91	225	28	42	629
		% within Text Type	31.0%	7.6%	14.5%	35.8%	4.5%	6.7%	100.0%
	Meeting	Count	50	18	14	56	9	6	153
		% within Text Type	32.7%	11.8%	9.2%	36.6%	5.9%	3.9%	100.0%
Monologic	Lecture	Count	335	40	30	87	29	2	523
		% within Text Type	64.1%	7.6%	5.7%	16.6%	5.5%	.4%	100.0%
	seminar / tutorial	Count	140	24	22	112	37	5	340
		% within Text Type	41.2%	7.1%	6.5%	32.9%	10.9%	1.5%	100.0%
Sum		Count	1053	323	289	808	162	171	2806
		% within Sum	37.5%	11.5%	10.3%	28.8%	5.8%	6.1%	100.0%

Table 5.24 shows that findings from the BNC (customised) largely confirm the view that *so* is more likely to achieve interpersonal functions in dialogues than in monologues. This is evident in multi-party texts such as meetings and interviews, where responsive *so* is the most frequently occurring function. In conversations in

the British data, however, the use of *so* in response does not seem to be in an overriding proportion. This is probably because of the difference in the composition of speakers in terms of familiarity level in the two corpora. While most conversations in the HKCSE are conducted between friends and acquaintances, a number of conversationalists in the BNC (customised) are couples. Since a high level of intimacy between participants may diminish the need to redress face damage, it is reasonable to suspect that fewer instances of *so* are used for politeness purposes, which leads to an overall reduction in the proportion of responsive *so* in the British data when compared with the Hong Kong corpus.

For the monologic talk in the BNC (customised), a high proportion of framing *so* is found in lectures, seminars and tutorials. The percentage is especially high in lectures (64.1%), where interactions are less frequent when compared with seminars or tutorials. The fact that the highest number of *so* in monologues is unanimously found in the educational setting for both corpora proves that *so* is seemingly indispensable in the structuring of teacher talk.

5.4.6 Summary of findings and comparisons between the HKCSE and the BNC (customised)

In general, the analysis above demonstrates that findings from the BNC (customised) are largely comparable to those from the HKCSE. This suggests that most of the statements made regarding the use of *so* based on findings from the HKCSE are confirmed by the British data. In terms of the frequency of *so*, the two corpora show very similar discourse-function ratios regardless of the first language of speakers, though the discourse rate of the word is lower in the BNC (customised) when compared with both Hong Kong Chinese and native speakers of English in the HKCSE. In other words, the question remains as regards which speaker group has a higher discourse rate of *so*. More importantly, however, is the fact that there is little difference in the functional distribution of *so* in relation to the linguistic background of speakers in both corpora. This provides strong evidence that Hong Kong Chinese and native speakers of English use *so* as a discourse particle in similar ways.

While the first language of speakers does not seem to be an influencing factor in the discourse functions of *so*, results from the BNC (customised) offer further support that lexical combination, position and contextual setting are important indicators of what function *so* achieves as a discourse particle. From the two corpora,

it is observed that the six discourse functions of *so* have different collocational patterns, positional as well as contextual preferences. This highlights the importance of taking various linguistic and contextual features into consideration in the interpretation of function of discourse particles.

Although results from the two corpora are mainly analogous, some minor variations are found which might point to differences in word choice, proportion of functional sub-types and the composition of speakers. Specifically, a number of collocational patterns are found to be less frequent in the customised corpus when compared with the Hong Kong corpus as British speakers seem to use a wider range of lexical items in combination with certain discourse functions of *so*. In addition, there are more instances of utterance medial *so* when compared with the HKCSE. This is probably due to the greater proportion of certain functional sub-types which usually occur in the middle of an utterance. Lastly, a lower proportion of *so* is used in response in the BNC (customised) which might be related to the higher level of intimacy between conversationalists in the British data.

5.5 Comparison of findings in authentic speech and textbooks

This section discusses the discourse use of *so* in the textbook database and draws a comparison between textbook descriptions and findings from the HKCSE and the BNC (customised). Three aspects of D-use *so* in the textbooks, namely frequency of occurrence, positions and functions, will be focused upon to investigate the similarities and differences in the use of *so* between authentic speech and teaching materials. A qualitative and quantitative functional analysis will be carried out for the pragmatic uses of *so* in the sample section, which contains presentations and discussions in the textbooks. For the teaching section which mostly contains short and detached examples and teaching points, the functional examination of the particle will only be qualitative.

5.5.1 Frequency of occurrence

In the sample section of the textbook database which contains 23,849 running words, there are 38 model presentations and 11 discussions. The total number of *so* found in the section is 156, making it the 22nd most frequently occurring word in the sample texts. Table 5.25 details the use of *so* in the sample texts and compares the figures with findings from the two corpora:

Table 5.25. The comparison of *so* in the textbooks, the HKCSE and the BNC (customised)

	Sample presentations in textbook database (N=38)	**Sample discussions in textbook database (N=11)**	**HKCSE (N=311)**	**BNC customised (N=258)**
Total number of words	10,817	13,032	949,972	948,920
Total number of *so*	47	109	8,296	6,258
Number of D-use	20	60	6,721	4,652
Number of P-use	27	48	1,525	1,487
Unclassified use	0	1	50	119
D-use / total use (%)	42.55	55.05	81.01	74.34
D-rate (per 10,000 words)	18.49	46.04	70.75	49.02

A comparison of the two parts in the textbook sample section reveals some interesting differences between the two sample text types regarding the use of *so*. Firstly, there is a greater total number of *so* in sample discussions when compared with presentations. While the total number of words in discussion texts (n=13,032) is only about 20% more than that in presentation texts (n=10,817), the total number of *so* in textbook discussions is more than twice that in presentations (n=109 vs. n=47). This suggests that the word occurs more frequently in textbook discussions. In addition, the discourse-function ratio is also higher for discussion texts. While 60 out of 109 tokens of *so* are D-use in discussions, only 20 out of 47 tokens are found to be D-use in presentations. In other words, a greater proportion of the word is used as a discourse particle in textbook discussions when compared with textbook presentations (55.05% vs. 42.55%), though this difference fails to reach statistical significance. Not only is there a higher overall number of *so* and a greater likelihood of it to be a particle when used, the discourse rate of *so* is also higher in textbook discussions. The rate of D-use *so* in textbook discussion texts is about 46 tokens per 10,000 words, as opposed to about 18 tokens in textbook presentations. This difference in the discourse rate of *so* between the two sample text types in textbooks achieves very high statistical significance ($x^2 \geq 13.420$; d.f. = 1; $p < 0.0005$; V = 0.024), which might suggest that *so* is more frequently occurring as a particle in textbook discussion texts.

When the findings of *so* in textbooks are compared with those from authentic data, the most notable difference could be found in the discourse-function ratio. In the HKCSE and the BNC (customised), the ratio of D-use to the overall use of *so* is 74.34% and 81.01% respectively. This means about three quarters or more tokens of *so* are discourse uses in the two corpora. In contrast, only about half of the total instances of *so* are used as a discourse particle in the textbook sample section. This highly significant difference ($x^2 \geq 138.137$; d.f. = 2; $p < 0.0005$; V = 0.097) in the discourse-function ratio of *so* between authentic speech and textbook data shows that the textbook materials underestimate the tendency of *so* being used as a discourse particle and over-emphasize the propositional use of the word.

As part of an attempt to determine whether the use of *so* in textbook presentations and discussions mirrors that in authentic presentations and discussions, a quantitative comparison in the discourse rate of *so* between textbook sample texts and similar text types in the two corpora is conducted. Table 5.26 shows the discourse rates of *so* in some major presentation text types in the two corpora and in the textbook presentations:

Table 5.26. The comparison of the discourse rate of *so* in presentations in the HKCSE, the BNC (customised) and the textbook database

Data source	**Text type**	**Discourse rate of *so* (per 10,000 words)**
HKCSE	business presentation	43.74
	lecture	121.96
	speech	21.14
BNC	business presentation	69.32
	lecture	86.13
	speech	15.26
Textbook database	sample presentation	18.49

As indicated by the figures from Table 5.26, sample presentations in textbooks appear to contain a lower discourse rate of *so* than most authentic presentations. Compared with the major presentation text types in the HKCSE and the BNC (customised), the discourse rate of *so* in textbook presentations is most similar to that in speeches in the two corpora. Given the fact that sample presentations in textbooks are supposed to be fairly spontaneous and only slightly prepared, it is unsure whether

speeches are the most suitable text type for comparison, as they are usually heavily scripted. In this sense, the nature of textbook presentations probably resembles lectures most among the three presentation text types, since they are normally only semi-prepared and unscripted. Unfortunately, Table 5.26 shows that there is a considerable discrepancy in the discourse rate of *so* between textbook presentations and lectures or even business presentations. When compared with the two text types in authentic data, textbook presentations have a noticeably lower discourse rate of *so*. As demonstrated earlier by the findings from the two corpora, the particle *so* serves important functions as a textual device, especially in monologues. Hence it remains unclear why such a low discourse rate of *so* is observed in textbook presentations if they truly reflect real usage of the word. This seems to imply that the use of *so* in largely monologic presentations, especially its textual aspect, is overlooked by textbook writers.

In comparison, authentic discussions and textbook discussions are more alike in terms of the discourse rate of *so*. Table 5.27 compares the discourse rate of *so* in discussion texts in the HKCSE, the BNC (customised) and the textbooks:

Table 5.27. The comparison of the discourse rate of *so* in discussions in the HKCSE, the BNC (customised) and the textbook database

Data source	**Text type**	**Discourse rate of *so* (per 10,000 words)**
HKCSE	business interview	81.18
	business meeting	90.15
	public TV talk show	57.14
BNC	business interview	55.79
	business meeting	38.28
	public TV talk show	20.96
Textbook database	sample discussion	46.04

It is observed earlier that the overall discourse rate of *so* is higher in the HKCSE when compared with the BNC (customised). This pattern is noted again in Table 5.27 when the discourse rates of *so* in business interviews, business meetings and public TV talk shows in the two corpora are compared. For textbook discussions, the rate of D-use *so* is about 46 tokens per 10,000 words. The figure is especially comparable to business interviews and meetings in the BNC (customised). This indicates that there

is a high level of similarity in the frequency rate of *so* as a discourse particle between authentic discussions and textbook discussions.

5.5.2 Positions

In the analysis of *so* in the HKCSE and the BNC (customised), it is found that there is a dominance of *so* as a discourse particle in utterance initial and medial positions. This is consistent with figures from the textbook sample section. Table 5.28 gives a comparison of the positional distribution of *so* as a discourse particle in an utterance in the three data sources:

Table 5.28. The positional distribution of D-use *so* in an utterance in the HKCSE, the BNC (customised) and sample texts in the textbook database

Position	**D-use *so* in HKCSE**	**D-use *so* in BNC (customised)**	**D-use *so* in sample texts in textbook database**
Final	82 (1.2%)	143 (3.3%)	0 (0.0%)
Initial	2,560 (38.1%)	1,292 (29.4%)	23 (28.8%)
Stand-alone	110 (1.6%)	41 (0.9%)	0 (0.0%)
Medial	3,969 (59.1%)	2,924 (66.5%)	57 (71.3%)
Total	6,721 (100%)	4,400 (100%)	80 (100%)

In fact, Table 5.28 shows that the proportion of initial and medial D-use *so* is even more overwhelming in textbook data. Of the 80 tokens of *so* used as a discourse particle, none of them is found free-standing or at the end of an utterance. All of them occur either at the beginning or in the middle of an utterance. Proportionally, textbook sample section contains the highest percentage of utterance medial D-use *so* among the three sources of data. Altogether 57 tokens are found embedded in discourse in textbooks, making up 71.3% of the total discourse use. In particular, all the instances of *so* as a discourse particle in textbook presentations occur medially. Example 56 shows an utterance medial *so* in a presentation text in textbooks:

(56)

T: … According to the passage, their first record sold over a million copies, ***so*** I suppose they've been very successful…

(textbook database, sample presentation 5.3, 103)

The occurrence of *so* in utterance initial position, in contrast, is less common. This confirms findings from the two corpora that *so* is less likely to occur initially. Altogether only 23 tokens of utterance initial *so* are found in sample texts, all of them in textbook discussions. They constitute approximately 28.8% of the total D-use. Example 57 contains an occurrence of *so* at the onset of an utterance in a discussion text in textbooks:

(57)

T1: … There's too much to learn, you know, there's too much in the syllabus. We just don't have enough time to learn it. And it's very difficult to remember things properly when you have to rush through them.

T2: ***So***, do you think it is the teachers' fault, then?

(textbook database, sample discussion 3.1, 421)

In sum, the positional distribution of the discourse use of *so* in textbook sample texts largely follows the pattern observed in the two corpora. This indicates that textbook authors provide a fairly accurate portrait of *so* as a discourse particle as far as its typical positions in an utterance are concerned.

5.5.3 Functional analysis

This section presents a detailed analysis of the discourse functions of *so* in textbooks, which is divided into two parts. Firstly, the teaching section of the textbooks will be studied thoroughly in order to give a qualitative report on whether and how the discourse use of *so* is described in the teaching materials. Secondly, the sample section in the textbook database will be examined to see how the different discourse functions of *so* are distributed quantitatively across textbook presentations and discussions.

Owing to the exam-oriented nature of the textbooks under investigation, the teaching section of the textbook database is mostly devoted to the description of oral skills which are most relevant to the assessment of the target examination. In other words, the bulk of the section focuses on illustrating how to build effective presentation and discussion techniques. Consequently, most tokens of *so* in the teaching materials are found in the examples of various speech functions which are considered essential in the two text types. Teaching points which specifically explain the use of *so*, or other discourse particles for that matter, are minimal.

Although most of the examples of speech functions are short and contextually void, a survey of such examples containing *so* could still provide indications as regards the kinds of functions textbook writers suggest to be associated with the particle. In the present analysis, all speech functions which involve the use of *so* in the teaching section could be classified into the functional categories developed from the two corpora. No other major functions which are absent in authentic speech are identified in the teaching materials, though variations of functional sub-types and linguistic representations are found.

In the teaching section of the textbook database, many instances of *so* are found to be associated with various textual functions. One speech function which is most frequently achieved with the presence of *so* is concluding or finishing. In these examples, *so* is commonly found at the beginning of the sentence to summarise a point in presentations or discussions or to draw a close. Example 58 shows the use of *so* in an example which illustrates how to make brief summaries which occur throughout the discussion:

(58)

T: ***So*** we've got three suggestions so far.

(textbook database, teaching section, 17388)

Example 59 shows a textbook suggestion of how to bring a presentation to an end:

(59)

T: ***So*** that's the end.

(textbook database, teaching section, 9392)

The above examples largely correspond to the conclusive use of *so* found in the corpora, which is a sub-type of the framing function. It is not at all clear though how decontextualised examples like the above could exemplify satisfactorily summarising or concluding if the preceding context is not provided. Another speech function in the textbooks which can vaguely point to the framing use of *so* is "managing the discussion" (Gran 1993:27). Again these examples mostly contain meta-discourse which serves to keep track of the progress of the talk, as in example 60 with the sequencers:

(60)

T: ***So*** first, … secondly… finally…

(textbook database, teaching section, 18110)

While the use of *so* in signalling topic change or digression is common in the corpus data, it is not found in textbooks. The lack of *so* in topic management in combination with the extensive use of *so* in prefacing conclusions and summaries in textbook descriptions seems to suggest that *so* is confined to closing and summing up in presentations and discussions, which is actually not the case in authentic speech.

As regards the linking and consequential textual functions of *so*, they are only mentioned briefly in the teaching section as teaching points though no examples are provided to illustrate how they are exactly realised. Terms given to *so* which highlight its connective aspect include "linking device" (Esser 1999:40) and "short, common connector" (Li 2003:26). In one of these descriptions, *so* is suggested to be a common connecting word in speech among others such as *and*, *or*, *then* and *though* (Leetch 2002). However, no further explanations are offered concerning the similarities and differences between these connecting words. Similarly, the consequential aspect of *so* is only dimly alluded to when the textbooks discuss how to indicate a result or an effect. Again it is cited among other lexical items such as *finally*, *consequently*, *hence* and *therefore* without any illustrations of how they are used.

In the interpersonal domain, *so* is found in a number of response-related speech functions. This echoes the responsive use of *so* in questions, comments and requests in corpus data. In particular, *so* is most commonly related to the speech function of checking understanding in textbooks. Many examples for checking and clarification purposes are in the form of questions and involve the use of *so*, including example 61:

(61)

T: ***So***, are you saying that…

(textbook database, teaching section, 8998)

In the teaching of presentation skills, *so* is also found to introduce personal reactions which are supposedly motivated by the speaker's prior talk, though the preceding context is never provided in such examples:

(62)

T: ***So*** as you see, I wholeheartedly agree with the passage and agree that we must do far more to save the Chinese dolphin…

(textbook database, teaching section, 10627)

In addition, *so* is also associated with examples which entail face-threatening acts in textbooks. They are usually found in the teaching of the skills for discussions, where making requests or suggestions and assigning responsibilities are common. Example 63 illustrates the use of *so* in making a request for starting the discussion:

(63)

T: I'm Candidate A, ***so*** I'll start, shall I?

(textbook database, teaching section, 8289)

In the above example, the marking of the request with *so* implies that the speaker's wish to initiate the discussion is based on some reason: the fact that s/he is Candidate A, who should be given the first opportunity to speak if an alphabetical order is followed. As a result the appeal is more legitimized, making it less face-threatening and less imposing on the hearer.

When compared with the interpersonal function of *so*, the interactional aspect of the particle is less adequately addressed in the textbooks. In particular, the use of *so* as a processing device is not found in the teaching section. As far as turn management is concerned, only the use of *so* in turn-yielding is found in speech functions in textbooks. Example 64 shows an instance of "inviting others to add some more points" in the textbook teaching section (Esser 1999:57):

(64)

T: ***So***, does anyone want to add some more points here?

(textbook database, teaching section, 13974)

Although the utterance position of *so* is not absolutely clear from the example above owing to the lack of contextual information, it obviously does not occur at the end of the utterance, which is the most typical position for turn-yielding *so* in the corpus data.

In the textbook sample section, *so* is found in texts which basically follow the structure of a presentation or a discussion. Consequently, more contextual information is available for the interpretation of discourse functions. Table 5.29 outlines the functional distribution of *so* in the two sample text types in textbooks:

Table 5.29. The functional distribution of *so* in the sample presentation and discussion texts in the textbook database

Text type in textbooks		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Sample presentation texts	Count	0	3	6	11	0	0	20
	% within Function	.0%	15.0%	30.0%	55.0%	.0%	.0%	100.0%
Sample discussion texts	Count	24	0	6	28	0	2	60
	% within Function	40.0%	.0%	10.0%	46.7%	.0%	3.3%	100.0%

As discussed earlier, *so* is chiefly a textual device in the two corpora. Its use in achieving textual functions is especially prevalent in monologic texts (see, for example, Section 5.4.5). Contrary to the findings from the two corpora, the particle *so* in textbook sample texts shows little difference in the proportion of textual and interpersonal functions. In particular, no instance of *so* is found as a frame in textbook presentations. This is in marked contrast to the results from the HKCSE and the BNC (customised), in which a large proportion of *so* is used in marking boundaries in monologues.

On the other hand, the interpersonal function of *so* appears to be sufficiently presented by the examples in the sample section. In both presentation and discussion texts, about half of the instances of *so* as a discourse particle preface responses. The proportion is especially high in textbook presentations when compared with authentic presentations.

In addition, a much lower proportion of *so* is found serving interactional functions in textbooks when compared with authentic speech. Only 3.3% of tokens of D-use *so* in textbook discussions are associated with turn management. All of them are used in turn-yielding. Moreover, no examples of *so* as a processing device are found in the two sample text types.

5.6 Pedagogical implications

In the previous sections, the frequency of occurrence, positions and functions of *so* in textbooks are studied and compared with the findings from the HKCSE and the BNC (customised). In terms of frequency of use, it is observed that the discourse-function

ratio of *so* in textbooks is lower than that in the two corpora. In addition, the sample presentations in textbooks have a lower discourse rate when compared with most authentic presentations, which suggests that the use of the discourse particle *so* in monologic presentations is largely overlooked by textbook authors.

In comparison, the positional distribution of D-use *so* in textbooks is generally consistent with that in the HKCSE and the BNC (customised). Most tokens of *so* in the sample section are found in utterance initial and medial positions. This echoes results generated from authentic speech. Slight variations are found in the higher percentage of medial *so* and the absence of final and free-standing *so* in textbooks, which do not seem to affect the general positional pattern. This indicates that the positions *so* occupies in textbook sample texts basically reflect authentic usage.

Functionally, the analysis of the teaching and sample sections in textbooks shows an unsatisfactory portrayal of *so*. Given the popularity of *so* in spoken discourse, it is rather surprising that writers of oral textbooks make little effort to elucidate how it is used. Detailed descriptions of how the particle is used are practically missing; even examples of speech functions containing *so* are mostly short and decontextualised. This presents an additional challenge to interpreting the discourse functions of *so* as it is difficult to determine whether a particular speech function is achieved by a given example without the provision of the neighbouring co-text. Although the use of *so* is found in all the three main functional domains in textbooks, its function as a processing signal is neither discussed in the teaching section nor realised in the sample section. Some sub-types of certain functional categories are also absent. Further, discrepancies are observed in the functional distribution of *so* across presentations and discussions between invented and authentic data, especially regarding the framing function of *so*. The importance of *so* as a textual device, which is evidenced by the occurrences from the corpora, is not made obvious in textbooks either by its quantity or its contextually deficient examples. This makes one doubtful whether the textbooks being examined reflect the real usage of *so* as a discourse particle, not to mention to impart that piece of knowledge to students.

Summary

In this chapter, a comprehensive analysis of *so* is conducted by a comparison of its use in three different sources of data. As the background to the present study, the chapter first reviews some previous studies of the particle *so* and explains with

examples how the discourse use and propositional use of the word are distinguished. It then presents a detailed examination of the use of *so* in the HKCSE. From the qualitative and statistical findings of the Hong Kong corpus, it is shown that the uses of *so* in expressing discourse and propositional meaning differ in terms of their positions and prosodic profiles. As a discourse particle, *so* is found to serve six major pragmatic functions in the textual, interpersonal and interactional domains, showing varying degrees of linguistic, sociolinguistic and contextual deviations.

In an attempt to further investigate the influence of linguistic background of speakers on the use of *so*, a customised corpus from the BNC is then examined and compared with findings from the HKCSE. This produces largely comparable results between the two corpora. Most importantly, the customised corpus provides additional proof that there is no significant difference between Hong Kong Chinese and native speakers of English as regards the discourse functions of *so*, though minor variations are found in the areas of collocation and the proportion of functional sub-types which may be attributed to the difference in word choice between the two speaker groups as well as the composition of the two corpora.

Lastly, the issue of whether teaching materials reflect authentic usage of *so* is addressed. The descriptions and occurrences of *so* in a textbook database are inspected and compared with findings from the two corpora. Results from the comparison indicate that although some similarities are observed between textbook data and authentic speech including the positional distribution of D-use *so*, the textbooks examined fail to provide a thorough account of the particle. Descriptions of some discourse functions of *so* are often not accompanied by examples. When examples of speech functions associated with *so* are provided, they are mostly short and decontextualised. In sum, the textbooks under investigation in the present study fall short of completely reflecting the real usage of the particle *so*, let alone of convincingly exemplifying to learners of English the various discourse functions of this highly common and versatile particle.

Chapter 6

Towards a functional framework of discourse particles

Abstract

In this chapter, a comparison is made between the discourse functions of *well* and *so* identified in the two spoken corpora to provide an account of the similarities and differences between the uses of the two particles in authentic data. Following a brief introduction (6.1), a two-part analysis which contrasts the discourse functions of *well* and *so* both qualitatively (6.2) and quantitatively (6.3) will be presented. This contrastive analysis serves to further our understanding of the ways in which *well* and *so* work in discourse and to demonstrate that the functional framework developed from the corpus data in the present study supplies a platform on which comparison could be drawn. This suggests the potential of applying the present functional framework to other discourse particles for a holistic approach to the description and interpretation of these linguistic items (6.4).

6.1 Are *well* and *so* interchangeable?

In the two previous chapters on *well* and *so*, it is found that the two discourse particles both operate in three functional domains: textual, interpersonal and interactional. In the textual domain, *well* and *so* are involved in the structuring and organisation of talk. Interpersonally, they are responsible for introducing speakers' attitudes, evaluations and feelings towards the previous discourse. As far as the interactional domain is concerned, the two particles are useful devices in the speakers' planning process and turn managing activities. From the findings discussed in these two chapters, it is also noted that the two discourse particles *well* and *so* share some of the functional categories, suggesting that some of their discourse functions are probably analogous to each other. Naturally it is tempting to ask, how similar and different are *well* and *so* as discourse particles? Are they freely interchangeable in discourse?

As a prerequisite for the above question, it is necessary to recapitulate the major discourse functions of *well* and *so* identified from the corpus data in the present study:

Table 6.1. Comparison of discourse functions of *well* and *so*

Functional categories	**Textual**			**Interpersonal**		**Interactional**	
	Framing	Linking	Consequential	Responsive	Emotive	Processing	Turn managing
well	+	+		+	+	+	+
so	+	+	+	+		+	+

It is observed from Table 6.1 that out of the seven major functional categories developed from the corpus data, *well* and *so* both realise functions in five categories, namely framing, linking, responsive, processing and turn managing. Only consequential and emotive functions are not achieved by both particles. Table 6.1 thus seems to suggest that *well* and *so* are to a large extent functionally equivalent. It confirms remarks made in earlier studies that the two discourse particles closely resemble each other (see, for example, Bolinger 1989; Schourup 2001). This conclusion, however, offers only a simplistic view. As demonstrated by the detailed functional analyses in the last two chapters, actual realisations of the functions could differ even if the two particles contain the same functional categories. Further, sub-types of functions could be present in one particle but not the other. The following sections provide a contrastive examination of the discourse functions of *well* and *so* based on the findings from Chapters 4 and 5. Key observations which were made in the previous two chapters individually regarding the two particles are taken together and studied from a comparative perspective. Instead of replicating what has already been presented, this contrastive analysis makes explicit the major similarities and differences between the discourse functions of the two particles, which in turn paves the way towards a fuller description of *well* and *so*.

6.2 Qualitative contrastive analysis of *well* and *so*

This section compares and discusses the partially overlapping ways in which *well* and *so* are used to achieve five discourse functions, which are grouped under the three functional domains: textual, interpersonal and interactional.

Textual functions

6.2.1 Frame (framing)

While framing is one of the major functions for both *well* and *so*, the two particles show varying degrees of associations with different types of breaks. Broadly

speaking, *well* is more commonly used to introduce disparate topics and discourse stages while *so* has a stronger tendency to signal moves between related aspects of the same topic. For example, while *so* is possible in the following scenario indicating a topic change, *well* is more frequently found to achieve this function in the data:

(1)

A: { = the [< DOCtor >] }

b: { \ the [< DOCtor >] } { \ [< ^ YEAH >] }

B: { \ [< Mhm >] }

((pause))

b: { \ ***well*** [C##'S] looking for a NEW < FLUTE > } { ? did they } { = [< DID >] } { = did they [< TELL >] you }

(HKCSE, C021, 1088)

On the other hand, *so* is more common with shifts from specificity to generality and vice versa. In these examples, *so* seems to convey the sense 'it naturally / logically follows that', which is not apparent with *well*. In example 2, the lecturer uses *so* to introduce a specific illustration following the general explanation of an abstract idea:

(2)

b: …{ \ [oKAY] quite < ABstract > } { \ [< REALly >] } { = [< AND >] } { \ [WHAT] we < NEED > } { = is [< REALly >] } { \ [SOME] sort of < conVERter > } { \ [< CIRcuit >] } { \ [THAT] can DO the power < conVERsion > } { \ [< _ oKAY >] } { \ [< ***SO*** >] } { = for [< ^ eXAMple >] } { = i take [< ER >] } { \ the [TYPE] TWO } { = [< _ ER >] } { = [< POWer >] flow } { = [< PATH >] } { / as an [< eXAMple >] }…

(HKCSE, A022, 4605)

Although for these transitions there is a difference in the relative tendency of use of the two particles, *well* and *so* are generally interchangeable in the examples above. There are, however, some types of boundaries which appear to be compatible only with one particle but not the other. The use of *well* as an open quotation mark to signal the beginning of direct speech, for example, does not seem to be found with *so*:

(3)

a2: …{ \ [< RIGHT >] } { = [< ERM >] } { \ [THEN] you say ***WELL*** i have this < PROblem > } { \ it is [^ NOT] a big PROblem < YET > }…

(HKCSE, B125, 3858)

This association with speaker orientation shift also results in the frequent combination of reporting verbs with *well*, but not with *so*. Similarly, the introduction of answers to questions seems to be a functional sub-type permissible only with *well*, as in the following example:

(4)

b: …{ = [< ER >] } { = [WHY] am i so < CONfident > } { / [< ***WELL*** >] } { = [WE] are a < VEry > } { = [< imPORtant >] } { = [< LIVEly >] } { = [< ACtive >] } { \ [FInancial] < _ CENtre > } { = [Others] are not going to rePLACE < THIS > }…

(HKCSE, P062, 2059)

Another framing sub-type which is found only with occurrences of *well* is its use as a verbal comma to divide segments of talk. In the following example, the separation of main clause following if-clause is realised by *well*, whereas it remains doubtful if that could be done using *so*:

(5)

A: …{ \ i mean you can [< reLY >] on him } { \ if you [GIVE] him < SOMEthing > } { \ ***well*** [NEXT] < TIME > } { \ he [< SEES >] you that } { \ he'll give it you [< BACK >] and }…

(HKCSE, C113, 1720)

On the other hand, the conclusive use of *so* in a three-stage pattern which often involves a *because*-clause and a *so*-clause seems to be irreplaceable by *well*. At the beginning of example 6 below, the speaker gives a statement that the business in Hong Kong will not be affected by the power station. Following the provision of the reason in the *because*-clause, the main idea first proposed is reiterated in the *so*-clause. In cases like this, it is difficult to imagine replacing *so* with *well*:

(6)

a2: …{ = i [^ THINK] in < THE > } { = [NEAR] to medium < TERM > } { \ there will [NOT] have any < IMpact > } { = on [< THE >] er } { = [< ER >] } { = [< HONG >] kong } { \ [< BUSIness >] } { = [< ER >] } { = for [C] l p < POWer > } { = [< beCAUSE >] we } { = we [alREAdy] <

HAVE > er } { = [< ER >] } { = a [< DEdicated >] field } { = [< YACHting >] } { = from [< SOUTH >] china sea } { = [supPLYing] a NAtural < GAS > for } { \ for the [BLACK] point POWer < STAtion > } { = [< ***SO*** >] } { = [< THEY >] will not } { = have [< MUCH >] } { \ [< sigNIficance >] } { \ as [FAR] as we are < conCERNED > }

(HKCSE, B156, 4663)

In sum, although both *well* and *so* act as frames in discourse, they show preferences for marking different kinds of boundaries. For some framing sub-types, the two particles are generally compatible and differ only in terms of degree of association. For other sub-types, however, only the use of one particle seems possible but not the other. This demonstrates the complexity and diversity of transitions that need to be signalled in discourse organisation.

6.2.2 Link (linking)

When compared with the framing use, the linking function appears to be less varied. Both *well* and *so* introduce additional information to the preceding discourse, that is to say, the appending function of both particles in connecting text segments is similar. What distinguishes *so* from *well*, however, is the presence of the element of sequence. The sequential aspect of *so*, which indicates what follows is the next action in a series, is not obvious with instances of linking *well*. In example 7, speaker B is talking about the procedure involved in ordering books from publishers. What is signalled by *so* in this example is that credit is given to the bookstore before the speaker goes over there. It could therefore be loosely glossed as *then*. In marking this sequential linkage, *so* does not seem to allow a swap with *well*:

(7)

B: …{ = [< THEY >] don't } { / [SEND] us any < ^ BOOKS > } { = but what [HAPpens] is they just < GIVE > } { \ [CREdit] to the < ^ BOOKstore > } { = ***so*** [^ WE] go over and < GET > }

(HKCSE, C033, 1885)

Briefly, while *so* could serve both appending and sequential functions as a linking device, the use of *well* in introducing the next element in a series of events is absent in the data. This probably restricts *well* in connecting discourse units and accounts

for the relatively lower proportion of linking *well* when compared with *so* in the data examined.

Interpersonal function

6.2.3 Responsive signal (responsive)

As a responsive signal, one of the most noticeable differences between *well* and *so* is the association with dispreferred responses. Characteristically, *well* prefaces qualified answers, indicating some undesirable aspects such as a sense of insufficiency or incompleteness in the following discourse. This is described by Brazil (1997:118) as "a dummy *well*" in prevarication. In such cases, it seems unfeasible to substitute *well* with *so*:

(8)

a1: { \ but we are [< NOT >] } { ? we are not } { \ we are not [alLOWED] to use the < DAta > are we }

B: { = [< ER >] } { = ***well*** [< THAT'S >] not } { = [< THAT'S >] not } { \ that's not [< CERtain >] } { = i need to [< FOLlow >] that up }…

(HKCSE, B060, 2006)

Unlike *well*, the particle *so* does not appear to occur at the beginning of answers to questions. This applies not only to dispreferred answers but also to direct answers. Example 9 illustrates an initial *well* in a direct answer. The particle *so* does not seem to be a suitable alternative here:

(9)

b1: { = [< UM >] } { = and i [< THINK >] that } { \ that's been [WELL] < reCEIVED > } { = [< HOW >] } { = [< ^ HOW >] did you } { = [arRIVE] at the six and a < HALF > million } { \ DOLlar < ^ FIgure > }

b2: { \ ***well*** [WE] have < ^ STUdied > } { \/ similar [< PROgrammes >] elsewhere } { = [< THEY >] are } { \/ [ON] going < PROgrammes > } { = [< IN >] } { = [< CAnada >] } { = in [< ausTRAlia >] …

(HKCSE, P126, 979)

While *so* does not preface answers, its major responsive use is to introduce questions as motivated by the preceding talk. This constitutes approximately half of the instances of *so* as a responsive signal in the data. In these cases, *so* is generally

replaceable by *well*. Both are used to "introduce 'genuine' information questions" (Schourup 2001:1040, quotes in original). Although the use of *well* in such examples is possible, it seems to be much less common when compared with *so* in the data:

(10)

a3: { \ we [< ^ FIND >] that } { = [< ER >] } (.) { = [TEAcher] < TALK > } { \ [STILL] < DOminate > }

B: { \ ***well*** when you [< SAY >] } { = teacher [< TALK >] } { \ [DOminates] what do you < MEAN > }

(HKCSE, B081, 699)

In a similar fashion, *well* prefacing comments or evaluations arising from prior talk is relatively less frequent than *so*. In addition, motivated requests, which are often brought out by an initial *so* for the sake of politeness, do not seem to permit a switch with *well*. Example 11 shows a typical request in service encounters initiated by *so* which presents the face-threatening act of asking for the client's signature as being justified:

(11)

a: …{ \ [THIS] is the HUNdred < DOLlars > } { = from [TAX] < ^ reFUND > } (.) { = [\ < ***SO*** >] } { = would you [PLEASE] SIGN your name < ^ HERE > }

(HKCSE, B027, 19)

In terms of word combinations, the fact that dispreferred responses are typically introduced by *well* but not by *so* means that collocation with negation is only a feature of *well*. On the other hand, as a consequence of its affinity with questions, the particle *so* is much more frequently found in front of interrogative words.

While both *well* and *so* are very useful in "the process of interpersonal positioning and repositioning taking place within a text" (Aijmer 2005:87), the two particles differ in their orientations to alignment with earlier text segments. As reflected in the discussion above regarding their tendencies to preface different kinds of responses, it appears that *well* is more likely to be used when speakers intend to disalign themselves with a previously stated proposition while *so* is more associated with alignment with the preceding discourse. While *well* could act as a cushion to mitigate and soften a criticism or disagreement, *so* signals that the current speaker's

view is in line with the previous speaker's. Both particles thus play an important role in the maintenance of comity.

Interactional functions

6.2.4 Processing device (processing)

When compared with *so*, *well* seems to be a more versatile processing device as it is more commonly found in a wide range of processing problems. As demonstrated in Chapter 4, *well* frequently accompanies self repair phenomena including word recovery problems, self correction and rephrasing. In contrast, *so* is only at home with reformulation of the preceding discourse unit but rarely before word search and error replacements. The use of *so* in place of *well* in the example below for substituting the mistake just made, i.e. *twenty sixth*, would possibly sound less natural:

(12)

B: { = [SO] we'll do THAT and then < PRObably > } { = on the twenty [< SIXTH >] } { \ [< ***WELL*** >] } { = twenty [FIFTH] then we'll have a < BIG > } { = [< DINner >] }…

(HKCSE, C111, 676)

Instead, *so* is more likely to be used simply as a delay tactic for planning and time stalling purposes. In this respect, although both particles are possible candidates for gaining extra time, *so* is proportionally more common in the data:

(13)

a2: { = you [< ^ DON'T >] } { \ [HAVE] to HAVE < HEELS > } { = [< ^ THAT >] } { = are [< NOW >] } { \ [< IN >] } { ? is the } { = [< ***SO*** >] erm } { ? i mean } { ? i'm not } { = i'm [< NOT >] wearing } { \ [< HEELS >] }…

(HKCSE, B125, 13711)

That *so* is found in greater proportion as a staller when compared with *well* means that its combination with fillers and pauses is more frequent. Consequently, it may account for the higher percentage of level tone with *so* than *well*, as level tone usually occurs in the surroundings of a hesitation or when there is a mismatch between tone unit boundaries and points of potential completion (Brazil 1997).

6.2.5 Turn management (turn managing)

In the area of turn management, both *well* and *so* are involved in keeping and taking the conversational floor. Sacks et al. (1978:32) describe *well* and *so* as "turn-entry devices", highlighting their importance in turn taking. Because of their lack of semantic meaning, *well* and *so* initiate a turn without giving away the propositional content of the utterance. Therefore, they could be conveniently used as attempts to take the turn even if the speaker does not have a solid plan of what to say (ibid.). This also explains their use in turn holding activities. More importantly, using semantically empty *well* and *so* to interrupt also means that in case of overlapping the propositional meaning or the interpretability of the utterance that follows will not be affected (cf. Sacks et al. 1978). Hence the two particles are both useful linguistic items for seizing and holding turns.

In promoting turn alternation, one difference between *well* and *so* rests on their association with turn yielding. In the data examined, the particle *so* regularly occurs in utterance final position as a turn giving signal. This particular sub-function does not seem to be shared by *well*. In example 14, the speaker uses *so* at the end of an utterance to signal that the message left unstated could be inferred from information given earlier, i.e. the speaker promoted her subordinate to the position of vice president (VP). At the same time, it indicates the speaker's willingness to take a more passive role and encourages the addressee to take the floor. A possible explanation that *well* does not fit here is that the resultative aspect of *so* is absent in *well* and thus could not be used to suggest that something is only implied:

(14)

a: { = [< AND >] } { = i [< ^ proMOted >] her } { \ to be the [V] < P > } { = [< beCAUSE >] } { = her [< COUNterparts >] } { \ [ALL] got promoted last < ^ YEAR > } … { = [< AND >] she } { \ [HAsn't] got promoted < YET > } (.) { = [< ***SO*** >] }

(HKCSE, B113, 895)

In sum, while both *well* and *so* have crucial turn organisational uses, the two particles differ in the extent to which they control the conversational floor. *Well* appears to play a more assertive part in turn management and engages speakers in actively assuming the speaker role by holding and taking the turn. Conversely, *so* seems to

take a milder and less assertive position as it is involved not only in turn taking and holding but also yielding, giving it a slightly more neutral flavour in terms of turn management.

6.3 Quantitative contrastive analysis of *well* and *so*

Following the detailed qualitative comparison of the functions of *well* and *so*, this section contrasts their spread across different discourse functions. Table 6.2 shows a comparison of the functional distributions of the two particles in the HKCSE and the BNC (customised). Figures presented below are aggregate numbers from the two sources with percentages of each function out of the total number of the particle examined given[55]:

Table 6.2. The comparison of the frequency distributions of the discourse functions of *well* and *so* in the corpus data

Functional categories	**Textual**			**Interpersonal**		**Interactional**	
	Framing	Linking	Consequential	Responsive	Emotive	Processing	Turn managing
well	2088 (34.1%)	331 (5.4%)	N/A	2265 (37.0%)	203 (3.3%)	868 (14.2%)	373 (6.1%)
so	4096 (37.3%)	1283 (11.7%)	1146 (10.4%)	3184 (29.0%)	N/A	757 (6.9%)	508 (4.6%)

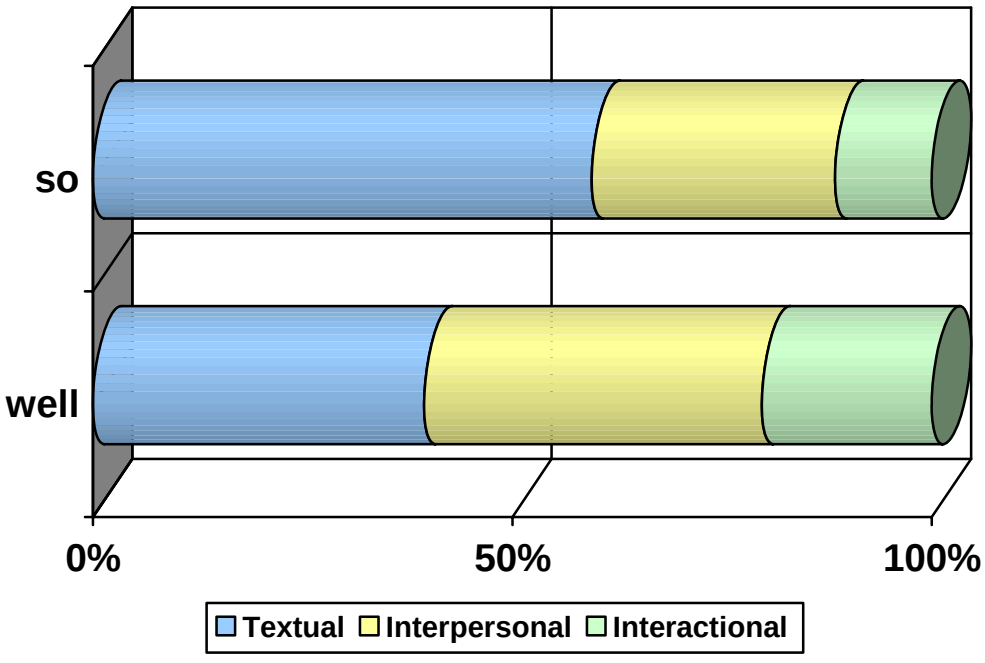
The above table indicates that while *well* and *so* both perform functions in five major categories, it can easily be seen that there are variations in terms of their frequency distribution. A few important observations could be made. First, in the textual domain, the linking function is considerably more common with *so*. More than one-tenth (11.7%) of all instances of *so* connect discourse units together while the proportion of *well* achieving this function is only about half of that (5.4%). A possible reason as discussed in the previous section is that *so* could signal the sequential dimension of discourse, which is not found in its counterpart *well*. Framing function is also slightly more frequent with *so* proportionally when compared with *well*, though the difference is not obvious. Moving onto the

[55] The total number of D-use *well* in Table 6.2 is 6,128, based on all the functionally classified instances in the HKCSE (n=1889, see Table 4.9) and the BNC (customised) (n=4239, see Table 4.21). The total number of D-use *so* in Table 6.2 is 10,974, based on all the functionally classified instances in the HKCSE (n=6574, see Table 5.8) and the BNC (customised) (n=4400, see Table 5.20).

interpersonal domain, responses are more likely to be associated with *well*. The proportions of *well* and *so* in marking responses are 37.0% and 29.0% respectively. This difference could be related to the fact that *well* could preface a variety of responses including dispreferred sequences, direct answers, follow-up questions and comments. Comparatively, *so* could only introduce a narrower range of responses. Another marked difference in the functional distribution of the two particles is found in the processing function. In the interactional domain, processing *well* constitutes approximately one-seventh (14.2%) of the total. This is double the proportion of *so* serving the same function (6.9%). Evidently, *well* is more frequently used as an editing device and a delay tactic when compared with *so*. This could be a result of its ability to signal a wider range of self-repair and hesitation phenomena as pointed out in the previous section. In turn management, the two particles have a similar ratio of occurrences. No considerable difference is found despite the fact that turn yielding is commonly associated with *so* but not *well*.

To further investigate the functional characteristics of *well* and *so*, the seven major functional categories are subsumed within three functional domains. Figure 6.1 illustrates the contrast between *well* and *so* across the textual, interpersonal and interactional functional domains:

Figure 6.1. The comparison of the frequency distributions of *well* and *so* across the three functional domains in the corpus data



A comparison of *well* and *so* in terms of their distributions in the three functional domains reflects the nature of the two particles. Figure 6.1 shows that *so* is primarily a textual device. As discussed earlier, not only is it more frequently found to achieve the textual functions of framing and linking, it is also used to express consequential relationship between adjacent text segments. The high proportion of *so* in textual functions results in more occurrences of this particle in utterance medial position as textual organisation is often more pressing in extended stretch of talk. It also accounts for the popularity of *so* in the academic domain since it contains the greatest number of unscripted and semi-planned long monologues among the four domains examined. Further, the overall aligning ambience of *so* and its higher tendency to accompany pauses probably give justification to the observation that *so* typically carries level tone when it is the tonic syllable. The fact that functions in the interactional domain only constitute a small fraction of the total number of *so* suggests that processing and turn managing uses are only auxiliary for the particle.

On the other hand, the particle *well* plays an important role both in the textual and the interpersonal functional domains, with a slight preference for the latter. In other words, *well* is relatively more active in the interpersonal domain when compared with *so*. The fact that *well* is more common in responses and expressing feelings as well as prefacing disparate topics and discourse stages explains its prevalence in utterance initial position. Because of the capacity of *well* for conveying emotions which is absent in *so*, *well* is more likely to occur as a stand-alone utterance and as a separate tone unit rather than as a pre-head when compared with *so*. Its closer affinity with interpersonal functions makes it a more indispensable item in conversations, where maintenance of social relationships and representation of personal attitudes are essential. The contrastive character of *well*, which has been widely documented in the literature especially when it is used as a marker of response, may give reasons for its stronger association with falling tone on the whole. This is in accordance with Brazil's (1997:80) remark on the social implications of proclaiming tone: "the speaker locates himself/herself outside the area of convergence, and foregrounds his/her own role as a non-assimilated participant in the interaction".

6.4 The general applicability of the functional framework arising from the present study

By conducting a qualitative and quantitative contrastive analysis of *well* and *so*, the present chapter provides a mapping of the functions of the two discourse particles. This mapping makes explicit the similarities and differences between *well* and *so* and elucidates what can be done by one particle but not the other which in turn deepens our understanding of the two particles. From the analysis above, it appears that the functional framework developed in the present study based on a large number of authentic examples of the two words offers a firm platform for individual studies and comparative analyses of discourse particles. The model proposed here, comprising three functional domains, seems to be sufficiently broad to encompass the various aspects of communication in the present investigation, which involves the study of many different contextual settings. At the same time, the functional categories account for what speakers intend to achieve in discourse and allow for meaningful comparison of functional sub-types if two particles share the same major function. The general applicability of the framework is thus supported by findings of two frequently occurring discourse particles in the English language. This suggests the potential of putting the proposed framework into practice with respect to other discourse particles and related items. More empirical studies and comparisons of discourse particles within this model may shed light on the complexity of functions of discourse particles and contribute towards a unified functional account of discourse particles.

Summary

A true understanding of the nature of discourse particles cannot be reached without careful examination from a variety of complementary perspectives. This necessitates the comparison between individual particles. By means of a contrastive study, this chapter demonstrates that although *well* and *so* overlap functionally, they are not freely interchangeable. On some occasions, they vary in their frequencies of association with different kinds of major functions and functional sub-types. In other cases, what could be done by one particle may not be possible by the other. The present chapter also shows that the data-driven functional typology established in this study seems to be generally applicable and could be extended to use in the description and interpretation of other discourse particles. This possibly leads us a

step closer to an overarching functional framework and a better understanding of discourse particles as a functional class.

Chapter 7

Conclusions

Abstract

This chapter, which consists of five major parts, concludes the thesis. It first gives a brief summary of all the previous chapters (7.1). The research questions of the study are then revisited and implications of the present study for particle research discussed (7.2). This is followed by some pedagogical suggestions for teaching materials (7.3). Finally, limitations of the study are described (7.4) with suggested areas for further research (7.5).

7.1 Summing up

At the beginning of Chapter 1, the genesis of the present study is provided. The rationale for the particular focus of the study is then given, which places extra emphasis on the importance of discourse particles, the value of intonation in particle studies, the prevalence of intercultural communication and the relation between discourse particles and pedagogy. This leads to the purpose of the study, i.e. to explore the use of discourse particles in real-life spoken intercultural interactions and in textbook descriptions, taking into account various linguistic, sociolinguistic and contextual factors in the interpretation and explanation of the findings of the study.

Following the introductory chapter, Chapter 2 discusses in detail previous studies of discourse particles, focusing on particular aspects which are especially relevant to the present research. It begins by addressing some controversial notions in the field, including the terminology, definition and classification of discourse particles. From the debate, it is shown that no consensus has been reached over the above issues. It is thus argued that more research needs to be done on individual linguistic items which could potentially be considered discourse particles before conclusive statements could be made about them as a class. Some of the most common properties of these items reported in the literature are described and evaluated based on their importance as defining features of discourse particles. How discourse particles have been analysed from different approaches is critically examined, which in turn suggests the inadequacies of some existing models. This is followed by a review of the discourse functions of particles, which shows that discourse particles are multi-faceted items

working in the textual and interpersonal meta-functional domains. The small number of studies concerning the prosodic profiles of discourse particles and their relation to language learning is then noted to highlight the significance of these areas, which have largely been overlooked thus far. Through an investigation of the descriptions of discourse particles in reference books, it is found that there has been a growing focus on these linguistic items over the evolution of grammar books and dictionaries. Discourse particles are no longer treated as peripheral elements in the language which deserve no mention in tool books. It is thus an appropriate time to look into the pedagogical aspect of these items. Although the in-depth literature review reports diverse and fruitful research outputs in the study of discourse particles, there are still gaps which remain to be filled. The present study, which examines factors such as prosodic patterns, text types and language background of speakers, is an attempt to deal with some of the neglected issues in the field and to contribute to the better understanding of discourse particles.

Chapter 3 provides a thorough description of the datasets and outlines the research process and analytical tools employed in the present study. The HKCSE was chosen for its intercultural nature, its prosodic mark-up and its carefully-annotated sociolinguistic and contextual variables. A customised corpus was compiled from the BNC spoken section as a reference corpus for comparing and cross-checking purposes. It serves as an independent data source to verify the reliability of findings generated from the HKCSE. To investigate the similarities and differences between authentic usage and textbook descriptions, a textbook database was created from a selection of upper-secondary English textbooks in Hong Kong. Throughout the course of the research process, a number of computer tools were used to facilitate the study, though a number of analytical issues required human interpretation.

The two subsequent chapters investigate the use of *well* and *so*, two of the most frequently occurring discourse particles in the English language. For each discourse particle, a number of linguistic, sociolinguistic and contextual factors are considered in the interpretation and explanation, including the collocational pattern, the syntactic position, the prosodic features, the language background and gender of the speaker as well as the domain and text type in which the particle is found. For both *well* and *so*, it is found that the syntactic positions and prosodic profiles are distinct for their discourse and grammatical functions. In other words, the syntactic looseness and prosodic detachedness are only characteristic of the two words when they are

discourse particles but not when they convey propositional meaning. This gives strong support for the syntactic and prosodic independence of discourse particles. For each particle, various discourse functions have been identified in the datasets, operating in three functional domains, namely textual, interpersonal and interactional, which are derived from thousands of occurrences examined. Although the linguistic and contextual features are found to give useful clues to the functional interpretation, no obvious one-to-one correlation is found between these features and discourse functions. Findings from the reference corpus generally substantiate conclusions drawn from the intercultural corpus regarding the use of the two particles. On the other hand, comparisons with the textbook database show huge discrepancies in the use of both particles between authentic examples and pedagogical descriptions. Specifically, the textbooks under investigation fail to give a comprehensive account of how discourse particles are used. Detailed descriptions are lacking and examples containing discourse particles are often short and decontextualised.

Chapter 4 presents a detailed examination of the use of *well* in the three datasets. From the analysis it is found that *well* serves six major pragmatic functions in the corpus data, with its role as a responsive signal being the most dominant. Two important sociolinguistic variations are observed. In terms of the first language background of speakers, it is shown that the discourse use of *well* is a more common feature in the speech of native speakers of English when compared with Hong Kong Chinese. Further, Hong Kong Chinese are more likely to use *well* as a processing device. As far as the gender of speakers is concerned, *well* is found to be more popular with men and is more likely to be used as a processing device by women. The fact that *well* is typically responsive leads to its high frequency in utterance initial position and in conversational talk.

Chapter 5 looks at another highly-versatile discourse particle *so*. Close examination of the data reveals that *so* is mainly used as a frame, though it also achieves five other pragmatic functions in the textual, interpersonal and interactional domains. Unlike *well*, no significant differences are found in relation to sociolinguistic factors. Hong Kong Chinese and native speakers of English use the particle *so* in similar ways with almost equal frequencies. The same applies to male and female speakers. The observation that *so* plays an important role in discourse structuring and organisation may account for its pervasiveness in the academic domain, which contains a large number of unscripted semi-planned texts.

In Chapter 6, a contrastive functional analysis is carried out to further examine the similarities and differences between the two discourse particles. The qualitative and quantitative comparison serves to further elucidate what *well* and *so* have in common as well as how they differ. This in turn promotes a better understanding of the characteristics of each individual particle. The contrast made also points to the possibility of applying the functional framework established from the present study to other potential members in the class for meaningful comparison between different particles.

7.2 Research questions revisited: Implications of the present study

This section revisits the research questions of the study and suggests how the present work contributes to the study of various aspects of discourse particles.

7.2.1 Are English discourse particles largely without meaning?

At the beginning of the thesis, the following question was raised: are English discourse particles largely without meaning? As the study draws to a close, it is timely to return to this issue, which is central to the present investigation. Along with evidence from numerous scholarly outputs in this area, the present study argues that, pragmatically, discourse particles are by no means largely without meaning. By conducting a large-scale analysis of two of the most frequently occurring discourse particles in the English spoken language, the present study demonstrates that discourse particles achieve a variety of discourse functions in many different settings. Based on the findings of *well* and *so* in this study, discourse particles thus could be defined as syntactically optional elements which have no or little propositional value but are rich in pragmatic meanings which vary according to the linguistic and situational context in which they occur. As noted by Wierzbicka (1976:327), discourse particles are abbreviatory devices “for expressing complex pragmatic meanings at minimal cost”. These short linguistic items convey highly condensed meanings which very much depend on the context. In Sinclair’s (2007a, 2007b) terms, they probably have a higher “meaning shifting” potential when compared with many other words in the language. As shown by the large number of authentic examples in this study, discourse particles are indispensable in regulating the flow of discourse, in signalling personal attitudes and feelings as well as in managing the interaction. Their high frequency in naturally-occurring texts convincingly proves

that they are a very popular group of linguistic items commonly employed in a wide range of contexts. The findings of *well* and *so* as presented in Chapters 4 and 5 have provided strong support for their importance in spoken discourse. The fact that *well* and *so* are not freely interchangeable, as illustrated in Chapter 6, has further demonstrated that discourse particles are not superfluous and thus cannot be casually replaced. Suffice it to say here that discourse particles are in no way pragmatically meaningless and they are "not simply a verbal 'crutch' indicating a lack of speaking proficiency" (He and Lindsey 1998:134, quotes in original), contrary to what has been suggested in the introductory chapter by the Hong Kong examination body. There are thus no reasons to dissuade students, or anyone for that matter, from using them without clearly elucidating to them by means of authentic examples how discourse particles are really used.

7.2.2 Do discourse particles have distinctive prosodic profiles and is there a correlation between their prosodic features and discourse functions?

By systematically examining the prosodic patterns of *well* and *so* in thousands of examples from corpus data, the present study concludes that the discourse use and the propositional use of the same word form have distinctive prosodic profiles. Specifically, the two discourse particles investigated in the present study enjoy a high degree of prosodic autonomy either as a pre-head or as a separate tone unit, which is unmatched by their grammatical counterparts (see Tables 4.3 and 5.3). As commented by Altenberg (1987:156) on the relation between what he calls discourse items such as *well* and *now* and tone unit (TU) position, "they often form TUs of their own in the text, and thus receive nuclear prominence as a matter of course". In other words, being one-word tone unit is typical of discourse particles. In addition, the frequent co-occurrence of the discourse uses of the word form with pauses but not the propositional uses is an indicator that only as discourse particles are *well* and *so* loosely attached to the syntactic structure of utterance (see Tables 4.7 and 5.7). Therefore, the present study demonstrates with a large quantity of authentic examples that prosodic detachedness could be used as a criterion for distinguishing discourse uses from propositional uses. This could be especially useful for identifying discourse particles which are not easily distinguishable from their grammatical counterparts. Although a systematic correlation is not found between prosodic features and all the discourse functions identified, some slight associations

and prosodic preferences are observed for certain functions, such as the stronger inclination of *well* to be the tonic syllable in a shared tone unit in marking direct speech (see Table 4.11) and the higher proportion of *so* as a separate tone unit in its turn managing function (see Table 5.9). As for the choice of tone, the findings that *well* and *so* typically carry fall and level tones respectively (see Tables 4.5 and 5.5) provide valuable empirical data for further investigating whether a specific set of discourse particles sharing the same tone are functionally similar. The present study thus shows that intonation is an important aspect in the study of discourse particles, not only in assisting in the identification of whether a certain occurrence is used as a discourse particle but also in the interpretation of their discourse functions.

7.2.3 How do the linguistic features of discourse particles, the sociolinguistic background of speakers and the context of communicative events influence the use of discourse particles in authentic speech?

Apart from showing that discourse particles are far more than mere empty fillers in a broad selection of spoken text types, the present study demonstrates the significance of taking into account the lexical associations and syntactic positions of discourse particles as well as the domain and text type of communicative events in the analysis of discourse particles. Through the examination of a considerable number of authentic examples displaying varied lexical and positional patterns from a wide range of text types in the academic, business, conversational and public domains, the study shows that different discourse functions of *well* and *so* tend to have different word combinations and positional preferences (see Sections 4.3.5.1-2 and 5.3.5.1-2). The situational context in which a discourse particle is found not only serves as an important functional clue, it also affects the rate at which the particle occurs (see Sections 4.3.6.3 and 5.3.6.3). As observed by Aijmer (2002), these linguistic and contextual features are indispensable in illuminating what discourse particles do in talk. In addition, the present study illustrates the similarities and differences in the use of discourse particles between speakers of dissimilar sociolinguistic backgrounds in ever-growing intercultural encounters. The fact that both the first language and gender of speakers have varying effects on the use of the two discourse particles examined in the present study shows the importance of individual analyses before collective judgment. It is crucial to discriminate between factors affecting the use of discourse particles as a collective functional group and factors which might only be

influential on individual members in the class (see, for example, Andersen 2001; Macaulay 2005; Stubbe and Holmes 1995, on social variation in the use of individual discourse particles). As demonstrated by the present study, not all discourse particles are susceptible to a given sociolinguistic variable to the same extent. While both the discourse rate and functions of *well* are subject to the first language and gender of speakers in the present study (see Sections 4.3.6.1-2), the use of *so* does not seem to be much affected by these two factors (see Sections 5.3.6.1-2). This exemplifies the significance of examining each lexical item which has the potential to be considered a discourse particle carefully before drawing any conclusions about discourse particles in general. The comparison of the use of discourse particles in the HKCSE with that in the BNC also hints at cultural differences between Hong Kong Chinese and British speakers of English in areas such as how dispreferred responses and turn taking are handled (see Sections 4.4 and 5.4).

7.2.4 Do textbook descriptions and presentations of discourse particles conform to how they are used as observed in corpus data?

In relation to pedagogy, a survey of the English textbooks collected for the present study shows two characteristic features of these teaching materials: that most of them take a communicative approach instead of a grammar-based approach and that they are highly task-oriented. Findings from the comparison between textbooks and authentic speech reveal considerable discrepancies between the two sources in terms of the frequencies and positions of the two discourse particles *well* and *so* and how they are used (see Sections 4.5 and 5.5). Textbook writers of the materials inspected in this study seem to have forgotten that apart from knowing how to transact in a target language, learners “at all levels also need to build relationships, express attitudes and affect, evaluate and comment, and make the propositional content of a message more person-oriented” (Carter 1998:50), all of which could be assisted by the use of discourse particles. Despite their pervasiveness in spoken discourse and the crucial textual, interpersonal and interactional functions they serve, discourse particles seem to play a minimal role in the textbooks examined. They are not depicted negatively as dysfluency signals, but there are only negligible descriptions associated with them. Their occurrences are mostly scattered across examples illustrating various speech functions in a rather haphazard manner, without any explicit explanations of why they are there and how they should be used. This is

alarming given the pragmatic significance of discourse particles and the lack of exposure of many foreign language learners to these linguistic items. In the worst case scenario, discourse particles may well be a linguistic resource that a language learner fails to master, even when they have already achieved a relatively high level of grammatical competence. As a result, language learners who have poor understanding of how discourse particles are used may run the risk of having their pragmatic competence seriously jeopardised.

Given the fact that English is acquired mostly through formal education in the local setting as well as in many EFL contexts and discourse particles are crucial in achieving pragmatic competence, there are no reasons why they should not be taught in classrooms. The present study shows that the descriptions and examples of discourse particles in the textbooks examined are far from satisfactory. Substantial revisions with the incorporation of authentic examples are required in order to present a more comprehensive picture to students concerning how discourse particles are used (see, for example, Carter et al. 2000, for the teaching of discourse particles using authentic examples, and McCarthy, McCarten & Sandiford 2005a; 2005b, for a corpus-informed syllabus of conversational strategies). Findings from the present study, which are generated from a large number of authentic examples in a wide range of settings, provide an empirical basis for the improvement of the descriptions of discourse particles in textbooks as well as in grammars and dictionaries. With the increasing number of research outputs showing the fruitful results of data-driven learning (DDL) (see, for example, Cheng, Warren & Xu 2003; Tribble 1997; Tribble and Jones 1990), it is high time to apply the methods of corpus linguistics to language teaching and learning, especially in areas such as the acquisition of discourse particles, where invented decontextualised examples could hardly elucidate to students the many discourse functions of these linguistic items in a variety of settings. In this respect, a corpus-based and data-driven approach to learning discourse particles with the use of concordancers may be useful (see, for example, Möllering 2004, on the teaching of German modal particles, and Zorzi 2001, on the learning of Italian discourse markers) and could probably provide the right resources for learners to explore and research in a more learner-centred way how discourse particles are actually used.

7.3 Teaching discourse particles: Pedagogical suggestions

In the light of the pedagogical implications arising from the present study, this section suggests two activities which could be incorporated into teaching materials to supplement and facilitate the learning of discourse particles. The first example, which aims at differentiating discourse particles from their grammatical counterparts, is probably more useful as an 'awareness-raising' activity for intermediate non-native learners. The second example, which deals with a specific discourse function of the particle *well*, is possibly more suitable for advanced learners of English and students with some linguistic knowledge[56].

Activity 1: Distinguishing discourse functions from grammatical functions

Study the following concordance lines which are examples of *well* serving discourse functions:

1 you let you let it goes everywhere A: **well** no he er he actually just live in the ca
2 e Hong Kong as a world city of Asia **well** let me tell you er some of the problem
3 nd you to this year's a1: will they B: **well** I don't know a1: ((laugh)) are you inte
4 ged now b: they are now changed B: **well** I don't know but it's (.) all the traffic li
5 don't (.) decorate your own place B: **well** I won't decorate won't decorate any (.)
6 to practise my Cantonese I I mean b: **well** but the thing is like you couldn't really
7 don't have very much time do we b: **well** I don't know how long the recording
8 all that's they want us to talk about b: **well** I guess I guess they don't want to put
9 e a1: mm you also think so (.) now er **well** at this point er we we will not go into
10 is press briefing so late in the evening **well** without further ado you know all the

Study the following concordance lines which are examples of *well* serving grammatical functions:

1 because of course in manufacturing as **well** as any other types of industry we are
2 item name volume preparation time as **well** as the list and amount of ingredients
3 o okay e- even though you did not do **well** during the workshop training you on
4 they the time within certain time how **well** it's going to be cooked (.) okay so by
5 u going to motivate people to perform **well** on their job (.) if (.) if and only if thi
6 ich is what (.) which is if you perform **well** you would (.) you would (.) you wor
7 need to go through them they're pretty **well** listed here perhaps the most expensi
8 ms work quite well it seems work quite **well** it seems work quite well but (.) er i
9 ard at the end of the day you will score **well** in this subject (.) alright this is the n
10 n (.) does not work well does not work **well** okay then you maybe (.) have more

[56] For illustrative purposes, the concordance lines in the activities display below only has a small word span. When these activities are carried out, preferably through a concordancing program on the computer, more concordance lines with a larger stretch of co-text may be needed to present more contextual information to students for interpretation.

Now discuss the differences between the discourse and grammatical functions of *well* in terms of:

a) how their omissions affect the structure of the utterance

b) their positional preference in the utterance

c) the words which are in their immediate surroundings (e.g. the word classes they belong to)

Based on the above information, decide whether the following instances of *well* achieve discourse functions or grammatical functions:

1)

b1: actually you have a consultant within your family ((laugh))

B: oh that's it yea oh yea yea (.) **well** you might have to marry the culture as **well** you see so (pause) I've never worked anywhere else really

2)

B: okay not for a while

a: yea er

B: **well** it's probably I I don't know Yuen Long very **well** but when you go out that way [(.) the air is cleaner

<u>Activity 2: Investigating the discourse function of *well* in dispreferred responses</u>

Study the following concordance lines which are examples of *well* serving a specific discourse function:

1 you let you let it goes everywhere A: **well** no he er he actually just live in the ca
2 flection of what actually going on a: **well** erm I don't think so I think those woul
3 nd you to this year's a1: will they B: **well** I don't know a1: ((laugh)) are you inte
4 ged now b: they are now changed B: **well** I don't know but it's (.) all the traffic li
5 don't (.) decorate your own place B: **well** I won't decorate won't decorate any (.)
6 to practise my Cantonese I I mean b: **well** but the thing is like you couldn't really
7 don't have very much time do we b: **well** I don't know how long the recording
8 all that's they want us to talk about b: **well** I guess I guess they don't want to put
9 ll affect one country two systems b3: **well** I doubt that er there're two two levels
10 d market er b: Selina is that true a: er **well** not quite in fact the er one thing is tr

Now answer the following questions:

a) Where do these instances of *well* usually occur in the utterance?

b) What are the words which are frequently found in the immediately surroundings of *well* in these instances?

c) What do these responses associated with *well* above have in common?

d) What are the possible effects on these responses if *well* is omitted?

Based on the above information, describe the properties of *well* serving this specific discourse function.

7.4 Limitations of the present study

This is one of the first, if not the only, large-scale attempts to systematically compare the use of discourse particles in two different corpora and in a textbook database. Given the scope of the present study, it was only possible to analyse two discourse particles in detail. Admittedly, the inclusion of more discourse particles would provide us with a more comprehensive account of how they overlap and differ. It does not mean, however, that the findings from the present study have limited implications only applicable to the two particles under investigation. On the contrary, the functional framework proposed in the study has the potential of accommodating the analyses of other discourse particles, thus allowing for contrasts to be made between different items in the class. This suggests that the model has wider currency outside the scope of the present study.

In terms of the data examined, it could be argued that the two corpora used are relatively small in size when compared with the text collections currently available on the market. Therefore questions might arise as regards how representative they are. However, as remarked by McEnery and Wilson (1996), a spoken corpus is only a glimpse of a given language variety at a restricted period in time. No matter how large the corpus is, it is not possible to represent the language as a whole. Despite the rather modest size of the datasets, they house a substantial number of naturally-occurring instances of *well* and *so* in a wide range of text types, making it possible to carry out an empirical study of discourse particles on a scale larger than customary.

7.5 Areas for further research

Over the past few decades the study of discourse particles has been expanding. Yet our understanding of these linguistic items in many respects is still inadequate. Issues which are pending to be resolved include cross-linguistic features of discourse particles in different languages, variation in discourse particles and synchronic and diachronic evolution of discourse particles. Do all languages share a set of discourse

particles with common basic features? Do variant forms of discourse particles function in the same way? What are the similarities and differences in the use of discourse particles between different varieties of the same language, for example, between world Englishes? What is the historical development of discourse particles? What is the relationship between other discourse particles and their grammatical counterparts? These are some of the interesting and important topics which have not yet been thoroughly addressed. The social and stylistic distribution of discourse particles, an area which is often neglected in the past, is attracting more and more attention (see, for example, Fuller 2003; Kyratzis and Ervin-Tripp 1999; Redeker 1990; Stubbe and Holmes 1995), though only a few large-scale studies have been attempted thus far (see, for example, Andersen 2001; Macaulay 2005; Müller 2005). With the aid of computer programs to generate collocational patterns, the largely ignored field of combination of discourse particles and other linguistic items is finally being systematically exploited (see Aijmer 2002). Further insights into the role of discourse particles in English language teaching and learning could also be gained by examining whether and how they are taught in language classrooms. With the rising recognition of discourse particles as a crucial pragmatic resource, we can be hopeful that many of these questions will be answered, leading us towards a more comprehensive understanding of discourse particles before long.

Appendix 1

Index of the HKCSE

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
A001	Academic	lecture	0:51:16	7,175	ENGL lecture
A002	Academic	lecture	0:50:00	4,593	CLU lecture
A003	Academic	lecture	0:49:04	6,344	ITC lecture
A004	Academic	lecture	0:50:48	5,135	HTM lecture
A005	Academic	lecture	1:26:00	10,879	MFG lecture
A006	Academic	lecture	0:46:00	6,091	MFG lecture
A008	Academic	lecture	1:30:00	8,010	MFG lecture
A009	Academic	lecture	1:30:00	7,579	MFG lecture
A010	Academic	lecture	1:13:00	6,509	MFG lecture
A013	Academic	seminar	0:47:02	6,656	ENGL seminar
A014	Academic	student presentation and Q&A	1:21:25	10,214	HTM seminar (student presentation and Q&A)
A015	Academic	student presentation and Q&A	1:20:26	10,103	HTM seminar (student presentation and Q&A)
A016	Academic	student presentation and Q&A	1:29:32	10,497	HTM seminar (student presentation and Q&A)
A017	Academic	student presentation and Q&A	1:32:56	11,922	ENGL seminar (student presentation and Q&A)
A018	Academic	student presentation and Q&A	2:01:53	15,625	BSE seminar (student presentation and Q&A)
A019	Academic	student presentation and Q&A	1:30:00	15,081	ENGL seminar (student presentation and Q&A)
A022	Academic	seminar	0:46:31	6,886	EIE seminar (teacher talk and Q&A)
A025	Academic	student presentation and Q&A	0:43:29	4,470	ABCT seminar (teacher talk and Q&A)
A028	Academic	seminar	1:51:00	11,677	Open University seminar

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
A029	Academic	student presentation and Q&A	1:49:06	13,165	ENGL BACEL Language and Society student presentation and Q&A
A036a	Academic	supervision	0:35:00	3,365	ENGL WAP supervision
A036b	Academic	supervision	0:13:00	1,319	ENGL WAP supervision
A036c	Academic	supervision	0:43:00	4,728	ENGL WAP supervision
A039	Academic	supervision	0:30:00	3,852	ENGL WAP supervision
A041	Academic	supervision	0:32:00	4,544	ENGL WAP supervision
A044	Academic	tutorial	1:26:00	7,283	ITC tutorial
A045	Academic	tutorial	0:05:28	984	ENGL tutorial
A047	Academic	workshop for staff	0:30:00	3,394	ENGL EETRC workshop
A050	Academic	seminar	0:45:00	5,124	VELA seminar discussion
B001	Business	service encounter	0:03:10	367	hotel (check out)
B002	Business	service encounter	0:03:10	149	hotel (service encounter)
B003	Business	service encounter	0:14:00	2,193	hotel (hotel ambassador)
B004	Business	service encounter	0:02:58	126	hotel (check out)
B005	Business	service encounter	0:00:38	144	hotel (service encounter)
B006	Business	service encounter	0:02:00	96	hotel (check out)
B007A	Business	service encounter	0:02:51	186	hotel (check out)
B007B	Business	service encounter	0:00:20	57	hotel (check out)
B008	Business	service encounter	0:02:40	130	hotel (check out)
B009	Business	service encounter	0:07:07	426	hotel (service encounter)
B010	Business	service encounter	0:03:24	309	hotel (service encounter)
B011	Business	service encounter	0:08:25	590	hotel (service encounter)
B012	Business	service encounter	0:00:30	31	hotel (check out)

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
B013	Business	service encounter	0:00:15	19	hotel (service encounter)
B014	Business	service encounter	0:01:28	77	hotel (service encounter)
B015	Business	service encounter	0:03:05	295	hotel (service encounter)
B016	Business	meeting	0:18:00	2,828	hotel (meeting)
B017	Business	meeting	0:40:13	6,188	hotel (meeting)
B018	Business	meeting	0:22:50	2,641	hotel (meeting)
B019	Business	meeting	0:10:00	1,435	hotel (meeting)
B020	Business	service encounter	0:01:39	160	hotel (service encounter)
B021A	Business	service encounter	0:02:10	486	hotel (service encounter)
B021B	Business	service encounter	0:02:00	405	hotel (service encounter)
B023	Business	meeting	0:14:30	2,341	hotel (meeting)
B024	Business	service encounter	0:03:07	453	airport (service encounter)
B025	Business	service encounter	0:00:21	68	airport (service encounter)
B026	Business	service encounter	0:01:21	143	airport (service encounter)
B027	Business	service encounter	0:00:35	101	airport (service encounter)
B028	Business	service encounter	0:00:30	31	airport (service encounter)
B029	Business	service encounter	0:04:38	573	airport (service encounter)
B030	Business	service encounter	0:02:15	440	airport (service encounter)
B031	Business	service encounter	0:00:05	24	airport (service encounter)
B032	Business	service encounter	0:08:32	669	airport (service encounter)
B033	Business	service encounter	0:04:23	482	airport (service encounter)
B034	Business	service encounter	0:01:00	136	airport (service encounter)
B035	Business	service encounter	0:05:00	748	airport (service encounter)
B036	Business	service encounter	0:01:00	137	airport (service encounter)
B037	Business	service encounter	0:05:25	446	airport (service encounter)
B038	Business	service encounter	0:03:46	201	airport (service encounter)

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
B039	Business	service encounter	0:00:30	60	airport (service encounter)
B040	Business	service encounter	0:00:10	47	airport (service encounter)
B041	Business	service encounter	0:02:00	422	airport (service encounter)
B042	Business	service encounter	0:03:15	162	airport (service encounter)
B043	Business	service encounter	0:02:03	274	airport (service encounter)
B044	Business	service encounter	0:02:13	162	airport (service encounter)
B045	Business	service encounter	0:00:40	79	airport (service encounter)
B046	Business	service encounter	0:00:10	31	airport (service encounter)
B047	Business	service encounter	0:00:33	56	airport (service encounter)
B048	Business	service encounter	0:01:55	142	airport (service encounter)
B049	Business	service encounter	0:04:27	573	airport (service encounter)
B050	Business	service encounter	0:01:45	128	airport (service encounter)
B051	Business	service encounter	0:01:42	288	airport (service encounter)
B052	Business	service encounter	0:01:25	216	airport (service encounter)
B053	Business	service encounter	0:00:53	109	airport (service encounter)
B054	Business	service encounter	0:01:04	140	airport (service encounter)
B055A	Business	service encounter	0:00:35	61	airport (service encounter)
B055B	Business	service encounter	0:05:10	609	airport (service encounter)
B056	Business	meeting	0:08:28	2,822	meeting
B058	Business	meeting	0:18:36	2,341	meeting
B059	Business	meeting	0:38:06	7,686	meeting
B060	Business	meeting	0:45:56	7,990	meeting
B061	Business	interview	0:45:00	7,207	placement interview
B062	Business	interview	0:18:00	3,464	placement interview
B063	Business	interview	0:35:00	6,594	placement interview
B064	Business	interview	0:09:00	1,706	placement

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					interview
B065	Business	interview	0:14:00	2,229	placement interview
B066	Business	interview	0:35:00	6,350	placement interview
B067	Business	interview	0:09:00	920	placement interview
B068	Business	interview	0:12:00	2,095	placement interview
B069	Business	interview	0:08:00	1,446	placement interview
B070(1)	Business	interview	0:15:29	2,557	placement interview
B070(2)	Business	interview	0:15:23	2,552	placement interview
B071	Business	interview	0:16:19	2,889	job interview
B072	Business	interview	0:25:26	4,584	job interview
B073	Business	workplace telephone talk	0:03:43	480	telephone conversation
B074	Business	workplace telephone talk	0:02:13	427	telephone conversation
B075	Business	informal office talk	0:30:40	6,336	informal office talk
B076	Business	interview	0:17:11	3,158	RA job interview
B077	Business	interview	0:10:41	1,446	RA job interview
B078	Business	interview	0:16:43	2,500	RA job interview
B079	Business	interview	0:10:30	1,496	RA job interview
B080	Business	interview	0:15:50	2,438	RA job interview
B081	Business	interview	0:21:22	4,714	RA job interview
B082	Business	interview	0:23:00	3,962	RA job interview
B083	Business	interview	0:17:51	3,099	RA job interview
B084	Business	interview	0:13:42	2,717	RA job interview
B085	Business	interview	0:20:50	3,434	RA job interview
B088	Business	presentation	0:05:20	711	Business After SARS Conference
B089	Business	presentation	0:09:05	1,138	Business After

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					SARS Conference
B090	Business	presentation	0:08:54	1,316	Business After SARS Conference
B094	Business	presentation and Q&A	0:37:40	7,889	Hong Kong General Chamber of Commerce Luncheon
B099	Business	presentation	0:22:10	3,644	HKGCC Pearl River Delta Roundtable Series
B104	Business	presentation	0:21:46	2,947	HKGCC Seminar
B106	Business	presentation	0:17:38	2,896	HKGCC Seminar
B108	Business	announcement and Q&A	0:57:30	9,075	Hang Seng Bank 2002 Annual Results Announcement
B109	Business	interview	0:10:00	1,257	workplace interview
B110	Business	presentation	0:12:40	1,663	China Light & Power
B111	Business	conference call/video conferencing	0:23:19	4,002	conference call with colleagues
B112	Business	workplace telephone talk	0:03:18	609	workplace telephone discussion
B113	Business	informal office talk	0:19:15	3,580	workplace talk at lunch
B114	Business	interview	0:34:46	5,629	RA job interview
B121	Business	presentation and Q&A	1:32:00	12,386	Speech: "The politics of English language provision"
B123	Business	presentation and Q&A	1:48:06	15,235	Speech: "Training professionals in intercultural awareness"
B125	Business	presentation and Q&A	1:49:30	15,708	Speech: "Professional

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					image for men and women"
B146	Business	informal office talk	0:56:28	9,250	informal office talk in a restaurant
B147	Business	conference call/video conferencing	0:10:22	2,015	video meeting
B148	Business	informal office talk	0:44:36	8,172	informal office talk in a restaurant
B150	Business	presentation	0:13:49	1,638	Speech: HKTDC
B151	Business	presentation	0:02:22	308	Announcement: HKTDC
B154	Business	presentation	0:29:11	3,859	Speech: HKTDC
B155	Business	announcement and Q&A	0:58:10	7,908	CLP 2001 Annual Results Announcement
B156	Business	announcement and Q&A	0:32:48	5,120	CLP 2002 Interim Results Announcement
C001	Conversational	conversation	1:14:54	4,158	Dinner party at a Chinese restaurant
C002	Conversational	conversation	1:14:50	11,207	home
C003	Conversational	conversation	0:03:59	390	Chinese restaurant
C004	Conversational	conversation	1:14:59	9,609	Chinese restaurant
C005	Conversational	conversation	0:23:12	4,737	home
C006	Conversational	conversation	0:05:07	749	Record 001 restaurant
C007	Conversational	conversation	0:17:57	4,445	Record 002 lounge
C008	Conversational	conversation	0:05:24	778	Record 003 fast food shop
C009	Conversational	conversation	0:24:55	3,844	Record 004 lounge
C010	Conversational	conversation	0:08:07	1,237	Record 005 staff club
C011	Conversational	conversation	0:39:34	6,493	Record 006 lounge
C012	Conversational	conversation	0:31:15	5,536	Record 007 university
C013	Conversational	conversation	1:03:13	11,305	Record 008

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					Oliver's
C014	Conversational	conversation	0:06:06	514	Record 009 the Peak
C015	Conversational	conversation	0:35:37	5,531	Record 010 university staff Canteen
C016	Conversational	conversation	0:11:15	1,830	0020 YMCA café
C017	Conversational	conversation	0:05:12	616	0021 food court/plaza
C018	Conversational	conversation	0:04:00	821	0022 restaurant
C019	Conversational	conversation	0:06:05	1,320	0023 Delifrance
C020	Conversational	conversation	0:19:11	4,706	0024 food Court
C021	Conversational	conversation	0:12:53	1,509	0025 airport
C022	Conversational	conversation	0:14:36	2,181	0027 food plaza
C023	Conversational	conversation	0:17:52	2,698	0028 restaurant
C024	Conversational	conversation	0:35:36	5,223	restaurant
C025	Conversational	conversation	0:18:38	3,541	office
C026	Conversational	conversation	0:15:38	2,373	restaurant
C027	Conversational	conversation	0:38:59	5,590	YMCA (TST)
C028	Conversational	conversation	0:14:50	2,344	Hayman restaurant
C029	Conversational	conversation	0:03:09	677	
C030	Conversational	conversation	0:07:20	1,198	DISC 29 café
C031	Conversational	conversation	0:06:36	1,102	TST Delifrance
C032	Conversational	conversation	0:26:04	3,027	restaurant
C033	Conversational	conversation	0:28:01	3,910	MD30 university staff canteen
C034	Conversational	conversation	0:04:06	841	Seibu food court
C035	Conversational	conversation	0:02:02	314	Seibu food court
C036	Conversational	conversation	0:36:55	5,274	Disk 2 Miranda TST Delifrance
C038	Conversational	conversation	0:16:01	2,562	0026 Pacific Place food court
C039	Conversational	conversation	0:10:00	2,111	church
C040	Conversational	conversation	0:21:00	3,853	church
C041	Conversational	conversation	0:22:49	4,311	church
C043	Conversational	conversation	0:27:54	2,832	Chinese restaurant
C044	Conversational	conversation	1:14:00	12,987	
C045	Conversational	conversation	0:13:50	2,741	private club
C046	Conversational	conversation	0:24:41	4,907	home
C048	Conversational	conversation	0:05:29	1,052	Admiralty Delifrance

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
C049	Conversational	conversation	0:05:34	401	YMCA
C050	Conversational	conversation	0:17:19	2,378	McDonald's
C051	Conversational	conversation	0:05:58	688	Disk 1 Admiralty Delifrance
C053	Conversational	conversation	0:01:08	298	TST YMCA
C054	Conversational	conversation	0:42:54	7,272	YMCA restaurant
C057	Conversational	conversation	0:06:26	865	4 persons
C058	Conversational	conversation	0:06:11	788	Delifrance 3 persons
C066	Conversational	conversation	0:10:53	1,494	street
C067	Conversational	conversation	0:17:07	3,434	school playground
C071	Conversational	conversation	0:02:54	314	A31 YMCA
C079	Conversational	conversation	0:19:06	4,212	car
C080	Conversational	conversation	0:37:31	7,137	office
C090	Conversational	conversation	0:45:00	8,228	
C091	Conversational	conversation	0:35:34	8,197	
C094	Conversational	conversation	0:20:38	3,730	
C109	Conversational	conversation	0:46:00	6,316	Chinese restaurant
C111	Conversational	conversation	0:36:00	5,525	
C112	Conversational	conversation	0:11:34	2,723	
C113	Conversational	conversation	0:21:16	4,204	
C114	Conversational	conversation	0:20:54	3,885	
C115	Conversational	conversation	1:22:15	10,080	
C116	Conversational	conversation	0:07:14	1,301	restaurant and on the street
C117	Conversational	conversation	0:21:35	6,372	
C119	Conversational	conversation	0:24:00	4,688	
C121	Conversational	conversation	0:19:30	3,408	
C122	Conversational	conversation	0:15:00	1,960	
P001	Public	speech	0:26:00	3,512	Joint Chambers' luncheon
P002	Public	speech	0:08:08	961	Global Summit of Women
P003	Public	speech	0:26:00	3,041	World Economic Forum's East Asia Economic Summit 2001
P004	Public	speech	0:09:00	1,033	XIIth World Productivity Congress
P005	Public	speech	0:07:00	938	2001 Hong Kong Awards

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					for Industry
P006	Public	speech	0:10:00	1,292	INTEGER Hong Kong Awards for Industry
P007	Public	speech	0:02:30	137	International Day of Remembrance for the tragic events in US
P008	Public	speech	0:05:30	612	Reception to welcome The Walt Disney Company
P009	Public	speech	0:04:00	480	Opening Ceremony of the 'Tourism Hong Kong'
P010	Public	speech	0:08:00	1,049	Hong Kong Salutes the World reception
P011	Public	speech	0:15:00	1,711	HKU 90th Anniversary Dinner
P012	Public	speech	0:03:15	452	Asia-Pacific Broadcasting Union
P013	Public	speech	0:23:00	2,796	Open Components Expo 2000
P014	Public	speech	0:04:00	405	14th General Meeting of the Pacific Economic Cooperation Council (PECC)
P016	Public	speech	0:12:40	1,719	Promotional speech for The Hong Kong Polytechnic University 2001 Education Info Day
P017	Public	speech	0:06:30	446	HKU Int'l Youth Summit Opening Ceremony and Plenary Session

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
P018	Public	speech	0:10:11	1,083	HKU Int'l Youth Summit Opening Ceremony and Plenary Session
P019	Public	speech	0:15:45	2,007	HKU Int'l Youth Summit Opening Ceremony and Plenary Session
P034	Public	interview	0:06:46	1,161	Radio interview with Yip Wing-sie
P036	Public	press briefing	0:06:22	819	Press briefing after presentation of Policy Address
P038	Public	speech	0:01:33	168	Start of Ceremony of International Chinese New Year Parade
P039	Public	speech	0:18:00	2,614	Foreign Correspondent's Club Luncheon
P040	Public	speech	0:09:40	1,127	SCMP Fund Manager of Year 2002 Awards Presentation Ceremony
P041	Public	press briefing	0:01:45	220	Media session after visit of Hospital Authority
P042	Public	press briefing	0:02:01	304	Stand-up briefing after visit of polling station
P043	Public	speech	0:07:54	791	Designation Ceremony of the Safe Communities
P044	Public	press briefing	0:02:22	219	Statement to members of the press regarding purchase of Lexis

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
P045	Public	press briefing	0:06:08	1,084	Stand-up briefing at Central Government Offices
P046	Public	speech	0:15:43	2,272	Joint Business Community Luncheon
P047	Public	speech and Q&A	0:21:38	3,088	CSFB Asian Investment Conference
P048	Public	speech and Q&A	0:44:16	5,772	Joint Chambers' luncheon
P049	Public	speech	0:13:35	1,747	Seminar on New Era of the Pearl River Delta: Further integration with the World Economy
P050	Public	speech	0:08:18	974	Hong Kong International Races Gala Dinner
P051	Public	speech	0:22:53	2,784	Hong Kong Bankers' Association Half-Yearly Dinner
P052	Public	speech	0:10:50	1,102	Leadership Forum 2002
P053	Public	press briefing and Q&A	0:57:20	9,332	Report on Securities and Futures Market Regulatory Structure
P054	Public	speech and Q&A	0:20:26	3,093	2002 Asian Venture Forum
P055	Public	speech	0:11:56	1,448	CNBC/TNT Asia Business Leader Awards Dinner
P058	Public	speech	0:13:48	1,745	Hong Kong Venture Capital Conference
P059	Public	speech	0:05:28	742	Inauguration Ceremony of the Cyberport

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					Institute of Hong Kong
P060	Public	speech	0:14:47	1,990	APEC Business Advisory Council (ABAC) Luncheon
P061	Public	speech	0:13:18	2,202	SARS Update
P062	Public	speech	0:22:02	2,948	Japan Society Luncheon
P066	Public	discussion forum	0:26:14	4,288	SARS Update Discussion Forum
P067	Public	speech	0:15:42	1,918	The opening of the 'Corporate Governance of the New Generation' Studying Programme cum Youth Summit
P068	Public	press briefing and Q&A	0:06:09	779	Media session on civil servant pay cut consensus
P069	Public	speech and Q&A	0.018796	3776	British Chamber of Commerce Luncheon
P070	Public	speech	0:06:30	757	ICAC Interpol Conference
P071	Public	speech	0:02:20	299	Switch On Ceremony of Hong Kong WinterFest
P072	Public	speech	0:16:57	2,139	5th Anniversary Dinner of HK Institute of Directors
P073	Public	speech	0:07:39	942	Award Presentation Ceremony of 2002 HK Awards for Industry
P074	Public	press briefing	0:03:53	575	Statement to members of the

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					press following delivery of Policy Adress by Chief Executive
P075	Public	press briefing	0:03:24	550	Press briefing regarding a cargo vessel with sick crews requesting to enter Hong Kong waters
P076	Public	interview	0:05:02	674	Interview on 'Key to good English: read and speak more' (news.gov.hk)
P077	Public	interview	0:06:01	950	Interview with Secretary for the Environment, Transport & Works (news.gov.hk)
P079	Public	speech	0:04:34	538	Letter to Hong Kong
P080	Public	speech	0:28:53	3,919	Asia Society Luncheon
P083	Public	speech	0:09:10	1,200	Letter to Hong Kong
P087	Public	speech	0:04:11	594	Speech made by General Manager of Lighthouse Technology
P088	Public	speech	0:32:55	4,860	Hong Kong/Pearl River Delta Road Show to Seoul and Suwon, Korea
P089	Public	speech	0:23:32	3,841	HKGCC breakfast workshops
P090	Public	interview	0:07:42	955	TDC Webpage Broadcast
P098	Public	speech	0:09:12	1,247	TDC Annual Dinner in

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					London
P101	Public	interview	0:24:26	4,743	Newsline, ATV World
P102	Public	interview	0:25:05	4,105	Newsline, ATV World
P103	Public	interview	0:24:33	4,441	Newsline, ATV World
P118	Public	interview	0:25:52	4,759	Newsline, ATV World
P119	Public	interview	0:24:05	4,589	Newsline, ATV World
P120	Public	interview	0:24:19	4,668	Newsline, ATV World
P121	Public	interview	0:24:09	4,084	Newsline, ATV World
P122	Public	interview	0:25:17	5,506	Newsline, ATV World
P123	Public	interview	0:24:38	4,791	Newsline, ATV World
P124	Public	interview	0:24:09	4,230	Newsline, ATV World
P125	Public	interview	0:24:54	3,610	Newsline, ATV World
P126	Public	interview	0:25:48	5,195	Newsline, ATV World
P127	Public	interview	0:24:43	5,604	Newsline, ATV World
P128	Public	interview	0:25:00	4,845	Newsline, ATV World
P129	Public	interview	0:25:01	4,994	Newsline, ATV World
P138	Public	speech	0:23:41	3,857	Hong Kong General Chamber of Commerce Luncheon "Health Care Cooperation"
P139	Public	discussion forum	0:12:50	2,411	Hong Kong General Chamber of Commerce Luncheon "Health Care Cooperation" (Q&A session)
P140	Public	speech	0:14:10	1,775	2003 SME

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					Market Day Opening Ceremony
P151	Public	speech	0:07:08	993	Letter to Hong Kong
P152	Public	speech	0:13:43	1,953	Hong Kong Chamber of Commerce roundtable workshop
P153	Public	radio announcement	0:00:30	88	Radio announcement of prevention of SARS
P154	Public	speech	0:08:07	1,127	Letter to Hong Kong
P155	Public	speech	0:08:00	1,197	Letter to Hong Kong
P156	Public	speech	0:07:08	1,005	Letter to Hong Kong
P157	Public	speech	0:09:30	1,196	Letter to Hong Kong
P158	Public	interview	0:03:19	489	Interview with Senior Nursing Officer (news.gov.hk)
P159	Public	speech	0:08:59	1,158	Letter to Hong Kong
P160	Public	speech	0:07:31	981	Letter to Hong Kong
P161	Public	speech	0:08:37	1,062	Letter to Hong Kong
P162	Public	interview	0:09:09	1,558	NewsHour - Conversation with Emily Lau by Elizabeth Farnsworth of PBS
P163	Public	interview	0:31:50	5,106	Shaping Hong Kong's Future - Interview with Anson Chan
P164	Public	speech	0:13:03	1,557	Hong Kong General Chamber of Commerce Luncheon
P167	Public	interview	0:10:49	1,875	NewsHour -

File ID	Domain	Genre	Duration	Word Total	Bibliographical Details
					Interview with Tung Chee Wah by Elizabeth Farnsworth of PBS
P168	Public	interview	0:12:06	1,830	NewsHour - Interview with Tung Chee Wah by Ray Suarez of PBS
P169	Public	interview	0:13:56	2,389	NewsHour - Interview with three finance ministers by Margaret Warner of PBS
P170	Public	speech and Q&A	0:44:17	6,410	Speech given by Chief Secretary for Administration at the National Press Club in Canberra Australia

Appendix 2

Index of the BNC (customised)

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
F8E	academic	complete	lecture	5,550	Newcastle Univ Department of Marine Biology and Coastal Management. Lecture on oceanography (Edu/inf). Rec. on 27 Oct 1992 with 2 partics, 88 utts
HE0	academic	complete	lecture	6,599	King's College London: lecture (Edu/inf). Rec. on 9 Dec 1993 with 2 partics, 169 utts, and lasting 1 hr 0 mins.
HUF	academic	complete	lecture	7,947	King's College London: political philosophy lecture (Edu/inf). Rec. on 9 Dec 1993 with 2 partics, 67 utts, and lasting 55 mins.
HYM	academic	complete	lecture	4,854	Nottingham Univ Economics Department: agricultural economics lecture (Edu/inf). Rec. on 6 Dec 1993 with 2 partics, 11 utts
HYT	academic	complete	lecture	6,148	Nottingham Univ Economics Department: lecture (Edu/inf). Rec. on 14 Dec 1993 with 3 partics, 41 utts
J8K	academic	complete	lecture	7,010	Bioenergetics: lecture (Edu/inf). Rec. on 9 Mar 1993 with 2 partics, 12 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
JP6	academic	complete	lecture	5,720	Computers lecture (Edu/inf). Rec. on ?? Jan 1994 with 2 partics, 7 utts
JT0	academic	complete	lecture	9,403	Aston Univ: lecture (Edu/inf). Rec. on ?? Feb 1994 with 2 partics, 33 utts
KS3	academic	complete	lecture	10,164	Lecture on philosophy. 1 partic, 1 utt
HYD	academic	complete	seminar and tutorial	9,320	King's College London: philosophy discussion class (Edu/inf). Rec. on 9 Dec 1993 with 2 partics, 199 utts, and lasting 1 hr 0 mins.
HYL	academic	complete	seminar and tutorial	5,306	Nottingham Univ Economics Department: tutorial (Edu/inf). Rec. on 6 Dec 1993 with 3 partics, 68 utts
HYN	academic	complete	seminar and tutorial	5,610	Nottingham Univ Economics Department: agricultural economics tutorial (Edu/inf). Rec. on 6 Dec 1993 with 3 partics, 145 utts
HYP	academic	complete	seminar and tutorial	6,509	Nottingham Univ Economics Department: tutorial (Edu/inf). Rec. on 6 Dec 1993 with 6 partics, 157 utts
J94	academic	complete	seminar and tutorial	5,390	Nottingham Univ, Economics Dept. — agricultural economics year 1: tutorial. (Edu/inf). Rec. on 12 Dec 1993 with 4 partics, 63 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
JJN	academic	complete	seminar and tutorial	14,595	Bristol Univ history department: seminar. 3 partics, 659 utts
F8F	academic	complete	consultation	1,020	Newcastle Univ Department of Marine Biology and Coastal Management. Tutorial. (Edu/inf). Rec. on 27 Oct 1992 with 4 partics, 84 utts
FMD	academic	complete	consultation	5,922	York Univ Careers Service: careers interview (Edu/inf). Rec. on 18 Mar 1993 with 2 partics, 602 utts
HDY	academic	complete	consultation	8,699	Careers guidance interview (Edu/inf). Rec. on 18 Mar 1993 with 2 partics, 662 utts
G4V	academic	complete	teaching session with more student talk	7,205	First-year undergraduate tutorial (Edu/inf). 5 partics, 453 utts
G4W	academic	complete	teaching session with more student talk	6,671	First-year undergraduate tutorial: linguistics (Edu/inf). 7 partics, 233 utts
HE2	academic	complete	teaching session with more student talk	8,237	London School of Economics: lecture (Edu/inf). Rec. on 8 Nov 1991 with 2 partics, 98 utts
HUJ	academic	complete	teaching session with more student talk	16,315	London School of Economics: lecture (Edu/inf). Rec. on 7 Nov 1991 with 10 partics, 312 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
JJL	academic	complete	teaching session with more student talk	16,284	Student seminar on absolute egalitarianism in China (Edu/inf). Rec. on 8 Dec 1993 with 2 partics, 642 utts
KGN	academic	complete	teaching session with more student talk	17,248	Student seminar on Hunan report (Edu/inf). Rec. on 1 Dec 1993 with 2 partics, 962 utts
KM6	academic	complete	teaching session with more student talk	14,806	Bristol Univ history department: [XXNOT employee training.] Rec. on 1 Jan 1993 with 2 partics, 786 utts
JJC	business	complete	service encounter	1,428	Estate agency: interview (Busn). Rec. on 13 Jan 1994 with 2 partics, 61 utts
KB0	business	028802	service encounter	522	farm shop, buying
KB8	business	069403	service encounter	444	shopping
KBK	business	002904	service encounter	95	husband and wife talk to waiter / reception / owner in restaurant about reservation made
KBK	business	002906	service encounter	87	husband and wife ordering meals with waiter in restaurant
KBK	business	002909(1)	service encounter	120	husband and wife inquire about take-away with waiter
KBK	business	002920(1)	service encounter	43	husband and wife ordering drinks and dessert
KBK	business	002920(2)	service encounter	32	waiter asking about ordering drinks and dessert
KBK	business	003004	service encounter	45	check, talk with waiter
KC0	business	033303	service encounter	246	shopping, buying food from butcher

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KC0	business	033304	service encounter	142	shopping, buying food from shop assistant
KC9	business	014009(1)	service encounter	86	shopping with daughter, talking to daughter and shop assistant
KC9	business	014009(2)	service encounter	1525	shopping with daughter, talking to daughter and two shop assistants who are friend and acquaintance
KC9	business	014305	service encounter	104	talk to unknown who sounds like shop assistant at shop
KC9	business	014602	service encounter	1062	ask for a refund at a shop for something ordered through catalogue with some general conversation when shopkeeper filling in a form
KCB	business	043003	service encounter	108	buying flour at a shop
KD4	business	048903	service encounter	359	talk with carpet fitter as carpet is fitted at home
KDL	business	016001	service encounter	1557	talk with shop assistant at boutique with grandma and mother
KP8	business	067203	service encounter	285	talk with stranger selling religious magazines at door at home
KBH	business	004002	service encounter	301	talk with shopkeeper, husband and daughter in shop
KBL	business	040506	service encounter	127	talk with shopkeeper in shop
KCK	business	073101(1)	service encounter	95	buying a biscuit in staff room
KCN	business	027404	service encounter	354	talk with staff at gas board

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KCN	business	027405	service encounter	92	shopping, buying a loaf
KCN	business	027504	service encounter	98	buying loaf at bakery
KCN	business	027505	service encounter	16	shopping in shop
KCS	business	062901	service encounter	218	talk at travel agent
KCS	business	063003	service encounter	125	talk to shopkeeper at hardware shop
KDM	business	030504	service encounter	152	talk to chemist to get some chemicals
KDM	business	030507	service encounter	94	shopping
KDM	business	030508	service encounter	161	shopping
KDM	business	030701	service encounter	66	filling petrol at garage
KDV	business	049804	service encounter	54	talk with shopkeeeper in shop
KE4	business	067308	service encounter	797	talk with telephone engineer about setting up phones at home
KPM	business	075603	service encounter	346	making enquries at bus office
F7A	business	complete	meeting	14,378	Tyneside Cinema Board meeting (Busn). Rec. on 29 Jul 1992 with 8 partics, 690 utts
H5D	business	complete	meeting	9,813	Careers Service: meeting (Busn). Rec. on 21 Apr 1993 with 2 partics, 1126 utts
JN7	business	complete	meeting	6,581	Air UK sales meeting (Busn). Rec. on 14 Jan 1994 with 6 partics, 353 utts
JN8	business	complete	meeting	1,796	Northern Development Company: meeting (Busn). Rec. on 18 Jan 1994 with 3 partics, 49 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
JT9	business	complete	meeting	2,347	Meeting at Southwell Racecourse (Busn). Rec. on 18 Jan 1994 with 3 partics, 172 utts, and lasting 15 mins.
KRY	business	complete	meeting	7,140	OUP Electronic Pub. Group: business meeting. 9 partics, 348 utts
J9X	business	complete	interview	12,249	Job interview (Busn). Rec. on 25 Jan 1994 with 2 partics, 733 utts
J9Y	business	complete	interview	11,279	Job interview (Busn). Rec. on 25 Jan 1994 with 2 partics, 498 utts
JA0	business	complete	interview	9,500	Interview (Busn). Rec. on 25 Jan 1994 with 2 partics, 589 utts
JA1	business	complete	interview	8,509	Interview (Busn). Rec. on 25 Jan 1994 with 2 partics, 611 utts
JA2	business	complete	interview	11,220	Interview (Busn). Rec. on 26 Jan 1994 with 2 partics, 1123 utts
JA3	business	complete	interview	8,333	Interview (Busn). Rec. on 26 Jan 1994 with 2 partics, 458 utts
JA4	business	complete	interview	13,931	Interview (Busn). Rec. on 26 Jan 1994 with 2 partics, 932 utts
JAE	business	complete	interview	6,180	Interview (Busn). Rec. on 26 Jan 1994 with 2 partics, 540 utts
HDE	business	complete	presentation	7,105	Enterprise Two Thousand: seminar (Busn). Rec. on 24 Mar 1993 with ?? participants, ?? utts.

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
HDF	business	complete	presentation	7,261	Enterprise Two Thousand: seminar (Busn). Rec. on 24 Mar 1993 with ?? participants, ?? utts.
JSD	business	complete	presentation	4,966	Coopers and Lybrand: seminar (Busn). 2 partics, 13 utts
HDV	business	complete	presentation and Q&A	8,852	Longman Group UK Ltd Lotus 123 seminar (Busn). Rec. on 26 Nov 1993 with 2 partics, 276 utts
HDX	business	complete	presentation and Q&A	5,502	Longman Group UK Ltd Lotus 123 seminar (Busn). Rec. on 26 Nov 1993 with 2 partics, 237 utts
HYW	business	complete	presentation and Q&A	3,602	Commercial Union training session (Busn). Rec. on 21 Nov 1993 with 2 partics, 123 utts
K6V	business	complete	presentation and Q&A	16,837	Tarmac Construction Ltd: training session (Busn). Rec. on 21 Jan 1994 with 2 partics, 179 utts
G4U	business	complete	presentation and Q&A	7,493	Sewage — Severn–Trent Water: talk (Leisure). 2 partics, 159 utts
HDG	business	complete	presentation and Q&A	4,977	Enterprise Two Thousand: seminar (Busn). Rec. on 24 Mar 1993 with ?? participants, ?? utts.
HM6	business	complete	announcement and Q&A	6,711	Pearson: annual general meeting and extraordinary general meeting (Busn). Rec. on 11 May 1990 with 5 partics, 92 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
HYE	business	complete	announcement and Q&A	12,415	Pearson preliminary results: presentation and press and analysts' questions (Busn). Rec. on 1 Apr 1992 with 10 partics, 144 utts, and lasting 1 hr 20 mins.
HYF	business	complete	announcement and Q&A	6,736	Pearson analyst presentation: analysts' meeting (Busn). Rec. on 3 Sep 1993 with 8 partics, 126 utts
J9N	business	complete	announcement and Q&A	5,373	MFI interim figures presentation: press announcement (Busn). Rec. on 27 Jan 1994 with 9 partics, 93 utts, and lasting 20 mins.
KB8	business	069501	informal office talk	4049	visit from friend (clerk to town council) on business at home
KBD	business	063906	informal office talk	1975	at work in nightclub, casual talk about eating and work-related talk about business at nightclub
KBK	business	002010	informal office talk	3312	business talk between export merchant and colleague (export assistant) at home, with some casual talk up to page 5 with respondent's wife
KBK	business	002012	informal office talk	459	business talk between export merchant and colleague (export assistant) at home with some talk about the recording

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KDU	business	046415	informal office talk	1321	company director talk with unknown about business at work
KDU	business	046416	informal office talk	379	company director talk with unknown about business at work
KSR	business	132205	informal office talk	848	talk between property agents at work about a colleague
KC6	business	055904	informal office talk	228	talk between colleagues at work on business matters
KC6	business	055905	informal office talk	486	talk between colleagues at work on business matters
KC6	business	056306	informal office talk	1191	talk between colleagues in office (newspaper agency), casual chat and business talk about stories in newspapers
KCK	business	073101(2)	informal office talk	454	talk with unknown generally about a student's spelling problem in teachers' staff room at school
KD5	business	034301	informal office talk	5222	talk with colleagues at computer on business and general matters
KD5	business	034302	informal office talk	3872	talk with colleagues at computer on business and general matters
KDK	business	016705	informal office talk	176	talk with colleague, going home from Heathrow Airport on bus, colleague taking a phone message for respondent

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KE3	business	010304	informal office talk	435	business related talk with male colleague about resources handbook in civil service
KE3	business	010803	informal office talk	172	business talk about moving things in office
KE3	business	010806	informal office talk	135	short talk about work and general matters
KE3	business	010808	informal office talk	100	short talk about work, finding a filing cabinet
KPJ	business	067707	informal office talk	1772	talk about a colleague and business matters about catering service
KDU	business	046403	workplace telephone talk	362	company director on phone at work
KDU	business	046404	workplace telephone talk	586	company director on phone at work
KDU	business	046413	workplace telephone talk	99	company director on phone at work, reserving a table
KSR	business	132206	workplace telephone talk	143	property agent on phone at work (one-sided talk only)
KC6	business	055704	workplace telephone talk	236	talk between colleagues, one is a photographer on business matters
KB0	conversational	028801	conversation	1069	meeting in street, talk between Margaret (retired) and visiting minister
KB0	conversational	028803	conversation	453	talk between farmer's wife and retired lecturer and missionary
KB0	conversational	028804	conversation	2762	talk between visiting minister and retired lecturer and missionary
KB8	conversational	069601	conversation	7956	having a coffee at neighbor's house

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KB8	conversational	069701	conversation	570	taking parcel to neighbor, neighbor's house
KB8	conversational	070003	conversation	1605	walking to school and having lift back, talk with four friends
KB8	conversational	070004	conversation	4751	having a cup of coffee at friend's home
KBC	conversational	071901	conversation	2320	having tea with husband (teacher) at home
KBC	conversational	071902	conversation	785	talk with nurse friend and husband at home
KBC	conversational	071903	conversation	57	working around the house, talking with acquaintance, 50+, clerk
KBC	conversational	071906	conversation	4152	visiting cousin in Blackpool
KBC	conversational	072101	conversation	1982	having tea with husband at home
KBC	conversational	072102	conversation	75	having tea with husband at home
KBC	conversational	072103	conversation	3496	having tea with husband at home
KBK	conversational	002001	conversation	288	talk with wife at home
KBK	conversational	002002	conversation	337	talk with wife at home
KBK	conversational	002003	conversation	245	talk with wife at home
KBK	conversational	002004	conversation	63	talk with wife at home
KBK	conversational	002005	conversation	65	talk with wife at home
KBK	conversational	002006	conversation	924	talk with wife at home
KBK	conversational	002007	conversation	460	casual talk with wife and colleague at home
KBK	conversational	002008	conversation	534	casual talk with colleague at home
KBK	conversational	002009	conversation	634	casual talk between colleague and wife

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KBK	conversational	002011	conversation	1910	talk with wife at home
KBK	conversational	002016	conversation	247	casual talk with wife and colleague at home
KBK	conversational	002119	conversation	162	casual conversation with colleague on spending the weekend and watering plants
KBK	conversational	002620	conversation	134	talk with neighbour at croquet club
KBK	conversational	002909(2)	conversation	426	talk with wife at Chinese restaurant
KBK	conversational	003006	conversation	529	talk with wife at Chinese restaurant
KBU	conversational	071401	conversation	4258	talking to friends in between and during lessons, at college
KBU	conversational	071501	conversation	6764	talking to friends in between and during lessons, at college
KBW	conversational	002103	conversation	778	talk with husband, housewife friend and unknown at church
KBW	conversational	022301	conversation	1590	talk with husband at home
KBW	conversational	022401	conversation	2205	chatting after evening service with friends, at home
KBW	conversational	022505	conversation	1197	friends have called at home
KBW	conversational	022506	conversation	281	talk with two friends about the piano lesson
KBW	conversational	022507	conversation	428	talking with a friend after piano lesson at home
KBW	conversational	023102	conversation	413	talk with housewife friends and unknown (toddlers club) at church
KBW	conversational	023202	conversation	766	talking with nursery owner and acquaintance at nursery

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KC0	conversational	033402	conversation	5214	talk with husband and a male friend who is an advertising executive at home
KC0	conversational	033403	conversation	1061	talk with husband and a male friend who is an advertising executive at home
KC0	conversational	033404	conversation	253	talk with husband and a male friend who is an advertising executive at home
KC0	conversational	033502	conversation	1142	talk with a male friend who is a retired musician
KC0	conversational	033503	conversation	1903	talk with husband and retired musician friend
KC0	conversational	033504	conversation	295	talk with husband and retired musician friend
KC0	conversational	033505	conversation	6352	talk with two friends who are a retired teacher and a butcher
KC3	conversational	051602	conversation	4994	chatting at home with two friends who are health visitor and chemist
KC3	conversational	051603	conversation	5494	chatting at home with a male friend who is a chemist
KC3	conversational	051801	conversation	5763	chatting after lunch at friend's home with housewife and teacher friends
KC8	conversational	019602	conversation	406	talk between respondent's husband (computer programmer) and two retired friends after lunch at home
KC8	conversational	019603	conversation	919	talk between husband and wife after lunch at home

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KC8	conversational	019703	conversation	2926	having coffee at a friend's house
KC8	conversational	019705	conversation	1591	having coffee at a friend's house
KC8	conversational	019706	conversation	1510	having coffee at a friend's house
KC9	conversational	014001	conversation	93	talk with husband who is a design engineer
KC9	conversational	014201	conversation	477	talk with a friend who is a hair salon owner while buying newspaper at newsagent
KC9	conversational	014307	conversation	3098	talk with a friend who is a hair salon owner at a shop
KC9	conversational	014501	conversation	294	talk with husband who is a design engineer
KC9	conversational	014503	conversation	74	talk with a friend who is a shop assistant
KC9	conversational	014505	conversation	209	talk with a friend who is a shop assistant
KC9	conversational	014506	conversation	4745	talk with husband with a few interactions with shop assistant friend
KCB	conversational	043102	conversation	2847	talk with three female friends who are housewife, coffee shop owner and coffee shop assistant in coffee shop
KCH	conversational	076403	conversation	607	talk with husband in the garden at home
KD4	conversational	048803	conversation	2757	casual talk between respondent and colleague (teacher), with a few interactions between colleague and respondent's three-year-old daughter towards the end

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KD4	conversational	048805	conversation	1606	talk with husband (environmental health officer)
KDR	conversational	071101	conversation	5380	talk with wife about packing
KDU	conversational	046501	conversation	149	picnic with wife and friends who are typist and retired
KPK	conversational	000201	conversation	1673	talk with friends and boyfriend, preparing a meal at home
KPK	conversational	000302	conversation	3228	talk with friends and boyfriend, social club after training ride in cycle club
KPV	conversational	130304	conversation	340	friends dinner party at friends' home
KPV	conversational	130601	conversation	3296	chatting with friends
KPV	conversational	130602	conversation	9511	chatting with friends
KSR	conversational	132201	conversation	568	casual talk with colleague, property agent on non-business matters at home
KSR	conversational	132202	conversation	639	talk with friend and friend's parents at home
KSR	conversational	132301	conversation	6696	talk with friend, friend's father and colleagues
KBG	conversational	033601	conversation	5219	talk with wife at home
KBG	conversational	033602	conversation	252	talk with wife at home
KBH	conversational	003904	conversation	1754	talk with husband and friend at home
KBH	conversational	003908	conversation	1117	talk with husband and friend at home
KBJ	conversational	052008	conversation	190	talk with a friend who is a hairdresser
KBJ	conversational	052009	conversation	190	talk with a friend who is a hairdresser
KBJ	conversational	052302	conversation	1093	talk with husband (area organiser for union) about writing out forms at home

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KBJ	conversational	052303	conversation	163	talk with husband about writing out forms at home
KBJ	conversational	052304	conversation	92	talk with husband about writing out forms and getting dinner ready at home
KBL	conversational	040509	conversation	1304	talk at lunchtime with colleague/friend on non-business matters
KC6	conversational	055604	conversation	1406	picking fiancée up from work and visiting friend in car and at friend’s house with fiancée and friend (hairdresser)
KC6	conversational	055605	conversation	1091	relaxing in friend's house with friend and fiancee
KC6	conversational	055701	conversation	790	talk with fiancee at home
KC6	conversational	055702	conversation	483	talk with fiancee at home
KC6	conversational	055703	conversation	181	talk with fiancee at home
KC6	conversational	056301	conversation	60	preparing to go to work, with fiancee at home
KC6	conversational	056302	conversation	187	preparing to go to work, with fiancee at home
KC6	conversational	056303	conversation	148	preparing to go to work, with fiancee at home
KCK	conversational	073701	conversation	4111	casual talk with colleagues on non-business matters (religion, alcohol etc), sitting and eating in staff room at school
KCS	conversational	062702	conversation	468	chatting with wife and neighbour at home

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KCS	conversational	062704	conversation	468	chatting with wife and neighbour at home
KCS	conversational	062804	conversation	7622	chatting with wife and neighbour at home
KCS	conversational	062902	conversation	1003	chatting with wife and neighbour at home
KCS	conversational	063002	conversation	1726	chatting with wife and neighbour at home, wife having hair done at kitchen
KCS	conversational	063004	conversation	3785	talk with friend on seafront
KD5	conversational	033901	conversation	1718	cooking dinner, talk with wife at home
KD5	conversational	033902	conversation	251	cooking dinner, talk with wife at home
KD5	conversational	033903	conversation	401	cooking dinner, talk with wife at home
KD5	conversational	033904	conversation	167	cooking dinner, talk with wife at home
KD5	conversational	034001	conversation	1010	walking dogs with wife, talk to wife and people walking dogs
KDJ	conversational	041201	conversation	1272	talk about work with friends
KDJ	conversational	041202	conversation	6606	talk about work with friends
KDJ	conversational	041203	conversation	7882	talk about work with friends
KDM	conversational	030002	conversation	4425	talk about television with wife and friend at friend's house
KDM	conversational	030509	conversation	2074	talk with wife and two neighbours at neighbour's
KDM	conversational	030510	conversation	6015	talk with wife and two neighbours at neighbour's
KDM	conversational	030602	conversation	7393	talk with friends at friend's house
KP5	conversational	130901	conversation	3994	talk at home and restaurant with friend and unknown

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KP5	conversational	130902	conversation	3873	talk at home and restaurant with friend and unknown
KPM	conversational	075601	conversation	3916	visit to parish church by local history group, talk about church-related issues
KPM	conversational	075602	conversation	572	talk while walking home on main street
KPM	conversational	075701	conversation	3527	meeting of scrabble club at friend's house
KPM	conversational	075702	conversation	720	meeting of scrabble club at friend's house
KPR	conversational	131401	conversation	672	reading with friend at home
KPR	conversational	131501	conversation	8932	talk with two male friends
KPU	conversational	131701	conversation	8670	talk with sister and two male friends at home
KPU	conversational	131702	conversation	3617	talk with sister and two male friends at home
FUR	public	complete	speech	1,831	Response to treatment in bladder cancers: seminar presentation (Busn). Rec. on 23 Jun 1993 with 2 partics, 10 utts
FUS	public	complete	speech	1,705	Results of prostatectomy survey: seminar presentation (Busn). Rec. on 23 Jun 1993 with 2 partics, 21 utts
HDN	public	complete	speech	9,263	Trade Union Annual Congress: trade union congress (Busn). Rec. on 6 Jun 1993 with 10 partics, 96 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
HDT	public	complete	speech	14,045	Trade Union Annual Congress: trade union congress (Busn). Rec. on 7 Jun 1993 with 10 partics, 147 utts
HLU	public	complete	speech	13,457	Trade Union Annual Congress (Busn). Rec. on 6 Jun 1993 with 10 partics, 116 utts
HLW	public	complete	speech	12,083	Trade Union Annual Congress (Busn). Rec. on 8 Jun 1993 with 10 partics, 101 utts
HUE	public	complete	speech	13,865	Trade Union Annual Congress: trade union congress (Busn). Rec. on 9 Jun 1993 with 10 partics, 142 utts, and lasting 1 hr 35 mins.
HV4	public	complete	speech	1,245	Inaugural address of new president of National Association of Round Tables of Great Britain and Northern Ireland (Leisure). Rec. on 8 May 1993 with 1 partic, 1 utt
J9F	public	complete	speech	4,085	Bradford Metropolitan Council: meeting (Pub/instit). Rec. on 21 May 1991 with an unknown number of partics, 135 utts, and lasting 1 hr 30 mins.

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
J9M	public	complete	speech	17,569	Leicestershire County Council: council chambers (Pub/instit). Rec. on 29 Sep 1993 with 10 partics, 259 utts, and lasting 2 hrs.
F89	public	complete	speech and Q&A	1,830	Birmingham college lecture: student union induction speech (Busn). Rec. on 14 Oct 1992 with 2 partics, 29 utts
F8N	public	complete	speech and Q&A	5,806	Atomic engineering: after-dinner speech at Rotary Club(Leisure). Rec. on 11 Jan 1993 with 6 partics, 37 utts
FUT	public	complete	speech and Q&A	8,922	Presentation on consumer rights (Pub/instit). Rec. on 27 Apr 1993 with 2 partics, 232 utts
G5G	public	complete	speech and Q&A	10,202	Meeting: selection of candidate for election to European Parliament (Pub/instit). Rec. on 25 Aug 1993 with 5 partics, 200 utts
JNL	public	complete	speech and Q&A	6,176	National Council for Voluntary Organisations annual general meeting (Busn). Rec. on 22 Nov 1991 with an unknown number of partics, 62 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
KJS	public	complete	interview	10,046	Call Nick Ross — radio phone-in debating programme: live broadcast. Rec. on 27 Aug 1991 with 17 partics, 184 utts, and lasting 1 hr 5 mins.
KRE	public	complete	interview	15,117	Ideas in Action programmes: radio broadcast. 9 partics, 206 utts
KRF	public	complete	interview	29,528	Ideas in Action programmes (02): radio broadcast. 9 partics, 390 utts
KRK	public	complete	interview	9,098	Bill Heine radio phone-in. 5 partics, 142 utts
HV1	public	complete	interview	5,851	Central Weekend Live — part 1: television broadcast (Leisure). Rec. on 29 Oct 1993 with 10 partics, 370 utts
HV2	public	complete	interview	5,723	Central Weekend Live — part 2: television broadcast (Leisure). Rec. on 29 Oct 1993 with 10 partics, 264 utts
HV3	public	complete	interview	3,866	Central Weekend Live — part 3: television broadcast (Leisure). Rec. on 29 Oct 1993 with 10 partics, 215 utts
HVC	public	complete	interview	6,683	Central Weekend Live — part 1: television broadcast (Leisure). Rec. on 5 Nov 1993 with 10 partics, 460 utts

File ID	Domain	Part number	Genre	Word Total	Bibliographical Details
HVD	public	complete	interview	5,563	Central Weekend Live — part 2: television broadcast (Leisure). Rec. on 5 Nov 1993 with 10 partics, 221 utts, and lasting 28 mins.
HVE	public	complete	interview	3,954	Central Weekend Live — part 3: television broadcast (Leisure). Rec. on 5 Nov 1993 with 10 partics, 276 utts, and lasting 20 mins.

Appendix 3.1

Two pages from the teaching section of the textbook database

Express your attitude that it is not surprising that this good idea is popular and can be used not only for celebrations, but business promotion and advertising.

You could use "The passage also warns us..." to introduce the ideas of the fourth paragraph.

Close by expressing your own enthusiasm for the idea (or, if you prefer, say it is silly).

Vagueness, etc.

Our examples of presentations are a bit unreal as they are written, not spoken. In actual speech we use the phrases *sort of* and *kind of*.

- *The printer uses some kind of edible ink.*
- *It's sort of a crazy idea.*
- *They use some sort of stuff you can eat – like rice paper.*
- *It's suitable for all kinds of parties and occasions.*

Here are two other common little phrases (that should not be overused):

- *I mean / I mean to say*
- *You know*

The first is used to clarify:

- *You use paper – I mean paper you can eat.*
- *It's a brilliant idea. I mean I'd love to have a birthday cake with my picture on it.*

The second encourages the listener to accept and agree:

- *You know, I just read a really interesting passage about a new sort of cake.*
- *You scan the picture and you print it, you know, just like you usually do, but...*

Reproduced from Leetch (2002:44)

Extras

Needing clarification

- *I'm sorry, but I'm not really following.*
- *I'm afraid I don't get your point.*
- *I'm not sure what exactly you mean.*

Occasional minor contributions

- *Fine*
- *Yes, I see.*
- *Mmm – Uhu*
- *Sure.*
- *Good thinking.*

Organising

- *Well, let's move on to the next point.*
- *Well, let's not rush it.*
- *Aren't we going off the point a bit?*

Interrupting

- *Well, O.K. – let's see what everyone else has to say about it.*
- *Can I just break in there?*

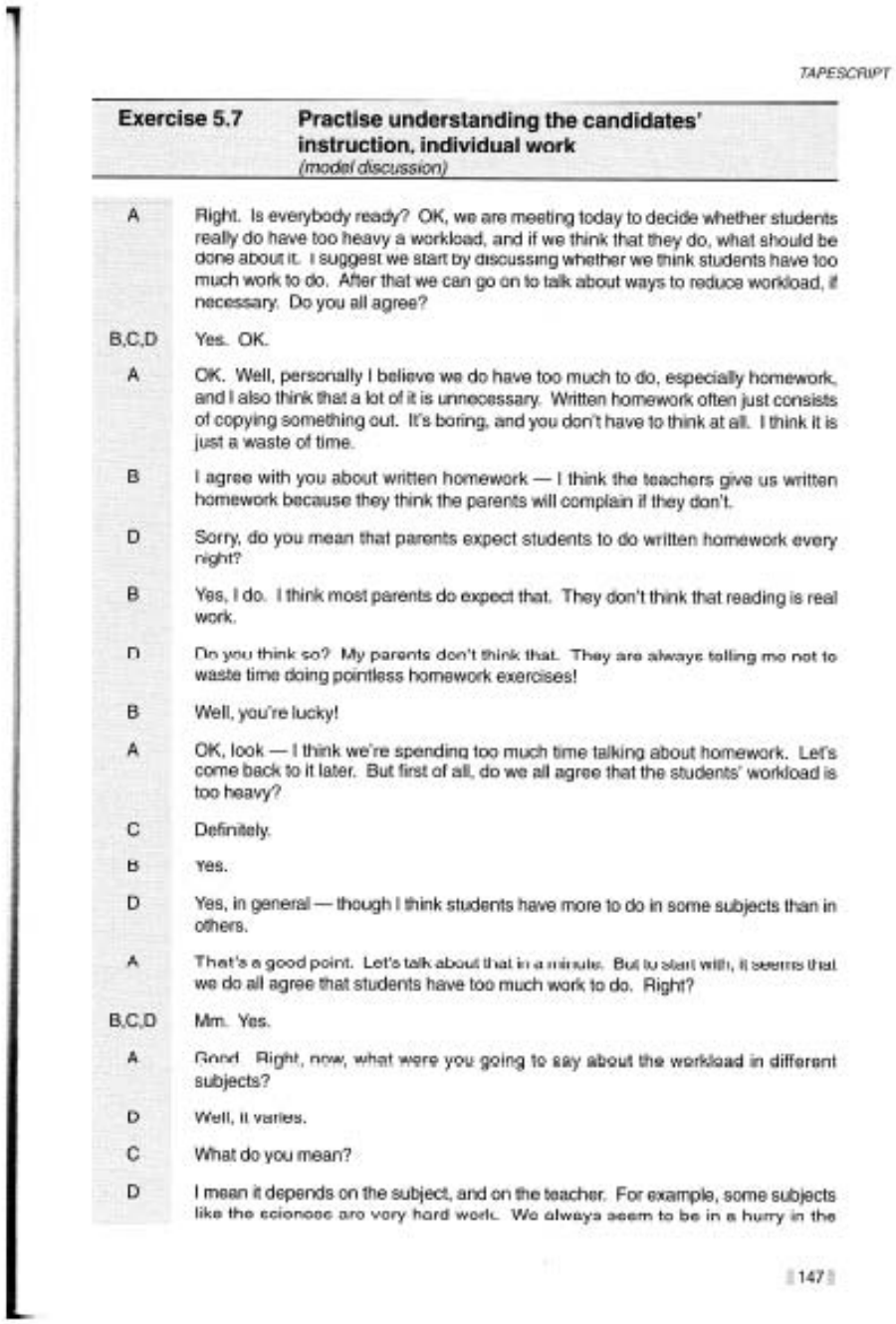
Ending

- *Well, that seems to be it, doesn't it?*
- *Time's running out so that will be our last suggestion.*

Reproduced from Leetch (2002:70)

Appendix 3.2

Two pages from the sample section of the textbook database



TAPESCRIPT

Exercise 5.7 **Practise understanding the candidates' instruction, individual work**
(model discussion)

A	Right. Is everybody ready? OK, we are meeting today to decide whether students really do have too heavy a workload, and if we think that they do, what should be done about it. I suggest we start by discussing whether we think students have too much work to do. After that we can go on to talk about ways to reduce workload, if necessary. Do you all agree?
B,C,D	Yes. OK.
A	OK. Well, personally I believe we do have too much to do, especially homework, and I also think that a lot of it is unnecessary. Written homework often just consists of copying something out. It's boring, and you don't have to think at all. I think it is just a waste of time.
B	I agree with you about written homework — I think the teachers give us written homework because they think the parents will complain if they don't.
D	Sorry, do you mean that parents expect students to do written homework every night?
B	Yes, I do. I think most parents do expect that. They don't think that reading is real work.
D	Do you think so? My parents don't think that. They are always telling me not to waste time doing pointless homework exercises!
B	Well, you're lucky!
A	OK, look — I think we're spending too much time talking about homework. Let's come back to it later. But first of all, do we all agree that the students' workload is too heavy?
C	Definitely.
B	Yes.
D	Yes, in general — though I think students have more to do in some subjects than in others.
A	That's a good point. Let's talk about that in a minute. But to start with, it seems that we do all agree that students have too much work to do. Right?
B,C,D	Mm. Yes.
A	Good. Right, now, what were you going to say about the workload in different subjects?
D	Well, it varies.
C	What do you mean?
D	I mean it depends on the subject, and on the teacher. For example, some subjects like the sciences are very hard work. We always seem to be in a hurry in the

147

Reproduced from Sutton (1999a:147)

A possible candidate presentation

I'm the proud owner of a lovely six-year-old black and white cat called Sunshine, so you can imagine how pleased I was to be given a passage about keeping a cat. The writer says that cats are very suitable as pets here in Hong Kong as our homes are too small for dogs – and certainly I think cats are more fun than fish!

The writer gives lots of good advice about having a cat. He tells you not to buy one but to go along to the SPCA and choose one of the cats there waiting for a home. You can choose a young kitten, or if you think they are too playful and active for you, you can have an older one. One advantage of going to the SPCA is that they check all the cats they receive, so you know you'll be receiving a healthy, neutered cat.

You don't really need to make many preparations for your cat. They don't need anything special. An old brush to keep its fur in good condition and a couple of old food bowls are okay, though of course, you've got to supply a toilet for the animal. Luckily cats are clean creatures and it is easy to train them to use a cat box. Certainly I have no trouble with my Sunshine – we just have to empty out the box once a day.

Cats have different likes and dislikes. Some like milk, some don't. Mine prefers water. You can give food leftovers and tinned cat food. Just remember not to feed them too much – or you'll have a cat like Garfield – and that they are basically meat-eaters.

Then we come to the most important part. Give your cat lots of attention and love so that it is happy with you. Certainly, I agree with the writer that having a cat is a wonderful thing. That's all. Thank you.

Reproduced from Leetch (2002:12)

References

Aijmer, K. (1988). “Now may we have a word on this”: The use of *now* as a discourse particle. In Kytö, M., Ihalainen, O. & Rissanen, M. (eds.) *Corpus linguistics, hard and soft: Proceedings of the Eighth International Conference on English language research on computerized corpora*. (pp. 15-34). Amsterdam: Rodopi.

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London; New York: Longman.

Aijmer, K. (2001). ‘It’s okay’. In Ljung, M (ed.) *Language structure and variation. A festschrift for Gunnel Melchers*. *Acta Universitatis Stockholmiensis. Stockholm studies in English 92*. (pp.1-17). Stockholm: Almqvist & Wiksell International.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam; Philadelphia: John Benjamins.

Aijmer, K. (2003). Discourse particles in contrast. In Bondi, M. & Stati, S. (eds.) *Dialogue analysis 2000: Selected papers from the 10th IADA anniversary conference, Bologna, 2000*. (pp. 63-78). Tübingen: Max Niemeyer Verlag.

Aijmer, K. (2005). Evaluation and pragmatic markers. In Tognini-Bonelli, E. & Del Lungo Camiciotti, G. (eds.) *Strategies in academic discourse*. (pp. 83-96). Amsterdam: John Benjamins.

Aijmer, K. & Simon-Vandenbergen, A-M. (2003). The discourse particle *well* and its equivalents in Swedish and Dutch. *Linguistics*, 41(6), 1123-61.

Altenberg, B. (1987). *Prosodic patterns in spoken English. Studies in the correlation between prosody and grammar for text-to-speech conversion. Lund studies in English 76*. Lund: Lund University Press.

Altenberg, B. (1990). Spoken English and the dictionary. In Svartvik, J. (ed.) *The London-Lund Corpus of Spoken English: Description and research*. (pp. 177-192). Lund: Lund University Press.

American Psychological Association (2001). *Publication manual of the American Psychological Association (5th ed.)*. Washington, DC: American Psychological Association.

Andersen, G. (1998). The pragmatic marker *like* from a relevance-theoretic perspective. In Jucker, A. H. & Ziv, Y. (eds.) *Discourse markers:*

Descriptions and theory. (pp. 147-170). Amsterdam; Philadelphia: John Benjamins.

Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation. A relevance-theoretic approach to the language of adolescents*. Amsterdam; Philadelphia: John Benjamins.

Andersen, E. S., Brizuela, M., DuPuy, B. & Gonnerman, L. (1999). Cross-linguistic evidence for the early acquisition of discourse markers as register variables. *Journal of Pragmatics*, 31, 1339-1351.

Atkinson, J. M. & Heritage, J. (eds.). (1984). *Structures of social action*. Cambridge: Cambridge University Press.

Bach, K. (1994). Conversational implicature. *Mind and Language*, 9, 124-162.

Bacon-Shone, J. & Bolton, K. (1998). Charting multilingualism: Language censuses and language surveys in Hong Kong. In Pennington, M. (ed.) *Language in Hong Kong at century's end*. (pp. 43-90). Hong Kong: Hong Kong University Press.

Bazzanella, C. (1990). Phatic connectives as interactional cues in contemporary spoken Italian. *Journal of Pragmatics*, 14, 629-647.

Bazzanella, C. & Morra L. (2000). Discourse markers and the indeterminacy of translation. In Korzen, I. & Marello, C. (eds.) *Argomenti per una linguistica della traduzione, On linguistic aspects of translation, Notes pour une linguistique de la traduction*. (pp.149-157). Alessandria: Edizioni dell' Orso.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Longman.

Bilbow, G. (1997). Cross-cultural impression management in the multicultural workplace: The special case of Hong Kong. *Journal of Pragmatics*, 28, 461-487.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (1988). *So* as a constraint on relevance. In Kempson, R. M. (ed.) *Mental representations. The interface between language and reality*. (pp. 183-195). Cambridge: Cambridge University Press.

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

Bolinger, D. (1977). *Meaning and form*. London: Longman.

Bolinger, D. (1989). *Intonation and its uses. Melody in grammar and discourse*. London: Edward Arnold.

Bolt, P. & Bolton, K. (1995). The international corpus of English in Hong Kong. (manuscript).

Brazil, D. (1985). *The communicative value of intonation*. Birmingham: English Language Research, University of Birmingham.

Brazil, D. (1997). *The communicative role of intonation in English*. Cambridge: Cambridge University Press.

Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin; New York: Mouton de Gruyter.

Brown, G. (1977). *Listening to spoken English*. London: Longman.

Brown, G. & Yule, G. (1983). *Discourse analysis*. Cambridge; New York: Cambridge University Press.

Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Bublitz, W. (1988). *Supportive fellow-speakers and cooperative conversations*. Amsterdam; Philadelphia: John Benjamins.

Burnard, L. (ed.) (2000). *The reference guide for the British National Corpus (World Edition)*. Oxford: Oxford University Computing Services.

Butler, C. (1985). *Statistics in linguistics*. Oxford: Basil Blackwell.

Carlson, L. (1984). *'Well' in dialogue games: A discourse analysis of the interjection 'well' in idealized conversation*. Amsterdam: John Benjamins.

Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52(1), 43-56.

Carter, R., Hughes, R. & McCarthy, M. (2000). *Exploring grammar in context. Upper-intermediate and advanced*. Cambridge: Cambridge University Press.

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

Cheng, W. (2003). *Intercultural conversation*. Amsterdam; Philadelphia: John Benjamins.

Cheng, W. (2007). The use of vague language across spoken genres in an intercultural corpus. In Cutting, J. (ed.) *Vague language explored*. (pp. 161-181). London: Palgrave Macmillan.

Cheng, W., Greaves, C. & Warren, M. (2005). The creation of prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic). *International Computer Archive of Modern English (ICAME) Journal*, 29, 5-26.

Cheng, W., Greaves, C. & Warren, M. (forthcoming). *A corpus-driven study of discourse intonation: The Hong Kong Corpus of Spoken English (Prosodic)*. Amsterdam; Philadelphia: John Benjamins.

Cheng, W. & Warren M. (2000). Using a corpus of intercultural spoken English as a learning and teaching resource: The case of the discourse marker *well*. In *Proceedings of the '98 International Conference on Teaching English at Tertiary Level in the Chinese Context*. (pp. 374-391). Beijing, China: Tsinghua University Press.

Cheng, W. & Warren, M. (2001). The functions of *actually* in a corpus of intercultural conversations. *International Journal of Corpus Linguistics*, 6(2), 257-280.

Cheng, W., & Warren, M. (2005). //well I have a DIFferent // THINking you know //: A corpus-driven study of disagreement in Hong Kong business discourse. In Gotti, M. & Bargiela, F. (eds.) *Asian business discourse(s).* (pp. 241-270). Frankfurt am main: Peter Lang.

Cheng, W. & Warren, M. (2006). "// R you need to be RUTHless //: Entertaining cross-cultural differences. *Language & Intercultural Communication*, 6(1), 35-56.

Cheng, W., Warren, M. & Xu, X. F. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System*, 31, 173-186.

Cobuild (1987). *Collins Cobuild English language dictionary*. London: Collins.

Cobuild (2002). *Collins Cobuild advanced learner's English dictionary*. London: HarperCollins.

Coulthard, M. (1977). *An introduction to discourse analysis*. London: Longman.

Cramer, D. (1998). *Fundamental statistics for social research. Step-by-step calculations and computer techniques using SPSS for Windows*. London: Routledge.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Crystal, D. (1975). *The English tone of voice. Essays in intonation, prosody and paralanguage*. London: Edward Arnold.

Curriculum Development Council. (1999). *CDC syllabus for use of English (sixth form)*. Hong Kong.

de Klerk, V. (2005). Procedural meanings of *well* in a corpus of Xhosa English. *Journal of Pragmatics*, 37, 1183-1205.

Erman, B. (1987). *Pragmatic expressions in English: A study of* you know, you see *and* I mean *in face-to-face conversation*. *Stockholm studies in English 69*. Stockholm: Almqvist & Wiksell.

Erman, B. (2001). Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics*, 33, 1337-1359.

Esser, D. (1999). *Teach & practice: AS-level oral English for form 6*. Hong Kong: Pilot Publications.

Evans, S. (1997). Teacher and learner roles in the Hong Kong English language classroom. *Educational Journal*, 25(2), 43-61.

Evans, S. & Green, C. (2001). Language in post-colonial Hong Kong: The roles of English and Chinese in the public and private sectors. *English World-Wide*, 22(2), 247-268.

Ferrara, K. W. (1997). Form and function of the discourse marker *anyway*: implications for discourse analysis. *Linguistics*, 35, 343-378.

Fischer, K. (2006). Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume. In Fischer, K. (ed.) *Approaches to discourse particles*. (pp. 1-20). Amsterdam: Elsevier.

Fischer, K. & Drescher, M. (1996). Methods for the description of discourse particles: contrastive analysis. *Language Sciences*, 18(3-4), 853-861.

Foolen, A. (1997). Pragmatic particles. In Verschueren, J., Östman, J. –O., Blommaert, J. & Bulcaen, C. (eds.). *Handbook of Pragmatics 1996 Installment*. (pp. 1-24). Amsterdam; Philadelphia: John Benjamins.

Forsyth, R. (2007). *Chi-squared program*. Retrieved May 7, 2007, from http://www.psychology.nottingham.ac.uk/staff/rsf/misc.html.

Forsyth, R., Clarke, D. & Lam, P. (2008). Timelines, talk and transcription: A chronometric approach to simultaneous speech. *International Journal of Corpus Linguistics*, 13(2), 225-250.

Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1-4), 19-33.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14, 383-395.

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

Fraser, B. (2006). Towards a theory of discourse markers. In Fischer, K. (ed.) *Approaches to discourse particles*. (pp. 189-204). Amsterdam: Elsevier.

Fuller, J. M. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, 35, 23-45.

Fung, L. & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410-439.

García Vizcaíno, M. J. & Martínez-Cabeza, M. A. (2005). The pragmatics of *well* and *bueno* in English and Spanish. *Intercultural Pragmatics*, 2(1), 69-92.

Goldberg, J. (1980). *Discourse particles: An analysis of the role of* y'know, I mean, well *and* actually *in conversation*. *PhD dissertation*. University of Cambridge.

Goodwin, R. (1999). *Personal relationships across cultures*. London; New York: Routledge.

Graddol, D. (1997). *The Future of English?* London: British Council.

Graddol, D. (2006). *English next*. London: British Council.

Gran, B. J. (1993). *Use of English oral practice papers form 7*. Hong Kong: Macmillan.

Greasley, P. (1994). An investigation into the use of the particle *well*: Commentaries on a game of snooker. *Journal of Pragmatics*, 22, 477-494.

Greaves, C. (2005). *iConc*©. Hong Kong: The Hong Kong Polytechnic University.

Grice, H. P. (1975). Logic and conversation. In Cole, P. & Morgan, J. (eds.) *Syntax and Semantics 3: Speech Acts*. (pp. 41-58). New York: Academic Press.

Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, Mass: Harvard University Press.

Gudykunst, W. B. (1994). *Bridging differences: Effective intergroup communication*. 2nd edition. Thousand Oaks, CA: Sage.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hansen, M. M. –B. (1998). *The function of discourse particles. A study with special reference to spoken standard French*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

He, A. (2002). On the discourse marker *so*. In Peters, P., Collins, P. & Smith, A. (eds.) *New frontiers of corpus research.* (pp. 41-52). Amsterdam, Netherlands: Rodopi.

He, A. W. & Lindsey, B. (1998). 'You know' as an information status enhancing device. Arguments from grammar and interaction. *Functions of Language*, 5(2), 133-155.

Healey, J. F. (2007). *The essentials of statistics: A tool for social research*. Belmont, CA: Thomson Wadsworth.

Heritage, J. (1984). A change of state token and aspects of its sequential placement. In Atkinson, J. M. & Heritage, J. (eds.) *Structures of social action* (pp. 299-345). Cambridge: Cambridge University Press.

Hewings, M. (ed.) (1990). *Papers in discourse intonation*. Birmingham: English Language Research, University of Birmingham.

Hewings, M. & Cauldwell, R. T. (1997). Foreword. In Brazil, D., *The communicative role of intonation in English*. (pp. v-viii). Cambridge: Cambridge University Press.

Hirschberg, J. & Litman, D. (1993). Empirical studies on disambiguation of cue phrases. *Computational Linguistics*, 19, 501-530.

Hofstede, G. (1994). *Cultures and organizations: Software of the mind*. London: Harper-Collins.

Holmes, J. (1984). Hedging your bets and sitting on the fence: Some evidence for hedges as support structures. *Te Reo*, 27, 47-62.

Holmes, J. (1995). *Meaning in interaction: An introduction to pragmatics*. London; New York: Longman.

Hong Kong Examinations Authority. (1998). *Guidelines for candidates: Use of English oral examination*. Hong Kong: Hong Kong Examinations Authority.

Hong Kong Examinations and Assessment Authority. (2003). *Syllabuses for the use of English. Advanced supplementary level*. Hong Kong: Hong Kong Examinations and Assessment Authority.

James, A. R. (1983). '*Well*' in reporting clauses: Meaning and form of a 'lexical filler'. *Arbeiten aus Anglistik und Amerikanistik*, 8(1), 33-40.

Jefferson, G. (1983). On a failed hypothesis: 'Conjunctionals' as overlap-vulnerable. *Tilburg Papers in Language and Literature*, 28, 29-33.

Johansson, S. (2006). How well can *well* be translated? On the English discourse particle *well* and its correspondences in Norwegian and German. In Aijmer, K. & Simon-Vandenbergen, A-M. (eds.) *Pragmatic markers in contrast*. (pp. 115-138). Amsterdam: Elsevier.

Jucker, A. H. (1993). The discourse marker *well*: A relevance theoretical account. *Journal of Pragmatics*, 19, 435-452.

Jucker, A. H. & Smith, S. W. (1998). *And people just you know like 'wow'*: Discourse markers as negotiating strategies. In Jucker, A. H. & Ziv, Y. (eds.) *Discourse markers: Descriptions and theory*. (pp. 171-201). Amsterdam; Philadelphia: John Benjamins.

Jucker, A. H. & Ziv, Y. (1998). Discourse markers: Introduction. In Jucker, A. H. & Ziv, Y. (eds.) *Discourse markers: Descriptions and theory*. (pp. 1-12). Amsterdam; Philadelphia: John Benjamins.

Kennedy, G. (1998). *An introduction to corpus linguistics*. Harlow: Longman.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746.759.

Kroon, C. (1995). *Discourse particles in Latin. A study of* nam, enim, autem, vero *and* at. Amsterdam: J.C. Gieben, Publisher.

Kyratzis, A. & Ervin-Tripp, S. (1999). The development of discourse markers in peer interaction. *Journal of Pragmatics*, 1321-1338.

Lakoff, R. (1973). Questionable answers and answerable questions. In Kachru, B. B., Lees, R. B., Malkiel, Y., Pietrangeli, A. & Saporta, S. (eds.) *Issues in linguistics. Papers in honor of Henry and Renée Kahane*. (pp. 453-467). Urbana, IL: University of Illinois Press.

Lam, P. (2006). *Well but that's the effect of it*: The use of *well* as a discourse particle in talk shows. *Sprache und Datenverarbeitung (International Journal for Language Data Processing)*, 30(1), 99-108.

Lam, P. (2007). A needle in a haystack: The making of a *BNC* customised corpus. (manuscript submitted to the journal *Corpora* under review). The Hong Kong Polytechnic University, Hong Kong.

Lee, D. Y. W. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the *BNC* jungle. *Language Learning & Technology, special issue on "Using Corpora in Language Teaching and Learning"*, 5 (3): 37-72.

Lee, D. Y. W. (December 23, 2003). The *BNCW Index*. Retrieved August 25, 2006, from http://clix.to/davidlee00.

Lee, I. (1998). Enhancing ESL students' awareness of coherence-creating mechanisms in writing. *TESL Canada Journal*, 15(2), 36-49.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.

Leech, G. & Svartvik, J. (2002). *A communicative grammar of English*. Harlow: Longman.

Leetch, P. (2002). *Use of English oral handbook*. Hong Kong: Free Press.

Lenk, U. (1995). Discourse markers and conversational coherence. In Warvik, B., Tanskanen, S. –K. & Hiltunen, R. (eds.) *Anglicana Turkuensia 14: Organization in discourse. Proceedings from the Turku Conference*. (pp. 341-352). Turku: University of Turku.

Lenk, U. (1998). *Marking discourse coherence: Functions of discourse markers in spoken English*. Tübingen: Gunter Narr.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Li, A. (2003). *Use English. Volume 2*. Hong Kong: Macmillan.

Ljung, M. (1990). *A study of TEFL vocabulary*. Stockholm: Almqvist and Wiksell International.

Local, J. & Kelly, J. (1986). Projection and 'silences': notes on phonetic and conversational structure. *Human Studies*, 9, 185-204.

Luke, K. K. (1987). The establishment of common ground in natural conversation: The case of the utterance particle *LA* in Cantonese. Paper read at the 1987 International Pragmatics Conference, Antwerp.

Luke, K. K. (1990). *Utterance particles in Cantonese conversation*. Amsterdam; Philadelphia: John Benjamins.

Macaulay, R. (2002). You know, it depends. *Journal of Pragmatics*, 34, 749-767.

Macaulay, R. (2005). *Talk that counts. Age, gender, and social class differences in discourse*. Oxford: Oxford University Press.

McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McCarthy, M., McCarten, J. & Sandiford, H. (2005a). *Touchstone. Student's book 1*. Cambridge: Cambridge University Press.

McCarthy, M., McCarten, J. & Sandiford, H. (2005b). *Touchstone. Student's book 2*. Cambridge: Cambridge University Press.

McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Meyerhoff, M. (1992). *We've all got to go one day, eh*: Powerlessness and solidarity in the functions of a New Zealand tag. In Hall, K., Bucholtz, M. & Moonwomon, B. (eds.) *Locating power: Proceedings of the Second Berkeley Women and Language Conference*, Berkeley, California. (pp.409-419). Berkeley: Berkeley Women and Language Group.

Mukherjee, J. (2001). *Form and function of parasyntactic presentation structures. A corpus-based study of talk units in spoken English*. Amsterdam: Rodopi.

Möllering M. (2004) *The acquisition of German modal particles. A corpus-based approach*. Bern: Peter Lang.

Müller, S. (2004). *'Well you know that type of person'*: Functions of *well* in the speech of American and German students. *Journal of Pragmatics*, 36, 1157-1182.

Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam; Philadelphia: John Benjamins.

Nelson, M. (2000). *A corpus-based study of the lexis of business English and business English teaching materials*. Unpublished thesis. Manchester: University of Manchester.

O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.

Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Owen, M. (1981). Conversational units and the use of '*well*…'. In Werth, P. (ed.) *Conversation and discourse: Structure and interpretation*. (pp.99-116). London: Croom Helm.

Östman, J. –O. (1982). The symbiotic relationship between pragmatic particles and impromptu speech. In Enkvist, N. E. (ed.) *Impromptu speech: A symposium.* (pp. 147-177). Åbo: The Research Institute of the Åbo Akademi Foundation.

Passot, F. (2007). A because B so A': Circularity and discourse progression in conversational English. In Celle, A. & Huart, R. (eds.) *Connectives as discourse landmarks*. (pp. 117-134). Amsterdam: John Benjamins.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: some features of preferred / dispreferred turn shapes. In Atkinson, J. M. & Heritage, J. (eds.) *Structures of social action*. (pp. 57-101). Cambridge: Cambridge University Press.

Potter, J. (2003a). *Steps & skills oral 6*. Hong Kong: Witman.

Potter, J. (2003b). *Steps & skills oral 7*. Hong Kong: Witman.

Prodromou, L. (2003). In search of the successful user of English. *Modern English Teacher*, 12(2), 5-14.

Quirk, R. (1955). Colloquial English and communication. *Studies in Communication* (pp. 169-180). London: Secker and Warburg.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1972). *A grammar of contemporary English*. London: Longman.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research*. San Francisco: Jossey-Boss.

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14, 367-381.

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

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In Fischer, K. (ed.) *Approaches to discourse particles*. (pp. 339-358). Amsterdam: Elsevier.

Risselada, R. & Spooren, W. (1998). Introduction: Discourse markers and coherence relations. *Journal of Pragmatics*, 30, 131-133.

Roach, P. (1983). *English phonetics and phonology. A practical course*. Cambridge: Cambridge University Press.

Romero Trillo, J. (2001). A mathematical model for the analysis of variation in discourse. *Linguistics*, 37, 527-550.

Romero Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, 34, 769-784.

Rouchota, V. (1998). Procedural meaning and parenthetical discourse markers. In Jucker, A. H. & Ziv, Y. (eds.) *Discourse markers: Descriptions and theory*. (pp. 97-126). Amsterdam; Philadelphia: John Benjamins.

Römer, U. (2005). *Progressives, patterns, pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam; Philadelphia: John Benjamins.

Sacks, H., Schegloff, E. & Jefferson, G. (1978). A simplest systematics for the organization of turn-taking for conversation. In Schenkein, J. (ed.) *Studies in the organization of conversational interaction*. (pp. 7-55). New York: Academic Press.

Schegloff, E. & Sacks, H. (1973). Opening up closings. *Semiotica* VIII(4), 290-327.

Schiffrin, D. (1982). *Discourse markers: semantic resource for the construction of conversation*. Ph.D. dissertation, University of Pennsylvania.

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

Schourup, L. (1985). *Common discourse particles in English conversation*. New York: Garland.

Schourup, L. (1999). Tutorial overview: Discourse markers. *Lingua*, 107, 227-265.

Schourup, L. (2001). Rethinking *well*. *Journal of Pragmatics*, 33, 1025-1060.

Scott, M. (2005). *Wordsmith Tools 4.0*. Oxford: Oxford University Press.

Siegel, S. & Catellan, N. J. Jr. (1988). *Nonparametric statistics for the behavioural sciences*. Second edition. New York: McGraw Hill.

Sinclair, J. (1991). *Corpus, concordance, collocations*. Oxford: Oxford University Press.

Sinclair, J. (1999). A way with common words. In Hasselgård, H. & Oksefjell, S. (eds.) *Out of corpora: Studies in Honour of Stig Johansson*. (pp. 157-179). Amsterdam; Atlanta, GA: Rodopi.

Sinclair, J. (2001). Preface. In Ghadessy, M., Henry, A. & Roseberry, R. L. (eds.) *Small corpus studies and ELT. Theory and practice* (pp. vii-xv). Amsterdam: John Benjamins.

Sinclair, J. (2004). Introduction. In Sinclair, J. (ed.) *How to use corpora in language teaching*. (pp.1-10). Amsterdam: John Benjamins.

Sinclair, J. (2007a). Collocation reviewed. (manuscript). Tuscan Word Centre, Italy.

Sinclair, J. (2007b). Defining the definiendom – new. (manuscript). Tuscan Word Centre, Italy.

Sinclair, J. & Coulthard, M. (1975). *Towards an analysis of discourse*. London: Oxford University Press.

Sinclair, J. & Mauranen, A. (2006). *Linear unit grammar: Integrating speech and writing*. Amsterdam; Philadelphia: John Benjamins.

Smith, S. W. & Jucker, A. H. (2000). *Actually* and other markers of an apparent discrepancy between propositional attitudes of conversational partners. In Andersen, G. & Fretheim, T. (eds.) *Pragmatic markers and propositional attitude*. (pp. 207-238). Amsterdam; Philadelphia: John Benjamins.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

SPSS Inc. (2006). *SPSS 15.0 for Windows*. Chicago, IL: SPSS Inc.

Stenström, A. -B. (1984). *Questions and responses in English conversation*. *Lund studies in English 68*. 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*. (pp. 137-175). Lund: Lund University Press.

Stenström, A. -B. (1994). *An introduction to spoken interaction*. New York: Longman.

Stenström, A. -B. (2002). What is going on between speakers. In Partington, A., Morley, J. & Haarman, L. (eds.) *Corpora and discourse*. (pp. 259-283). Bern: Peter Lang.

Stenström, A. -B. (2006). The Spanish discourse markers *o sea* and *pues* and their English correspondences. In Aijmer, K. & Simon-Vandenbergen, A-M. (eds.) *Pragmatic markers in contrast*. (pp. 155-172). Amsterdam: Elsevier.

Stenström, A. -B. & Andersen, G. (1996). More trends in teenage talk: A corpus-based investigation of the discourse items *cos* and *innit*. In Percy, C. et al. (eds.) *Synchronic corpus linguistics*. (pp. 141-152). Amsterdam: Rodopi.

Stubbe, M. & Holmes, J. (1995). *You know, eh* and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication,* 15(1), 63-88.

Stubbs, M. (2000). Using very large text collections to study semantic schemas: a research note. In C. Heffer & H. Saunston (eds.), *Words in context*. Birmingham University. [CD-ROM.]

Sutton, M. E. & Duncan, J. (1999). *Speaking precisely 1*. Hong Kong: Precise Publications.

Svartvik, J. (1980). *Well* in conversation. In Greenbaum, S., Leech, G. N. & Svartvik, J. (eds.) *Studies in English linguistics for Randolph Quirk*. (pp. 167-177). London: Longman.

Swan, M. (1996). *Practical English Usage*. Second Edition. International Student's Edition. Oxford: Oxford University Press.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: Morrow.

Tannen, D. (1994). *Gender and discourse*. New York: Oxford University Press.

The British National Corpus, version 2 (BNC World). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium.

Traugott, E. C. (1995). Subjectification in grammaticalization. In Stein, D. & Wright, S. (eds.) *Subjectivity and subjectivisation in language*. (pp. 31-54). Cambridge: Cambridge University Press.

Traugott, E. C. & König, E. (1991). The semantics-pragmatics of grammaticalization revisited. In Traugott, E. C. & Heine, B. (eds.) *Approaches to grammaticalization. Volume 1. Focus on theoretical and methodological issues*. (pp. 189-218). Amsterdam; Philadelphia: John Benjamins.

Tribble, C. (1997). Improvising corpora for ELT. Quick and dirty ways of developing corpora for language teaching. In Lewandowska-Tomaszczyk, B. & Melia, P. (eds.) *Palc '97: Practical applications in language corpora*. (pp. 106-118). Lödz: Lödz University Press.

Tribble, C. & Jones, W. (1990). *Concordances in the classroom: A resource book for teachers*. Harlow: Longman.

Tsui, A. B. M. (1995). *Introducing classroom interaction*. London: Penguin English.

Tsui, A. B. M. (2004). Medium of instruction in Hong Kong: One country, two systems, whose language? In Tollefson, J. W. & Tsui, A. B. M. (eds.) *Medium of instruction policies: Which agenda? Whose agenda?* (pp.97-106). Mahwah, N.J.: Lawrence Erlbaum Associates.

Warren, M. (2006). *Features of naturalness in conversation*. Amsterdam: John Benjamins.

Watts, R. J. (1986). Relevance in conversational moves: A reappraisal of *well*. *Studia Anglica Posnaniensia,* 19, 37-60.

Watts, R. J. (1988). A relevance-theoretic approach to commentary pragmatic markers: The case of *actually*, *really* and *basically*. *Acta Linguistica Hungarica,* 38, 235-260.

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, 203-237.

Wichmann, A. (2001). Spoken parentheticals. In Aijmer, K. (ed.) *Gothenburg Studies in English 81. A wealth of English: Studies in honour of Göran Kjellmer*. (pp. 177-193). Göteborg: Acta Universitatis Gothoburgensis.

Wierzbicka, A. (1976). Particles and linguistic relativity. *International Review of Slavic Linguistics,* 1(2-3), 327-367.

Wierzbicka, A. (1986). Precision in vagueness. The semantics of English approximatives. *Journal of Pragmatics,* 10, 597-614.

Wierzbicka, A. (1991). *Cross-cultural pragmatics. The semantics of human interaction*. Berlin: Mouton de Gruyter.

Zar, J. H. (1999). *Biostatistical analysis*. 4th edition. Englewood Cliffs, N.J.: Prentice Hall.

Zorzi, D. (2001). The pedagogic use of spoken corpora: Learning discourse markers in Italian. In Aston, G. (ed.) *Learning with corpora*. (pp. 85-107). Houston, TX: Athelstan.

Zwicky, A. M. (1985). Clitics and particles. *Language,* 61, 283-305.