

**Discourse markers in Saudi English and British English:  
A comparative investigation of the use of English discourse markers**

**Sami Algouzi**

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of  
Philosophy

University of Salford, Salford, UK

School of Arts and Media

September 2014

## Table of Contents

Table of contents.....	i
List of tables.....	iv
List of figures.....	v
Dedication.....	vii
Acknowledgements.....	viii
List of abbreviations and acronyms.....	ix
Abstract.....	x
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
1.1 Introduction.....	1
1.2 Corpus linguistics.....	2
1.3 Aims of the study.....	4
1.4 Hypotheses and research questions.....	5
1.5 Significance of the study.....	8
1.6 Summary and outline of the thesis.....	9
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>11</b>
2.1 Introduction.....	11
2.2 Previous research on discourse markers .....	11
2.2.1 Terminology.....	12
2.2.2 Characteristics of discourse markers .....	13
2.3 Some specific functions of discourse markers.....	27
2.4 Major approaches to discourse markers.....	29
2.4.1 Discourse markers in the coherence-based approach .....	30
2.4.2 Discourse markers in the relevance-based framework .....	35
2.5 Discourse markers and language learning .....	38
2.6 The present study .....	45
<b>CHAPTER 3: DATA DESCRIPTION AND RESEARCH METHODOLOGY.....</b>	<b>47</b>
3.1 Introduction.....	47
3.2 Data description .....	47
3.2.1 Corpora .....	47
3.3 Research methodology.....	70
3.3.1 Functional analysis.....	73
<b>CHAPTER 4: <i>SO</i> IN THE DISCOURSE OF THE SLC AND LOCNEC.....</b>	<b>83</b>
4.1 Introduction.....	83
4.2 Functions of <i>so</i> in the literature .....	83
4.3 Functions of <i>so</i> in the SLC and LOCNEC.....	93
4.3.1 Non-discourse marker functions of <i>so</i> .....	93
4.3.2 <i>So</i> indicating a result.....	97

4.3.3 <i>So</i> marking a main idea unit.....	104
4.3.4 <i>So</i> introducing a new sequence .....	107
4.3.5 <i>So</i> marking transition .....	110
4.3.6 <i>So</i> introducing a turn/discourse.....	116
4.3.7 Unclear instances .....	118
4.3.8 Summary of the functions of <i>so</i> .....	120
4.4 Quantitative results for functions of <i>so</i> .....	121
4.5 Comparison of findings in the corpus and textbooks.....	127
4.6 Summary and conclusion.....	130
<b>CHAPTER 5: <i>YOU KNOW</i> IN THE DISCOURSE OF THE SLC AND LOCNEC.....</b>	<b>132</b>
5.1 Introduction.....	132
5.2 Functions of <i>you know</i> in the literature.....	132
5.3 Functions of <i>you know</i> in the SLC and LOCNEC .....	147
5.3.1 Non-discourse marker functions of <i>you know</i> .....	147
5.3.2 <i>You know</i> signalling lexical or content search .....	148
5.3.3 <i>You know</i> for restart and repair .....	152
5.3.4 <i>You know</i> introducing an explanation .....	155
5.3.5 <i>You know</i> for speaker/hearer shared knowledge .....	157
5.3.6 <i>You know</i> for generally shared knowledge .....	162
5.3.6 <i>You know</i> in narratives .....	166
5.3.7 Unclear instances .....	170
5.3.8 Summary of the discourse marker functions of <i>you know</i> .....	171
5.4 Quantitative results for <i>you know</i> .....	172
5.5 Comparison of findings in the corpus and textbooks.....	177
5.6 Summary and conclusion.....	179
<b>CHAPTER 6: <i>LIKE</i> IN THE DISCOURSE OF THE SLC AND LOCNEC .....</b>	<b>181</b>
6.1 Introduction.....	181
6.2 Functions of <i>like</i> in the literature .....	181
6.3 Functions of <i>like</i> in the SLC and LOCNEC.....	189
6.3.1 Non-discourse marker functions of <i>like</i> .....	190
6.3.2 Searching for the appropriate expression.....	193
6.3.3 Marking an approximate number.....	197
6.3.4 <i>Like</i> introducing an example .....	202
6.3.5 <i>Like</i> marking focused information .....	206
6.3.6 <i>Like</i> preceding a restart .....	210
6.3.7 Unclear instances .....	211
6.4 Quantitative results for functions of <i>like</i> .....	213

6.5 Comparison of findings in the corpus and textbooks.....	218
6.6 Summary and conclusion.....	219
<b>CHAPTER 7: CONCLUSIONS .....</b>	<b>220</b>
7.1 Introduction.....	220
7.2 Summary of the study .....	220
7.3 Research questions revisited .....	222
7.4 Limitations of the study .....	230
7.5 Recommendations for further research .....	232
7.6 Concluding remarks .....	234
<b>APPENDIX 1.....</b>	<b>235</b>
<b>APPENDIX 2.....</b>	<b>236</b>
<b>APPENDIX 3.....</b>	<b>237</b>
<b>APPENDIX 4.....</b>	<b>242</b>
<b>APPENDIX 5.....</b>	<b>244</b>
<b>APPENDIX 6.....</b>	<b>245</b>
<b>APPENDIX 7.....</b>	<b>247</b>
<b>APPENDIX 8.....</b>	<b>250</b>
<b>APPENDIX 9.....</b>	<b>251</b>
<b>APPENDIX 10.....</b>	<b>252</b>
<b>APPENDIX 11.....</b>	<b>259</b>
<b>APPENDIX 12.....</b>	<b>260</b>
<b>APPENDIX 13.....</b>	<b>261</b>
<b>REFERENCES.....</b>	<b>262</b>

## List of Tables

Table 2.1 Inventory of discourse markers in modern English .....	18
Table 2.2 Discourse marker functions .....	28
Table 2.3 Planes of discourse marker functions .....	32
Table 3.1 Number of interviews and words in LOCNEC and the SLC.....	51
Table 3.2 Discourse markers found in the pilot data: raw and normalized frequencies .....	55
Table 3.3 Universities and college in which data were collected .....	56
Table 3.4 Interviewers' gender, mother tongue and institute .....	59
Table 3.5 Frequency of discourse markers per 10,000 words for 10 recordings in the SLC ..	66
Table 3.6 Number of most frequent discourse markers in the 10 recordings and their frequency per 10,000 words.....	67
Table 3.7 Contingency table for word frequency (Rayson & Garside, 2000, p.3) .....	73
Table 4.1 Semantic realization of <i>so</i> and <i>because</i> in discourse .....	84
Table 4.2 Functions of <i>so</i> and their abbreviations .....	121
Table 4.3 Numbers of tokens of <i>so</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses) .....	122
Table 4.4 Functions of <i>so</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses).....	123
Table 4.5 Number of tokens of <i>so</i> in the SLC according to gender (frequencies per 10,000 words in parentheses).....	124
Table 4.6 Functions of <i>so</i> in SLC according to gender (frequencies per 10,000 words in parentheses).....	125
Table 4.7 Summary of <i>so</i> in the SLC and LOCNEC .....	131
Table 5.1 Functions of <i>you know</i> and their abbreviations .....	172
Table 5.2 Numbers of tokens of <i>you know</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses).....	173
Table 5.3 Distribution of <i>you know</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses) .....	174
Table 5.4 Number of tokens of <i>you know</i> in the SLC according to gender (frequencies per 10,000 words in parentheses).....	175
Table 5.5 Functions of <i>you know</i> in the SLC according to gender (frequencies per 10,000 words in parentheses).....	176
Table 5.6 Summary of <i>you know</i> in the SLC and LOCNEC .....	180
Table 6.1 Functions of <i>like</i> and their abbreviations .....	213
Table 6.2 Numbers of tokens of <i>like</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses) .....	213
Table 6.3 Distribution of the discourse marker <i>like</i> in the SLC and LOCNEC (frequencies per 10,000 words in parentheses).....	215

Table 6.4 Number of tokens of <i>like</i> in the SLC according to gender (frequencies per 10,000 words in parentheses).....	<b>216</b>
Table 6.5 Functions of <i>so</i> in the SLC according to gender (frequencies per 10,000 words in parentheses).....	<b>216</b>
Table 7.1 Most frequent discourse markers in the SLC (frequencies per 10,000 words in parentheses).....	<b>224</b>
Table 7.2 All discourse marker functions in the SLC according to their levels .....	<b>226</b>
Table 7.3 Number of tokens of the three discourse markers in the SCL and LOCNEC (frequencies per 10,000 words in parentheses).....	<b>227</b>
Table 7.4 Example of a suggested activity to introduce discourse markers in textbooks .....	<b>233</b>

## List of Figures

Figure 3.1 Core components of learner corpus research (Granger, 2009, p. 15) .....	53
Figure 3.2 Map of Saudi Arabia .....	57
Figure 3.3 Screenshot of AntConc showing concordance lines of <i>like</i> in LOCNEC .....	63
Figure 3.4 Screenshot of an Excel spreadsheet showing the analysis of the discourse marker <i>like</i> in the SLC .....	64
Figure 4.1 Distribution of the linguistic item <i>so</i> – SLC versus LOCNEC .....	122
Figure 4.2 Distribution of discourse marker functions of <i>so</i> – SLC versus LOCNEC .....	124
Figure 4.3 Distribution of the linguistic item <i>so</i> in the SLC according to gender .....	125
Figure 4.4 Distribution of discourse marker functions of <i>so</i> in the SLC according to gender .....	126
Figure 5.1 Meta-knowledge concerning speaker/hearer shared knowledge .....	137
Figure 5.2 Distribution of the linguistic item <i>you know</i> – SLC versus LOCNEC .....	173
Figure 5.3 Distribution of <i>you know</i> as discourse marker – SLC versus LOCNEC .....	174
Figure 5.4 Distribution of the linguistic <i>you know</i> in the SLC according to gender .....	175
Figure 5.5 Distribution of DM functions of <i>you know</i> in SLC according to gender .....	176
Figure 6.1 Distribution of the linguistic item <i>like</i> – SLC versus LOCNEC .....	214
Figure 6.2 Distribution of <i>like</i> as discourse marker – SLC versus LOCNEC .....	215
Figure 6.3 Distribution of the linguistic <i>like</i> in the SLC according to gender .....	216
Figure 6.4 Distribution of discourse marker functions of <i>like</i> in the SLC according to gender .....	217
Figure 7.1 Frequency of <i>like</i> , <i>so</i> and <i>you know</i> in the SLC .....	224
Figure 7.2 Distribution of the three discourse markers between the SLC and LOCNEC ....	227

## **Dedication**

To my late father, Ibrahim Bin Abdullah, and my mother, Alwafyiah Bint Hamad, I dedicate  
this work.



## **Acknowledgements**

First and foremost, I would like to express my gratitude to my supervisors, Iván García-Álvarez and Nicholas Smith, for their academic support and meaningful ideas. I would like to express my gratitude to all the interviewers from the four universities where I collected the data for this study. Another ‘thank you’ goes to the participants who agreed to be interviewed. To all of them I say, without your cooperation this thesis could not have been completed.

My sincere thanks to my mother for her constant support and motherly call every day to make sure I was fine and doing well. I would also like to thank my wife Sharifah and my children Mohammed, Alwafiyah, Ibrahim, Ahmed and Hanin who tolerated all the time of stress and bad moods. I am grateful for their unceasing emotional and moral support throughout the journey of my PhD studies. Sincere thanks also go to my brothers, Mohammed, Akrum, Musaab and Anas, who were always there for me whenever I needed them.

## List of abbreviations and acronyms

DM*	Discourse marker
SLC	Saudi learners corpus
LINDSEI	Louvain International Database of Spoken English Interlanguage
LOCNEC	Louvain Corpus of Native English Conversation
NSs	Native speakers
NNSs	Non-native speakers
L1	First language
L2	Second language
CIA	Contrastive interlanguage analysis
CA	Contrastive analysis

\* For ease of reading, I do not abbreviate discourse marker (except in tables), but it is abbreviated in certain quotations.

## Abstract

Based on two corpora, one of Saudi learners and the other of native English speakers, this thesis investigates qualitatively and quantitatively the use of English discourse markers in the speech of advanced Saudi learners of English in the third and fourth years of undergraduate study of English and compares it to the use of discourse markers by native speakers of English. Three of the most frequently occurring discourse markers in the spoken language, namely *so*, *you know* and *like*, are analysed. Qualitatively, the results from the Saudi learners' corpus show that the three discourse markers under investigation serve a variety of discourse functions. In particular, they show that native speakers of English use *so* and *like* more frequently than Saudi learners. *You know* is used more frequently by Saudis. These results introduce to the research field of discourse markers a new conceptualization of how non-native Saudi English language learners use discourse markers in their speech. Even though the results from the Saudi learners' corpus show that the three discourse markers are used with a variety of discourse functions, the analysis of the textbooks shows that of the three discourse markers, *so* is the only one introduced. This makes it difficult to make a strong claim about the connection of the local pedagogy and the use of the discourse markers. Saudi English learners are probably able to acquire them through their exposure to the media and through their interaction in English with their peers.

## CHAPTER 1: INTRODUCTION

### 1.1 Introduction

There have been noticeable advances in the fields of discourse analysis and pragmatics during the last few decades. Brown and Yule (1983) define the term *discourse analysis* as the analysis of language in use (p. 1), whereas *pragmatics*, as defined by Yule (1996, p. 3), is concerned with the study of meaning as communicated by a speaker (or writer) and interpreted by a listener (or reader). Advances in these two major fields of linguistics, i.e. discourse analysis and pragmatics, have afforded an increased understanding of discourse markers, and the meanings and appropriate use of discourse markers in spoken discourse – the focus of this thesis – have attracted the attention of many linguists (for example, Blakemore, 1987; Brinton, 1996; Erman, 1987; Fraser, 1996; Östman, 1981; Schiffrin, 1987; Schourup, 1985). Many of these linguists argue that these linguistic items are not only considered to mark discourse but also enable the speaker to produce a coherent and intact utterance. The use of discourse markers and their linguistic features will be discussed in more detail in the literature review in chapter 2.

The majority of language learners, if not all, have many aims in learning a language. One of these aims, which probably comes number one on their list, is to speak the language appropriately. This view is supported by Sinclair and Mauranen (2006) who claim that the spoken mode of any language is used before the written mode and therefore the emphasis in descriptions of language should be on spoken language. By extension, discourse markers are of importance as one of the common features of spoken English (Carter & McCarthy, 2006) and indeed very many languages (Beeching & Detges, 2014). This, of course, does not mean

that the use of discourse markers in written discourse is not valuable. As far as the Arabic language is concerned, the amount of research which has addressed discourse markers in spoken language is less than the amount of research which has focused on written Arabic; studies such as those undertaken by Al Kohlani (2010), Al-Saif and Markert (2010), Alhuqbani (2013) and Alghamdi (2014) focus on discourse markers in written Arabic. Nonetheless, the focus of this study is on spoken language.

This PhD thesis aims to investigate the English spoken by foreign language speakers in a foreign language context, namely Saudi Arabia. The investigation is focused on the use of English discourse markers by advanced Saudi students who are in their third and fourth years of undergraduate study of English at four Saudi universities. Because of the highly heterogeneous nature of discourse markers (Brinton, 1996; Schiffrin, 1987), the study does not aim to analyse every single discourse marker expression found in Saudi English discourse. The focus will be on the three discourse markers which occur most frequently in the Saudi learner corpus (henceforth SLC), which have different semantic and pragmatic characteristics, and which have previously been covered in the literature on discourse markers. The use of discourse markers by the students in this study will be compared to those used by native speakers (NSs).

## **1.2 Corpus linguistics**

One of the approaches to studying spoken language is to investigate a ‘corpus’ or (in the plural) ‘corpora’ of running text data. Corpus linguistics refers to the study of language based on a collection of naturally occurring ‘real life’ examples (Baker, 2010; Hunston, 2002; McEnery & Wilson, 1996). The use of natural data is one of the main criteria in any corpus.

Baker (2010) argues that natural examples of language help to ‘derive rules or explore trends about the ways in which people actually produce language’ (p. 94). Corpora are designed so as to be representative of some kind of language (Biber, 1993). In addition, corpora have been used in the design of dictionaries, such as *Collins COBUILD Advanced Learner’s English Dictionary* (2006) and the *Longman Dictionary of Contemporary English* (2009). They have also been used in several areas in Applied Linguistics (Hunston, 2002).

Hunston (2002) lists eight different types of corpora which are commonly used:

- specialized corpora, which contain texts of a particular type, e.g. newspaper editorials;
- general corpora, which consist of texts of many types; comparable corpora, containing two or more corpora in different languages or different varieties of the same language, e.g. British English and American English;
- parallel corpora, in which the same texts are produced in two or more languages;
- learner corpora, which comprise texts produced by learners of a language;
- pedagogic corpora, which consist of all the language – including coursebooks, readers, tapes, etc. – to which a learner has been exposed;
- historical corpora, in which texts from different periods of times are collected;
- monitor corpora, which are designed to track changes in a language.

The main corpus used in this study is a learner corpus. It is debatable whether the learner corpus used, the Saudi Learner Corpus (SLC), satisfies Baker’s criterion of ‘naturally occurring examples’, because the data were elicited through the somewhat formal technique of a structured interview. However, a set of guidelines for conducting the interviews was applied in order to increase the level of spontaneity and naturalness of these interviews (see

section 3.2.1). Moreover, considerable effort was made to ensure that the participants sampled were broadly representative of Saudi university-level learners of English.

Building a spoken corpus is time-consuming. However, in a study such as this, based in the Saudi context, there was no available spoken corpus to investigate, which led to the compilation of my own corpus to be able to answer the research questions posed in this thesis (see 1.2). A major additional feature of this project is that the corpus will form part of the Louvain International Database of Spoken English Interlanguage (LINDSEI) project.<sup>1</sup> The common design of LINDSEI provides a good model for comparison of different non-native varieties of English with those of NNs (the strengths and weaknesses of LINDSEI will be examined in depth in chapter 3). This corpus model was selected because it provides the same setting and task for all participants, natives and non-natives, so that the resulting transcripts can be perfectly comparable and are thus ideally suited to the purpose of my research.<sup>2</sup> In addition, being a partner in this project allows me and other scholars a wide range of opportunities for comparison of Arabic with native and other non-native varieties of English, both in terms of discourse markers and other linguistic features.

### **1.3 Aims of the study**

The main aim of this study is to investigate the use of English discourse markers by advanced Saudi English language learners and to systematically compare these discourse markers to the same discourse markers used by native speakers of English in order to determine the similarities and differences between them and possible reasons for those differences. For this purpose, I will: a) use both a corpus of Saudi learner English, compiled by myself, and a

---

<sup>1</sup> For further details see: <http://www.uclouvain.be/en-cecl-lindsei.html>

<sup>2</sup> My thanks go to Dr. Nicholas Smith for recommending these corpora and for introducing me to the LINDSEI team.

widely available native-speaker reference corpus called LOCNEC (Louvain Corpus of Native English Conversation)<sup>3</sup>; and b) examine the ways in which discourse markers are represented in Saudi textbooks. More specifically the aims of this study are:

1. To identify the frequency and use of the selected discourse markers by Saudi learners of English.
2. To analyse the functions that the three selected discourse markers fulfil in the discourse of Saudi learners of English.
3. a) To analyse the frequency and use of the selected discourse markers in LOCNEC.  
b) To compare the similarities and differences (including frequency and functions) in the use of the selected discourse markers between the Saudi learners of English and native speakers of English.
4. To examine some of the independent variables which influence the use of the selected discourse markers (e.g. age, gender, time spent in an English-speaking country, etc.).
5. To compare the representation of discourse markers in Saudi English textbooks with their usage in the corpus of Saudi learners in order to examine the extent to which attention is given to these lexical items in the local pedagogical setting.

#### **1.4 Hypotheses and research questions**

This research sets out to test a number of hypotheses related to the use of discourse markers by Saudi non-native speakers (NNSs) of English and NSs. First, I hypothesize that there will be variations of discourse marker use (quantitative and qualitative) not only between the Saudi English language speakers and English NSs, but also between the Saudi speakers themselves based on certain factors (linguistic/non-linguistic), for example gender, years of

---

<sup>3</sup> All native speakers in LOCNEC were British students majoring in English language and/or linguistics at the University of Lancaster.



English study and learning English in an informal context. Second, I also hypothesize that the linguistic output of students taught in the Saudi educational system (elementary, intermediate and secondary level) will be affected/influenced directly by their exposure to the input of their teaching both in terms of the frequency with which they use discourse markers and the range of functions that they exhibit.<sup>4</sup>

My first hypothesis is motivated by the increase in interest in the description of how learners write and speak English (Hunston, 2002). It is also inspired by the research findings on the use of discourse markers by a number of studies, although the conclusions in these studies relate to different language backgrounds (Fuller, 2003; Fung & Carter, 2007; Hays, 1992; Moreno, 2001; Müller, 2005; Sankoff et al, 1997; Trillo, 2002) and the studies do not include speakers whose first language (L1) is Arabic.

In support of the second hypothesis, it has been noted that discourse markers seldom attract attention in the formal language classroom (Müller, 2004; Trillo, 2002). Instead, as pointed out by Hellermann and Vergun (2007) ‘many language learners have had “grammatical” language as the primary goal of their language learning experiences’ (p. 158). For example, some studies (Kasper, 1982; Kasper, 1989; Lörcher, 1986) have shown that grammar-centred classroom instruction results in insufficient politeness marking by the students because they do not gain sufficient practice in the conversational strategies needed in contexts outside the classroom (Kasper & Schmidt, 1996, p. 161). Moreover, it is noticeable that language learners often get corrected when they use non-standard morphology or syntax,

---

<sup>4</sup> Saudi English language textbooks are the same at all education levels for both boys and girls, i.e. primary, intermediate and secondary.

but they are much less likely to be corrected when discourse markers are used inappropriately (Svartvik, 1980, as cited in Müller, 2004, p. 1158). Svartvik (1980) comments:

‘...if a foreign language learner says *five sheeps* or *he goed*, he can be corrected by practically every native speaker. If, on the other hand, he omits a *well*, the likely reaction will be that he is dogmatic, impolite, boring, awkward to talk to, etc., but a native speaker cannot pinpoint an “error”.’ (p. 171)

These hypotheses are also based on a number of other factors. One factor is my own informal observation during my experience as an English language teacher and supervisor in Saudi Arabia for more than 15 years. Saudi students seem to struggle to achieve the pragmatic appropriateness of native speakers with the use of discourse markers. The situation is similar to that described by Hellermann and Vergun (2007), Sankoff et al (1997) and Trillo (2002) in that the use of discourse markers is not explicitly taught in most Saudi classrooms. However, pragmatic knowledge is teachable and its instruction does make a difference in learners’ development (Bardovi-Harlig, 2001; Fukuya et al, 1998; Matsuda, 1999; Rose & Kasper, 2001). In addition, my hypotheses are informed by analysis of the spoken data collected for my MA dissertation, which consisted of two recordings of two English teaching classes (Algouzi, 2008).<sup>5</sup> Although the focus of that dissertation was not on discourse markers (it was on barriers to students developing speaking skills), nevertheless the study suggested that discourse markers were surprisingly rare in the data and that where they were found, they were often used inappropriately.

---

<sup>5</sup> These recordings are not quite the data I am looking to use in a study such as this – they relate to classroom discourse in secondary schools – but it was the point which engaged my motivation to undertake this study.

Thus, to test my hypotheses, the following research questions are explored in the thesis:

1. To what extent, if at all, do advanced Saudi learners of English (as represented by third and fourth year undergraduate students) use discourse markers in their spoken English language?
2. Which discourse markers are used most frequently in the discourse of advanced Saudi learners of English?
3. What discourse functions do the selected discourse markers fulfil in the discourse of advanced Saudi learners of English?
4. To what extent are Saudi students similar to or different from – both quantitatively and qualitatively – native speakers of English in their use of the selected discourse markers?<sup>6</sup>
5. What are the variables, also called ‘linguistic/non-linguistic’ factors by Müller (2005), which might influence the Saudi use of the selected discourse markers (e.g. gender, speaking other languages, time spent in an English-speaking country)?
6. To what extent are the selected discourse markers represented in Saudi textbooks? How is this connected to the actual output of these speakers?

### **1.5 Significance of the study**

This study is significant because it is the first study to investigate discourse markers in the speech of Saudi NNSs and thus will contribute to the research on Saudi speakers of English. The results of this study will, for the first time, provide an insight into how Saudi learners of English language use discourse markers. The analysis of the functions of the selected discourse markers (those most commonly used) will raise awareness among Saudi speakers

---

<sup>6</sup> Data from native speakers of English are based on LOCNEC (the Louvain Corpus of Native English Conversation), <http://www.uclouvain.be/en-cecl-lindsei.html>.

concerning what functions each of the selected discourse markers has in the spoken discourse. I believe that when it comes to NNSs, understanding/knowledge of the functions of discourse markers may contribute to better usage. Another significant aspect of this study is the use of corpus methodology. It is expected that this study will not only benefit linguists in the field of discourse marker research but also will benefit teaching English in Saudi Arabia. The investigation of Saudi textbooks aims to uncover the extent to which discourse markers are represented.

### **1.6 Summary and outline of the thesis**

This introductory chapter provides an overview of the research. It begins with a brief background to the topic. It illustrates the reason why I have chosen to study discourse markers in spoken discourse and not in written discourse. It then sets out the research hypotheses and questions and describes how this study is to be carried out by collecting particular data which will enable me to answer the research questions.

The remaining chapters of the thesis are as follows. Chapter 2 presents a literature review of some of the key studies in the field of discourse markers which are relevant to this study. It offers an overall presentation of the term discourse marker and the major characteristics. The chapter also discusses the general functions of discourse markers. This is followed by a discussion of the major frameworks employed in relation to discourse markers. I also review some previous studies which have addressed discourse markers, opening up the way to identify the gap into which this study fits.

Chapter 3 introduces a detailed description of the data, the process used for data collection, and data analysis procedures. It first gives an overview of the history of the corpora and what they can provide researchers with. It then provides a description of this study's corpus. In this chapter, I also discuss the database of textbooks and why examining them is of significance. Finally, the research methodology followed in this study is introduced.

Chapters 4, 5 and 6 provide in-depth analyses of the English discourse markers *so*, *you know*, and *like* based on the empirical results derived from my corpus. These chapters will make reference to textbooks to examine the extent to which the discourse markers under investigation are represented in these textbooks. The last chapter, chapter 7, presents the main findings and conclusions of this study, relating them to my research questions, as well as the limitations and some recommendations for further research.

## CHAPTER 2: LITERATURE REVIEW

### 2.1 Introduction

This chapter aims to present the key issues and frameworks which address discourse markers. Section 2.2 gives an overview of the research field. Section 2.2.1 sheds light on issues of terminology and 2.2.2 presents the characteristics of discourse markers reported in earlier studies. In section 2.3 the functions of discourse markers are discussed. In 2.4 major studies which consider discourse markers are examined. In the review of previous studies in section 2.5 there is an emphasis on the relationship between discourse markers and language learning. The final section, 2.6, shows how the present study may fill a gap in the current literature.

### 2.2 Previous research on discourse markers

Since the 1970s, discourse markers have been the focus of a number of studies. These studies, as Schourup (1999, p. 228) states, were interested in how discourse markers are produced by speakers and received by listeners. However, the field of research on discourse markers is ‘far from being homogeneous or unified’ (Lenk, 1998, p. 37). Researchers still disagree on how this class of linguistic expressions should be defined. Let us consider the following comments which Lenk (1998, p. 38) says is the full extent of Levinson’s treatment of discourse markers in English:

‘...there are many words and phrases in English, and no doubt most languages, that indicate the relationship between an utterance and the prior discourse. Examples are utterance-initial use of *but, therefore, in conclusion, to the contrary, still, however, anyway, well, besides, actually, all in all, so, after all*, and so on. It is generally conceded that such words have at least a component of meaning that resists truth-

conditional treatment [...] what they seem to do is indicate, often in very complex ways, just how the utterance that contains them as a response to, or a continuation of, some portion of the prior discourse.’ (Levinson, 1983, p. 87 et seq.)

This quotation gives a comprehensive description of discourse markers. However, in this quotation, Levinson (1983) does not give a name to these *words and phrases in English*. Indeed, researchers in the field of discourse markers do not always agree on how to refer to the object of their interest or what items they should be studying. These disagreements have yielded different perspectives on discourse markers and various approaches through which they are investigated. The literature on discourse markers shows that some studies have examined more than one discourse marker (for example, Aijmer, 2002; Fraser, 1990; Lenk, 1998; Schiffrin, 1987; Schourup, 1985), whereas other studies have focused on a single discourse marker (Erman, 2001; Lakoff, 1973, as cited in Müller, 2005). Moreover, there have been differences in the frameworks employed to investigate these discourse markers. Some are ‘top-down’ studies, focusing on how the discourse markers under investigation could potentially fit in a particular theoretical framework (for example Lenk, 1995; Lenk 1998). In some other cases, discourse markers have been investigated using a ‘bottom-up’ or data-driven approach which analyses each single instance before making a generalization (see, for example, Müller, 2005). More discussion of approaches to the study of discourse markers will be provided later in this chapter (cf. section 2.4) following an overview of the terminology and the characteristics and functions of discourse markers.

### **2.2.1 Terminology**

The term ‘discourse marker’ may be seen as a ‘fuzzy concept’ (Jucker & Ziv, 1998, p. 2) because of the variety of definitions/labels applied by scholars. For instance, terms for

discourse markers include cue phrases (Knott & Dale, 1994), cue words (Horn et al, 2001), discourse connectives (Blakemore, 1987; Blakemore, 1992), discourse operators (Redeker, 1990; Redeker, 1991), discourse particles (Aijmer, 2002; Hansen, 1997; Hansen, 1998; Schourup, 1986), pragmatic connectives (Stubbs, 1983; Van Dijk, 1979), pragmatic markers (Anderson, 2001; Brinton, 1996; Fraser, 1996), and discourse markers (Fraser, 2006; Müller, 2005; Schiffrin, 1987). As the latter seems to be the term most widely understood and used, it is that used in this study. In fact, I agree with Schourup (1999) who states that ‘the term DM used in this review is merely the most popular of a host of competing terms used with partially overlapping reference’ (p. 228). In my opinion, there is another reason why the term discourse marker is most appropriate for describing these elements, namely that it can be compared to the term ‘landmark’, defined as ‘An object or feature [...] that is easily seen and recognized from a distance, especially one that enables someone to establish their location’ (Stevenson, 2010, p. 992), which is exactly what discourse markers do in conversation. They act as signs provided by speakers for listeners to recognize and – based on the function discourse markers carry – they can take a position in the conversation, e.g. turn-exchange (cf. Fischer, 2000; Levinson, 1983; Schiffrin, 1987).

### **2.2.2 Characteristics of discourse markers**

Regardless of the different terms applied to discourse markers, they share some specific defining characteristics. These include: syntactic independence from the sentence structure, initial position within the clause, having vague or no referential meaning, and phonological reduction, i.e. shortened and produced with ‘articulatory undershoot’<sup>7</sup> (Brinton, 1996, pp. 33–35; Schiffrin, 1987, p. 328).

---

<sup>7</sup> cf. Gahl et al (2012).



The following quotation from Schiffrin (1987) introduces two prominent characteristics of discourse markers. She states:

‘Although markers often precede sentences, [...] they are independent of sentential structure. Removal of a marker from its sentence initial position, in other words, leaves the sentence structure intact. Furthermore, several markers – *y’know, I mean, oh, like* – can occur quite freely within a sentence at locations which are very difficult to define syntactically.’ (Schiffrin, 1987, p. 31 et seq.)

Discourse markers such as *well, you know, right, ok, I mean*, etc., are typically used to ‘signal transitions, [...] index the relation of an utterance to the preceding context and indicate an interactive relationship between speaker, hearer and message’ (Fung & Carter, 2007, p. 411). Hölker (1991) provides four features defining discourse markers (or pragmatic markers, as he calls them). Semantically, they neither affect the truth conditions of an utterance nor do they add anything to the propositional content of an utterance. Pragmatically, they are related to the speech situation (i.e. the context of the utterance) rather than the situation discussed. They can also be defined functionally as they tend to have an emotive, expressive function rather than a referential, denotative, or cognitive function.

Brinton (1996) adds some important characteristics not discussed by Schiffrin (1987), stating that discourse markers:

- are typically used in oral discourse rather than written;
- form a separate tone unit;
- are syntactically optional;
- are difficult to place within a traditional grammatical word class;
- are multifunctional. (1996, pp. 33–35)

In the following sub-sections, some of the key characteristics of discourse markers are discussed in greater detail.

### *Syntactic position*

That discourse markers often occur sentence initially is considered to be one of their ‘most noticeable features’ (Lenk, 1998, p. 51). For some researchers, such as Fraser (1990; 1996), Redeker (1991), Hansen (1997) and Schiffrin (2001), this feature is one of the defining characteristics of this class of linguistic expressions. However, it has been argued that this is not necessarily always the case because some of these expressions may also appear sentence medially and finally (Andersen, 2001; Fung & Carter, 2007; Müller, 2005; Romaine & Lange 1991). Schourup (1999) argues that many discourse markers can appear ‘parenthetically within clauses with functions fundamentally identical to those they serve initially’ (p. 233). For example, we may have either *after all* introducing the sentence as in (1a) or falling within the sentence boundaries as in (1b):

(1)

- a) After all, Corgis are an intelligent breed.
- b) Corgis, after all, are an intelligent breed.

(Schourup, 1999, p. 233, examples 7 and 8)

Although occurring sentence initially seems to be the norm for discourse markers, it is still perfectly possible to use them sentence medially or finally. In these cases, the function may be different from that of the sentence initial position. Schiffrin (1987) demonstrates this with the discourse marker *y’know*: when it appears sentence initially, it precedes/introduces new information; in sentence final position, it follows old information (Schiffrin, 1987). Other different functions for *y’know* are suggested by Crystal (1988). He states that when the

discourse marker *y'know* appears at the beginning of a sentence, it softens the force of the utterance, as in the warning in example (2a), or it clarifies or amplifies the meaning of what one has just said when it comes in the middle, as in (2b), or it acts as a kind of tag question or as a check that the listener understands what is being said, as in (2c):

(2)

- a) *Y'know*, you should be more careful.
- b) He's just got a new BMX – *you know*, one of those tough little bikes ...
- c) He's bought a BMX – *you know*?

(Crystal, 1988, p. 47)

Regarding the functions of *you know* suggested by Crystal (1988), it would seem that his interpretation of example (2c) is questionable in identifying it as a tag question. The term 'tag question' as defined in the Routledge Dictionary of Language and Linguistics (2006) is 'Short question added to a statement which requests assurance or affirmation regarding what is expressed in the main clause: isn't it? [...]' (2006, p. 1174). Thus, if '*you know*?' in example (2c) does comprise a question, I would suggest that it is a short form of '*do you know*?' and that it is therefore, arguably, not a discourse marker. Rather, the function in Crystal's example is that of 'a check that the listener is understanding what is being said' (1988, p. 47).

However, further research is needed in this area to explain how meaning is affected by the position in which discourse markers may occur. Meanwhile, this study will consider discourse markers occurring in all three positions: sentence initial, medial and final.

### ***Lack of a single word class***

In addition to the disagreement on terminology, researchers have not yet reached a consensus on what items the class of discourse markers should include (Brinton, 1996, p. 31; Jucker, 1993, p. 436; Lenk, 1998, p. 39). For example, if we look at the list of discourse markers given by Schiffrin (1987; 2001), we find that it includes *oh, well, but, and, or, so, because, now, then, I mean* and *y'know*. On the other hand, if we look at another study by Lenk (1998), we find that it includes *anyway, actually, however, incidentally, still* and *what else*. This difference in terms of the inclusion or exclusion of terms is essentially based on varying definitions of what constitutes a discourse marker. Another example is that of Fraser (1990, p. 392) who excludes two of Schiffrin's (1987) expressions – *y'know* and *I mean* – because for him they are content words that are not relational, whereas discourse markers should signal discourse relations and have no content meaning. However, regardless of any disagreement on issues such as terminology, features and class membership, most researchers seem to agree on one property, which is their function of connecting units of discourse (Blakemore, 1987, p. 125; Biber, 2006, p. 66; Fraser, 1990, p. 387; Fraser, 1998, p. 21; Fung & Carter, 2007, p. 411; Goldberg, 1980, p. 141; Redeker, 1990, p. 372; Redeker, 1991, p. 1168; Schiffrin, 1987, p. 31).

As mentioned earlier, according to Brinton (1996), discourse markers are difficult to place within a traditional grammatical word class. Siepmann (2005, p. 44) argues that there is no morphosyntactic or semantic criteria through which discourse markers can be delimited as a word class in the traditional sense. Schiffrin (1987, p. 64) shows that discourse markers are syntactically diverse. For instance, they can be classified as adverbs (e.g. *now, then*), verbs (e.g. *say, like*), coordinating conjunctions (e.g. *and, but, however*), interjections (e.g. *well, oh*), or can include entire phrases and clauses (e.g. *you know, I mean*). Table 2.1 below presents an inventory of discourse markers in modern English which, according to Brinton

(1996), has attained partial consensus because the discourse markers included have received detailed scholarly attention.

**Table 2.1 Inventory of discourse markers in modern English**

<i>Ah</i>	<i>if</i>	<i>right/all right/that's right</i>
<i>Actually</i>	<i>i mean/think</i>	<i>So</i>
<i>after all</i>	<i>just</i>	<i>Say</i>
<i>Almost</i>	<i>like</i>	<i>sort of/kind of</i>
<i>and</i>	<i>mind you</i>	<i>Then</i>
<i>and {stuff, things} like that</i>	<i>moreover</i>	<i>Therefore</i>
<i>Anyway</i>	<i>now</i>	<i>uh huh</i>
<i>basically</i>	<i>oh</i>	<i>Well</i>
<i>Because</i>	<i>o.k.</i>	<i>yes/no</i>
<i>But</i>	<i>or</i>	<i>you know (y'know)</i>
<i>go "say"</i>	<i>really</i>	<i>you see</i>

(Brinton, 1996, p. 32)

### ***Orality***

In terms of style, the use of discourse markers is generally agreed to be a feature of oral discourse and important in colloquial speech (Brinton, 1996; Sankoff et al, 1997; Schiffrin, 1987; Schiffrin, 2001). Thus, the ability to use and apply them appropriately is undoubtedly an important aspect of sociolinguistic and intercultural communicative competence. Regarding communicative competence, Celce-Murcia et al (1995) point out that the first model of communicative competence was proposed by Canale and Swain (1980). Celce-Murcia et al (1995) consider that there are four components of communicative competence.

The first component is *grammatical competence*, which they simply put as knowledge of the language code. This component is the corner stone of communicative competence as without it one does not have communicative competence at all. The second component is *sociolinguistic competence*, which refers to how the language should operate socially, i.e. to understand the social context in which language is used (Alptekin, 2002, p. 58). The third component is *discourse competence*, which can be interpreted as the ability to incorporate the language in context. The last component is *strategic competence*, which means the ability to adapt language in an authentic or real communicative situation. Of course, this model has received some criticism (cf. Schachter, 1990) but this is not the place to discuss the issue in detail. In this study, I adopt the same line as Alptekin (2002) that communicative competence is ‘essential in order for foreign language learners to participate fully in the target language culture’ (p. 58).

The nature of the relationship between spoken and written language has been the subject of considerable research (Biber & Conrad, 2009; Brown & Yule, 1983; Chafe & Danielewicz, 1987). However, since the 1980s there appears to have been increasing acceptance that the relationship is not absolute opposition, but rather a scale or continuum ranging from casual conversation at one end to the most formal (official or academic) types of prose at the other (see e.g. Baron, 2000; Biber & Conrad, 2009; Chafe & Danielewicz, 1987; Leech, 1982). Spoken styles are characterized by a higher level of interactiveness and a lower level of planning and editing. Conversely, written styles are characterized by a less interactive, more impersonal style, with more focus on presenting or discussing information and higher levels of planning. Although there are cases of overlap within this continuum (e.g. a political speech is likely to be more carefully edited and impersonal than an office memo), the distributional differences of many linguistic features between speech and writing are quite strong (as shown

in corpus studies such as that of Biber et al, 1999). This is particularly the case for discourse markers. While there are a few discourse markers, notably *moreover* and *nonetheless*, which are restricted to more formal, written formal discourse, it is very clear, as Brown and Yule (1983) and Carter et al (2000) note, that the vast majority of discourse markers are limited to spoken discourse types. Everyday speech contains a large number of prefabricated fillers such as *well*, *erm*, *you know* etc. that have a range of functions, such as giving the speaker time to plan their contribution. This is an important factor in my decision to focus on spoken – not written – data for my investigation of discourse markers by Saudi learners.

Discourse markers are generally considered to be an important component in the organization of native speaker discourse (Carter & McCarthy, 2006; Hellermann & Vergun, 2007). Indeed, Aijmer (2002) states:

‘Discourse particles seem to be dispensable elements functioning as signposts in the communication facilitating the hearer’s interpretation of the utterance on the basis of various contextual cues. This does not mean that discourse particles are meaningless decorations or a verbal “crutch” in discourse indicating a lack of speaker proficiency, but they are better dealt with in pragmatics or in discourse analysis than in semantics.’  
(p. 2)

Weydt (2006) considers that those speakers who use discourse markers are perceived to be friendly and sociable, whereas speech without discourse markers sounds strange (p. 208). Crystal (1988) thinks of discourse markers as ‘the oil which helps us perform the complex task of spontaneous speech production and interaction smoothly and efficiently’ (p. 48). Appropriate usage of discourse markers is one of the criteria for a non-native speaker (NNS) to sound native-like (Hlavac, 2006, p. 1872). It has additionally been claimed that high

frequency of usage of discourse markers is ‘the hallmark of the fluent speaker’ (Sankoff et al, 1997, p. 191) and that underuse of discourse markers renders speech unnatural or non-native-like, as indeed does overuse (Müller, 2005, p. 13; Siepmann, 2005, p. 245). As Müller herself admits (2005, p. 26), repeated use of discourse markers can be experienced as ‘irritating’ and this applies to native speakers (NSs) and NNSs alike.

It is relevant here to discuss briefly what is meant by pragmatic fluency as this may offer some way of evaluating the appropriateness of talk by NNSs. As used by House (1996, p. 228), the term *pragmatic fluency* refers to a dialogic ability which combines pragmatic appropriateness and holding smooth ongoing talk. For NNSs to be considered pragmatically fluent, it is essential that they meet the expectations of NSs and that their speech represents acceptable language behaviour (House, 1996). He suggests assessing the following:

1. Use of routinized pragmatic phenomena focused on in the instructional treatment, namely, gambits, discourse strategies, and speech acts (for a definition, explanation, and exemplification of these analytical categories).
2. Ability to initiate topics and topic changes using appropriate routines.
3. Ability to ‘carry weight’ in a conversation manifest in substantive comments.
4. Ability to show appropriate uptaking, as well as replying and responding behavior [...], anticipation of end-of-turns as evident in latching, overlapping, and the use of appropriate ‘second pair parts’ in routinized reciprocation.<sup>8</sup>
5. Appropriate rate of speech, types of filled and unfilled pauses, and frequency and function of repairs.

(House, 1996, p. 229)

Even though the Louvain International Database of Spoken English Interlanguage (LINDSEI) considers learners of English to be advanced based on the external criterion that they are third

---

<sup>8</sup> See Goffman (1981).



and fourth year students, I argue that this is not sufficient. In determining who comprised advanced learners of English and who did not when I collected my data, I also took into consideration House's (1996) areas of assessments.

### ***Phonological features***

Östman (1982, p. 149; 1995) and Brinton (1996) argue that discourse markers are short items and always phonologically reduced. Schiffrin (1987) puts it in another way. She points out that they prototypically have a range of prosodic contours/clues such as 'tonic stress and followed by a pause, phonological reduction' (p. 328). (cf. Fung and Carter, 2007). This feature has been acknowledged by other researchers (see Aijmer, 2002; Müller, 2005; Traugott, 1995). Similar to these researchers, Andersen (2001) starts his definition of discourse markers as 'a class of short, recurrent linguistic items that generally have little lexical import but serve significant pragmatic functions in conversation' (p. 39).

### ***Optionality***

Optionality is another feature that most studies have attributed to discourse markers. For example, Schiffrin (1987, p. 64) points out that discourse markers are never obligatory, i.e. 'any utterance preceded by a marker may also have occurred without that marker' (Schiffrin, 2001, p. 57). Similarly, Schourup (1999) considers discourse markers to be syntactically and semantically optional. Fraser (1988, p. 22) points out that the absence of discourse markers does not render the sentence ungrammatical or unintelligible even if it does remove a powerful clue between the current utterance and the prior discourse.

To illustrate what is meant by optionality, I here employ two examples of the discourse marker *like* from my data. The word *like* in example (3) below cannot be omitted without

causing ungrammaticality; however, in example (4), *like* can be omitted without affecting the utterance syntactically or semantically.

(3) <A> we should talk to you about this thank you very much </A>

<B> obviously the= these people are a couple and he is drawing her and she is not very pleased with the drawing </B>

<A> (uhu) </A>

<B> a= and then he changed the drawing completely and the drawing doesn't even look *like* her and she is very please and telling her friend that my my my husband or whoever he is has </B>

(SLC-AR04)

(4) <A> oh and (erm) .. is it very different from living on campus when you live in town what do you prefer </A>

<B> (erm) </B>

<A> . if you have any preferences <laughs> </A>

<B> it it is it is quite nice living off campus because erm .. you don't feel erm . stuck on campus all the time it can get a little bit claustrophobic I found . last year </B>

<A> mhm </A>

<B> and you you go home to a different place at the[i:] end of the day . but there are disadvantages <begin\_laughter> *like* it's more <end\_laughter> expensive </B>

<A> yeah you have to take the bus .. I guess </A>

(LOCNEC-E01)

However, the optionality of discourse markers does not undervalue their communicative importance. Brinton (1996) argues that 'they are not pragmatically optional or superfluous:

but they serve a variety of pragmatic functions' (p. 35). She adds that when discourse markers are missing from discourse, even though the discourse will remain grammatical, it will be judged 'unnatural', 'awkward', 'disjointed', 'impolite', 'unfriendly', or 'dogmatic' within the communicative context (pp. 35–36). Furthermore, Fraser (1990) points out that the absence of discourse markers may increase the chances of 'communicative breakdown' (p. 390). The previous examples seem to support this view.

### ***Multi-functionality***

Before explaining what is meant by multi-functionality, it is worth highlighting what Jucker and Ziv (1998a) say about it in relation to discourse markers. They state:

'Whether a specific linguistic element is monofunctional or polyfunctional is not a useful criterion in deciding whether it is a discourse marker or not because of the obvious analytical vicious circularity it entails. Many studies actually set out to argue explicitly for the monofunctionality or polyfunctionality of specific markers, thus nullifying this as a valid criterion.' (Jucker & Ziv, 1998a, p. 4)

However, Müller (2005) argues that even though Jucker and Ziv (1998) may be correct in objecting to multi-functionality as a defining criterion, research has shown that many discourse markers fulfil more than one function or at least have sub-functions.

Fraser (1988) examines the function of the discourse marker *so* in six utterances and concludes that '*so* as a discourse marker permits a wide range of interpretations, all of which arguably emerge from a core sense. Starting with this core meaning, the specific interpretation [...] in a given instance is the result of enriching this general signal in light of the details of the particular discourse context' (p. 23).

Brinton (1996) contends that discourse markers are multifunctional in the sense that they can operate at local and global levels simultaneously and also textually and interpersonally within the pragmatic component. Similarly, Schiffrin (1987) points out that for an element to be a discourse marker, it should work both locally and globally. This was what Hansen (1998) noted when she defined discourse markers as ‘non-propositional linguistic items whose primary function is connective, and whose scope is variable’ (p. 73). What she means by variable scope is that ‘the discourse segment hosting a marker may be of almost any size or form, from an intonational pattern indicating illocutionary function [...] through subsentential utterances [...] to a segment comprising several utterances [...]’ (1998, p. 73).

Under the functional properties of discourse markers, Aijmer (1996) discusses global and local discourse markers. She points out that the distinction between them is considered basic in that:

‘Global discourse markers comment on the relationship between larger discourse units. They accompany the introduction of new topics, serve a role in turn-taking and segment discourse into larger units; they mark the opening of conversation, order points in a discussion sequentially, etc. Local discourse markers signal intersentential connections and comment on the expectedness or validity of the new message; they underline its importance or simply facilitate the transition from one turn to another.’ (1996, p. 222)

To clarify what is meant by local and global discourse markers, Aijmer (1996) gives some examples of both. Regarding local discourse markers, as in the following example, she explains that the discourse marker *actually* helps a speaker to clarify or sharpen a point he/she has recently made. In her treatment of *actually*, she argues that it marks whether the following proposition is in line with expectations:

(5) A so I went to the school of Applied Linguistics

In Winchester

B m

A *actually*

I applied to British Council

and failed to get in

(Aijmer, 1996, p. 223, example 33)

Regarding global discourse markers, Aijmer (1996) argues that they help the speaker to guide the hearer to global aspects of the utterance and to solve problems in communication. Globally, discourse markers are helpful when the speaker '[makes] corrections, adds a point to the discussion, inserts a comment or a digression, summarizes information, recapitulates or clarifies a point' (Aijmer, 1996, p. 226). For instance, the discourse marker *briefly* helps the speaker to summarize information globally, exemplified as follows:

(6) A now *briefly*

These four members of staff at the moment

Three of us

Have ongoing research

(Aijmer, 1996, p. 227, example 50)

To summarize, even though canonical discourse markers would typically exhibit the six characteristics above, this is not always necessary. For the purpose of this study, I consider the discourse markers of this study to:

- be syntactically optional,
- occur at the beginning, middle, or end of the discourse unit,
- have little or no propositional meaning,

- be multifunctional,
- occur more in oral than in written discourse.

### **2.3 Some specific functions of discourse markers**

The literature on discourse markers has acknowledged different functions based on the frameworks used to investigate them (e.g. coherence-based, relevance-theoretic accounts, etc.). Aijmer (1996) suggests that discourse markers ‘function as cues or guides to the hearer’s interpretation’ (p. 210). Schiffrin (2001), however, argues that discourse markers can indicate semantic, pragmatic and social relations. The primary function of discourse markers for Östman (1981) is that they ‘implicitly anchor the utterance in which they function, to the speaker’s attitudes towards aspects of the on-going interaction’ (p. 5). In another paper, she argues that the function of discourse markers is tied to culture and the society. She states that discourse markers anchor utterances ‘to the communicative restraints of a culture and society’ (1995, p. 100) such as politeness, discourse coherence or involvement (Aijmer & Simon-Vandenberg, 2011). It can be noted that Östman’s (1982) description of the functions of discourse markers is rather broad.

Briton (1990; 1996) provides a more specific list of functions which also incorporates the most commonly recognized functions of discourse markers. Following Halliday and Hasan’s (1976, pp. 26–28) functional-semantic components, Brinton (1996) classifies the functions into textual and interpersonal or interactional functions as Müller (2005) calls it.

**Table 2.2 Discourse marker functions**

<b>Textual Functions</b>	<b>Interpersonal Functions</b>
a) initiate and close discourse, b) mark a boundary in discourse, c) serve as a filler or delaying tactic, d) aid the speaker in holding the floor, e) repairs one's own or others' discourse, f) bracket the discourse either cataphorically or anaphorically, g) mark either foregrounded or backgrounded information.	h) preface a response or a reaction, i) effect an interaction or sharing between speaker and hearer.

(Brinton, 1990; Brinton, 1996; Brinton, 1998)

According to Brinton (1996), as Table 2.2 illustrates, the functions (a–g) are part of the textual functions of discourse markers, whereas the second category (h–i) forms part of the interpersonal functions. Textually, according to Brinton (1996), the need to initiate and close discourse, to mark topic shifts, to indicate new and old information and to constrain the relevance of adjoining utterances are part of the textual functions of discourse markers. In the textual mode of language, the use of discourse markers helps the speaker to structure meaning as text, create passages of discourse cohesively and relate language to context. Interpersonally, discourse markers are considered as a means through which speakers can express their attitudes, evaluations, judgments, expectations and demands, as well as the nature of the social exchange, the role of the speaker and the role assigned to the hearer (Brinton, 1996). For Jucker and Ziv (1998, p. 4), when discourse markers have been analysed they have been examined as text-structuring devices, as modality or attitudinal indicators, as markers of speaker–hearer intentions and relationships, and finally as instructions on how given utterances are to be processed. In her paper *Pragmatic Markers in Spoken Interlanguage*, Aijmer (2004) ascribes some tactical functions to discourse markers or what she calls ‘pragmatic markers’. She says that they have the function of ‘checking that the

participants are on the same wavelength or of creating a space for planning what to say, making revisions, etc.’ (p. 177). Several of these functions will be referred to later in the thesis.

## **2.4 Major approaches to discourse markers**

Fischer (1998) avers that discourse markers ‘have been analysed in many different frameworks and from different perspectives with often highly different results’ (p. 113). Müller (2005) argues that coherence and relevance theory are the two basic frameworks which have been employed to look at discourse markers. There have been a number of studies which have adopted those frameworks. For example, in coherence-based studies, such as those of Schiffrin (2001), Lenk (1998) and Halliday (2004), discourse markers are described as signals of coherence relations that link textual units (Schourup, 1999, p. 240). However, while Schiffrin’s (2001) analysis focuses on the functions of discourse markers in constructing local coherence between immediately adjacent units in text, Lenk (1998) analyses these expressions in terms of their role in coherence at the global level of text (p. 28). Lenk (1998) argues that discourse markers bring coherence at a more global level within the discourse, for example, linking to earlier topics, topics to be followed, or knowledge outside the context of the conversation. She points out that coherence in conversations depends not only on what Schiffrin (2001) refers to as relationships between two immediately adjacent items but also on global relationships linked by discourse markers for facilitating the interlocutors’ process of establishing coherence.

On the other hand, there have been also a number of studies that have adopted relevance theory (for example, Andersen, 2001; Ariel, 1998; Blakemore, 1988; Blass, 1990; Jucker, 1988). Working within the theory proposed by Sperber and Wilson (1986), Blakemore (1987;



2001) argues that these expressions are employed to signal to the receiver how one discourse segment is relevant to another. Unlike the coherence-based approaches, which are concerned with discourse and the relations between its segments, the object of study in relevance theory is not discourse itself, but the ‘cognitive processes underlying successful linguistic communication’ (Blakemore, 2002, p. 5).

The literature on discourse markers shows the diversity of the approaches taken in studying these items. There have been a number of attempts to provide a definition of this pragmatic category. While the coherence framework is said to focus more on the textual functions (e.g. Andersen et al, 1999, p. 1339), relevance theory focuses on cognitive processes (see Sperber & Wilson, 1986). In the following sub-section, I provide an overview of these two important frameworks in the analysis of discourse markers.

#### **2.4.1 Discourse markers in the coherence-based approach**

Schiffrin’s (1987) influential study of English discourse markers is one of the most important detailed studies that has addressed these linguistic items. Her analysis included 11 English discourse markers (*oh, well, but, and, or, so, because, now, then, I mean and y’know*) as they appeared in a corpus of sociolinguistic interviews with Jewish Americans.

Schiffrin (1987) defines discourse markers as ‘sequentially dependent elements which bracket units of talk’ (p. 31). This definition introduces two significant aspects of discourse markers. The first significant aspect refers to the fact that discourse markers are elements that work at the level of discourse and consequently are dependent on the sequence of discourse. Schiffrin (1987) argues that this sequential dependence can be seen where discourse markers combine two units which do not belong to the same syntactical category, as in the following

example in which the discourse marker *and* links a declarative sentence with an interrogative sentence:

(7)

Debby: I don't like that

Zelda: I don't like that. And, is he accepting it?

(Schiffrin, 1987, p. 38, example 8)

As for bracketing, Schiffrin (1987) claims that she deliberately chooses the term 'units of talk' to avoid excluding cases in which markers appear in non-traditionally defined units, such as ideas, turns, speech acts or tone units. Thus, she states that:

'a unit which focuses on how linguistic structure, meaning, and act are phonologically realized in speech might seem to be a more promising basis for our definition of markers.' (Schiffrin, 1987, p. 33)

This broad definition allows Schiffrin to 'use presence of a marker as an indicator of some yet undiscovered unit of talk' (1987, p. 36). Based on her analysis of English discourse markers in unstructured interviews, Schiffrin (1987) proposes a model of discourse coherence consisting of five planes of analysis: exchange, action, ideational structures, information state and participation framework, which collectively she refers to as a Model of Discourse Coherence. Below is a summary of Schiffrin's (1987, p. 316) categories of markers used on each of the planes of discourse. Table 2.3 illustrates that discourse markers are multi-functional as certain markers are used more on one plane of discourse than on the other. Each column in the table represents the different structures of the discourse.

**Table 2.3 Planes of discourse marker functions**

<b>Information state</b>	<b>Participation framework</b>	<b>Ideational structure</b>	<b>Action structure</b>	<b>Exchange structure</b>
oh	oh		oh	
well	<b>well</b>	well	well	well
		<b>and</b>	and	and
		<b>but</b>	but	but
		<b>or</b>		or
So	so	<b>so</b>	so	so
because		<b>because</b>	because	
	now	<b>now</b>		
them		<b>then</b>	then	
I mean	<b>I mean</b>	I mean		
<b>you know</b>	you know	you know		you know

Note: the primary plane is represented in bold font

(Schiffrin, 1987, p. 316)

The ‘exchange structure’ indicates the sequence of conversational roles and how turn changes interrelate. The ‘action structure’ shows how different speech acts of speakers are sequenced and determined, reflecting participants’ identities, social factors and actions. The ‘ideational structure’ differs from the exchange structure and the action structure, which are pragmatic, in that it consists of semantic units which Schiffrin calls ‘ideas’. It reflects different semantic relationships among ideas within the discourse. This plane includes three kinds of relations: cohesive, topical and functional (Schiffrin, 1987, p. 26). The ‘participation framework’ indicates the different ways in which speakers relate to each other. Finally, there is the ‘information state’ which is concerned with what the speaker and the hearer share in terms of cognitive capacities; according to Schiffrin (1987), this plane ‘involves the organization and management of knowledge and meta-knowledge’ (p. 28). Fraser (1999) cites these planes with some adaptations:

**Exchange Structure**, which reflects the mechanics of the conversational interchange (ethnomethodology) 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, including cohesive relations, topic relations, and functional relations;

**Participation Framework**, which reflects the ways in which the speakers and hearers can relate to one another as well as orientation toward utterances; and

**Information State**, which reflects the ongoing organization and management of knowledge and metaknowledge as it evolves over the course of the discourse.

(Fraser, 1999, p. 934)

Schiffrin (1987) makes the point that discourse markers can function at different levels of discourse structure (linguistic or non-linguistic). They can operate at the ‘ideational’ (informational) structural level in the sense that they indicate relations between ideas in discourse or, in other words, they mark the organization of ideas in discourse. For instance, a discourse marker such as *but* indicates that what follows it contrasts with what precedes it. They can also operate in relation to the participation framework (discourse exchange and interaction) in the sense that they play a role in controlling the conversational labour between speakers and hearers as is the case with *oh* and *well*.

However, these coherence planes have not gone unchallenged. Redeker (1991) criticizes Schiffrin’s (1987) coherence planes and proposes an alternative, suggesting a revised model of discourse coherence. This model includes three components: ideational structure, rhetorical structure and sequential structure. These three components suggested by Redeker are roughly equivalent to Schiffrin’s ideational and action structures and ‘an extended variant

of [Schiffrin's] exchange structure' (Redeker, 1991, p. 1167). Redeker's sequential structure is different from Schiffrin's exchange structure in that it can account for sequential transitions in monologic talk as well as dialogic discourse. Another difference between the two models, as Redeker claims, is that Schiffrin defines the planes of her coherence model in relation to the markers' functions and does not provide an independent definition of each plane (Redeker 1991, pp. 1167–1168), whereas Redeker defines the components of her model irrespective of the functions of the discourse markers within them.

However, I agree with Müller (2005) that Redeker's (1991) model does not include all the functions of discourse markers and does not define them precisely. Müller states:

'[...] I do not find her model more convincing. It does not accommodate all functions of discourse markers either (for example in turn-transitions or the negotiation of interpersonal relationships) and does not seem to be any more precise in its definitions than Schiffrin's model.' (2005, p. 30)

Redeker (1991) disagrees with Schiffrin's (1987) conclusion that only two discourse markers *so* and *well* can work on all planes of talk as shown in Table 2.3 above. Redeker (1991) argues that if Schiffrin's model is applied to a wider range of data, such as narrative and spontaneous talk, nine out of the 11 markers can actually function on all five planes of discourse. Thus, Redeker (1991) states:

'I will argue that DS's [Schiffrin's] minimalist approach to the semantics of discourse markers places too heavy a burden on the syntactic and contextual determination of marker meanings. If the coherence model is applied consistently to a slightly wider range of talk, the model ceases to discriminate between the markers.' (p. 1139)

Furthermore, Redeker (1991) criticizes the ‘vaguenesses and inconsistencies of theoretical concepts and terminology’ as well as the failure of Schiffrin’s work to provide clear and reliable quantitative information concerning the discourse markers investigated in her study (Redeker, 1991, pp. 1160–1161). In this study, I address Redeker’s concerns by providing frequencies of the use of discourse markers by different categories of speaker.

In addition, Redeker (1991) believes that two of the planes suggested by Schiffrin (information structure and participation framework) do not directly contribute to discourse coherence. Redeker states:

‘[...] the components information structure and participation framework are clearly not on par with the other three planes. The cognitions and attitudes composing these two components concern individual utterances, while the building blocks at the other three planes are relational concepts’ (Redeker, 1991, p. 1162)

#### **2.4.2 Discourse markers in the relevance-based framework**

The second main approach that has been used to study discourse markers is relevance theory (see Sperber & Wilson, 1986). Sperber and Wilson (1986) consider relevance theory a general theory of communication based on cognitive principles. Anderson (1998) points out that in relevance theory utterances are seen as ‘inputs to inferential processes which affect the cognitive environment of the hearer’ (p. 150). Unlike the coherence approach, which focuses on the textual functions of the discourse markers in creating coherent discourse, relevance theory stresses the cognitive processes involved in interpretations of utterances and the role of discourse markers in constraining these interpretations. Blakemore (1987; 1992; 1998; 2002; 2004) is one of the pioneering scholars who has analysed discourse markers within the relevance theory framework. Blakemore (1987) argues that the role of discourse markers is to

constrain ‘the interpretation of the utterance that contain them by virtue of the inferential connections they express’ (1987, p. 105). Blakemore (1992) further contends that the relevance-theoretic account allows a more specific claim to be made about linguistic structure and pragmatic interpretation. There are three ways in which information conveyed by an utterance can be relevant:

- 1) It may allow the derivation of a contextual implication
- 2) It may strengthen an existing assumption<sup>9</sup> (by providing better evidence for it)
- 3) It may contradict an existing assumption

(Blakemore, 1992, p. 138)

In view of this, as the following will demonstrate, three types of discourse connectives have been identified (Blakemore, 1992, pp. 138–142), each of which is discussed and illustrated using an example below:

- a) Discourse connectives which introduce contextual implications (*so, therefore*). For example:

(8) A: Your clothes smell of perfume.

B: So (what)?

(Blakemore, 1992, p. 139, example 20)

In this example, Blakemore (1992) assumes that speaker B is asking speaker A rhetorically what conclusion he is expected to draw on the basis of A’s utterance.

---

<sup>9</sup> Assumptions, as defined by Sperber and Wilson (1986), are ‘thoughts treated by the individual as representations of the actual world’ (p. 2).

b) Discourse connectives concerned with strengthening (*after all, besides, moreover, furthermore, utterance-initial also, indeed*). For example:

(9) A: Will you make pancake?

B: I haven't really got time tonight. Besides there's no milk.

(Blakemore, 1992, p. 140, example 25)

The aim of speaker B using *besides* is to strengthen his/her own commitment to the assumption conveyed in the first utterance that the request '*Will you make pancake?*' will be met.

c) Discourse connectives which introduce denials (*however, still, nevertheless, but*). For example:

(10) David is here. However, you can't see him.

(Blakemore, 1992, p. 141, example 30)

Blakemore (1992) includes this example as a case in which the speaker must indicate that his utterance is relevant as a denial. Blakemore points out that in this example '*however* indicates that the proposition it introduces is inconsistent with a proposition that the speaker assumes the hearer has derived as a contextual implication from the first utterance' (Blakemore, 1992, p. 141).

After reviewing the two major approaches and the literature of studies based on them, Risselada and Spooren (1998) point out that discourse markers have most commonly been studied within the coherence approach (for example, Fuller, 2003; Hansen, 1998; Hays, 1992;



Holmes, 1986; Lee-Wong, 2001; Lenk, 1995; Redeker, 1990; Schiffrin, 1987; Takahara, 1999). Risselada and Spooren (1998) state that:

‘[A] characteristic which most approaches to discourse markers have in common, and which forms also one of the recurrent themes in the contributions to this special issue, is the key role that are attributed to coherence relations.’ (p. 131)

One powerful reason for choosing a coherence approach to the analysis of discourse markers is that there has been a wide range of previous studies that have applied the approach very effectively to the analysis of authentic discourse data (Holmes, 1986; Knot & Dale, 1994; Schiffrin, 1987; Schiffrin, 2001), including some based on corpora (Lenk, 1995; Lenk, 1998; Risselada & Spooren, 1998; Schiffrin, 1987). The insights from these studies have helped me make sense of many of my own corpus examples by drawing attention to the use of discourse markers as coherence relations that link textual units. Of course, this does not mean that I agree with everything that previous coherence approach-based researchers have said. The relevance-based approach, although it has some intuitive appeal and has been around almost as long, has not been applied extensively to authentic spoken or written texts (a very rare exception is Anderson, 2001). For this reason, I find it less attractive than the coherence approach in terms of identifying the different functions and levels of usage of discourse markers in my own corpus data.

## **2.5 Discourse markers and language learning**

As earlier sections have demonstrated, discourse markers in the English language have received a considerable amount of attention over the last two decades. The use of discourse markers is widespread across many languages. If NNSs wish to sound like NSs, they have to know ‘how things are said’ and how to use ‘conventional expressions’ and standard ways of

talking (de Cock, 1999, p. 52; de Klerk, 2005, p. 1148). Thus, in what follows, I discuss some of the research on discourse markers used by NNSs. This will allow comparison of the results of this thesis to those found in previous research.

Sankoff et al (1997) investigated the use of discourse markers in English and French by Anglophone learners of French as a second language in Montreal. The result of their analysis demonstrated great variation not only in individual repertoires but also in the frequency of the use of discourse markers. They found that those who were more integrated into the local Francophone community exhibit more native-like use of discourse markers, especially those who had been exposed to French since childhood. Sankoff et al (1997) claim that a higher frequency of discourse marker use is ‘the hallmark of the fluent speaker’ (p. 191). They conclude that there is a correlation between the frequency of discourse marker use and the speakers’ knowledge of French grammar and more native-like control of discourse marker use in the second language (L2) second language. They state that ‘the more successful L2 speakers were those who could control native discourse markers in a native-like fashion’ (1997, p. 213). Thus it is worth considering the claimed implications of these findings in this study.

Similarly, Lee (2004), in a corpus-based study, examined the spoken English of Korean immigrants to the United States. The term corpus-based refers to a methodology in which the corpus is mainly used to test or refine existing hypotheses taken from other sources (Tognini-Bonelli, 2001; Baker, 2010). Age and immigrant generation were the variables Lee considered. With regard to gender, his findings showed that women do not use discourse markers more often than men. Lee (2004) found that the discourse marker *you know* was used with great frequency by men. These men were from the second generation immigrants,

considered NSs of English. However, as noted by House (2009), the results of this study challenge those of many previous studies (for example, Brinton, 1996; Holmes, 1986; Lakoff, 1973; Ostman, 1981) which found that women use discourse markers such as *like* more frequently than men.

A study by Gross and Salmons (2000) examined the historical background to the replacement of discourse markers in German dialects spoken in the United States with English discourse markers. They concluded that the borrowing of English discourse markers in German-American dialects was the result of code-switching. Their findings suggested that English discourse markers entered the German dialects first as emblematic switches (e.g. interjections, fillers, tags and idiomatic expressions).<sup>10</sup> After that, discourse markers became established borrowings. This borrowing prepared for partial or complete replacement of the whole set of German discourse markers. Gross and Salmons (2000) state that code-switching is one crucial mechanism which explains how ‘structural interference comes about’ (p. 471).

Trillo (2002) used a corpus-driven approach to examine the use of discourse markers in English by native and non-native Spanish children and adults. Unlike corpus-based approach, in corpus-driven approach a corpus is used to ‘support linguistic argument or to validate theoritical statement’ (Tognini-Bonelli, 2001, p. 84). Trillo (2002) concluded that native and non-native children show a similar pattern in their use of discourse markers, whereas non-native adults fossilize in their L2 pragmatic development due to the lack of discourse marker instruction. Based on his findings, he urged the teaching of discourse markers in the curriculum in Spain so that learners could attain pragmatic competence. This conclusion

---

<sup>10</sup> Emblematic switching is more commonly used by semi-bilinguals or those who speak one language significantly less fluently than the other.

highlights the significance of investigating Saudi textbooks to examine their potential impact on the use of discourse markers.

Other corpus-driven studies were undertaken by Müller (2004; 2005). Her studies compared the use of English discourse markers by German non-native speakers of English to American native speakers. She incorporated qualitative and quantitative analysis of the discourse markers *so*, *well*, *like* and *you know* as they appeared in the interaction between the German and American speakers. The analysis interestingly showed that both groups of speakers employed the four markers in their talk. However, the groups were different in terms of each marker's functional use. Some functions were never used by the Germans (e.g. the use of *you know* to introduce quoted material) whereas the Americans did use them. On the other hand, some functions (e.g. *well* to introduce a summary/conclusion and *well* to continue expressing an opinion) were only used by the German speakers (Müller, 2005). Thus, it was concluded in this study that NNSs and NSs are not alike. The implications of this study for my own research on Saudi learners of English are uncertain. While it is possible that Müller's findings will be replicated by those of this study and it is certainly worth comparing, it has to be remembered that German is closer to English than is Arabic.

Furthermore, Fuller (2003a) analysed the use of discourse markers by native and non-native German speakers of English in two different contexts: interviews and conversations. She included five discourse markers, *well*, *oh*, *y'know*, *like* and *I mean*, because they fit her two main criteria. The first criterion was optionality of the discourse marker in the utterance. Optionality in her study means 'if the DM is removed from the utterance, the semantic relationship between the elements they connect remains the same. [...] without the DM, the grammaticality of the utterance must still be intact' (Fuller, 2003a, p. 186). The findings of

her study showed that both NSs and NNSs used the discourse markers *oh* and *well* more frequently in the conversation data. The study also demonstrated that the NSs used *like* and *you know* more frequently in the interview data. However, NNSs did not show a difference in the use of *like* and *you know* across contexts. Fuller (2003a) reached the conclusion that there are differences between NSs and NNSs in the use of discourse markers in relation to speech context, i.e. in terms of register, or type of discourse and frequency rate.

Moreno (2001) examined the use of *well* as a discourse marker by Spanish students of English in interaction with NSs. NNSs were undergraduate students in their 3rd, 4th, or 5th year of an English Language and Literature course at the University of Seville, aged 21 to 25 years. The NSs were two American students from Wisconsin and an English teacher from London. The data analysis for the compiled corpus showed the low occurrence of discourse markers in the Spanish students' discourse. For example, students in the 3rd year did not use the discourse marker *well* at all and this was almost the case with the other years, 4 and 5. Besides, there were some students who used the marker *well* several times, but their usage was inappropriate (using a discourse marker that did not fit appropriately) in some cases. For example, one of the participants used the discourse marker *okay* where *well* should have been used instead.

Employing Schifffrin's (1987) discourse model, Hays (1992) examined the use of discourse markers in English interviews by Japanese students in Japan. The results of this study illustrated that most speakers demonstrated the ability to use *and*, *but* and *so*. Hays (1992) argues that the ability to use these particular discourse markers is mainly a) because of their crucial nature in developing ideas and b) the fact that they are usually explicitly taught (p.

29). However, for other markers on other planes, Hays states that they ‘would not be expected unless there has been exposure to that discourse community’ (1992, p. 29).

Hasselgren (2002) investigated discourse markers or what she refers to as ‘small words’. In her study, she compared quantitatively and qualitatively the use of lexical items such as *well* and *you know* used by Norwegian learners of English to that of NSs of English. Her main hypotheses aimed to establish the role of discourse markers in contributing to learner fluency. Her results showed that the more fluent her participants the closer they were to the NSs.

Fung and Carter (2007) compared the production of discourse markers by NSs from a corpus of spoken British English with NNSs from a corpus of classroom discourse in Hong Kong. They found a considerable discrepancy in the use of discourse markers between NSs and learners. NNSs used discourse markers at a very restricted level and with limited functions.

Regarding discourse markers in the Arabic language, Gaddafi (1999) discussed discourse markers in Libyan Arabic. His thesis focused on two discourse markers, *garaft* (equivalent to the English discourse marker *you know*) and *yaGni* (I mean). The purpose of his study was to identify some of the functional roles of discourse markers in informal conversational interaction. Following Schiffrin’s (1987) discourse model, Gaddafi (1999) concluded that *garaft* and *yaGni* performed a wide variety of functions on different planes of talk.

Saeed and Fareh (2006) investigated the problems translators and Arab learners of English encounter when translating Arabic sentences containing *fa* into English. After surveying several texts in Arabic newspapers and Arabic grammar references, five functions of discourse marker *fa* were identified: explanatory *fa*, consequential *fa*, causal *fa*, sequential *fa*

and adversative *fa*. They then designed a translation task to find out how this discourse marker could be translated into English. The study involved 50 English major Arab students from the University of Sharjah. The findings of this study showed that students found it more difficult to translate the explanatory and sequential *fa* than the consequential *fa*, the causal *fa* and the adversative *fa*.

Another study which compared Arabic discourse markers to English discourse markers was that of Kurdi (2008). In her PhD thesis, *The Use of Discourse Markers by Syrian Arabic Learners of English*, she investigated the use of the English discourse markers *so*, *you know* and *I mean* by Syrian Arabic speakers. The aim of her study was to examine the influence of L1 on the production of English discourse markers. To verify this, she compared the English discourse markers with their equivalents in the Arabic language. Her overall conclusion was that there was no apparent influence from Arabic on the use of the English discourse markers. However, this was not the case with the discourse marker *so*. Kurdi (2008) found that her informants' use of *so* as a marker of transitions was as a result of the influence of the Arabic marker *fa*, which exhibits a similar function.

Hussein (2009) studied the use of discourse markers in English and Arabic using relevance theory as a theoretical framework. He included 10 discourse markers, five used in Standard Arabic – *lakinna*, *bainama*, *lakin*, *bal* and *fa* – and the other five used in Syrian Arabic – *bass*, *la-heik*, *la-ha-sabab*, *ma'nāt-o* and *bi-ittal*. His reason for using these 10 discourse markers was that they can be compared and contrasted to Blakemore's (1987; 2002) two discourse markers *but* and *so*. Hussein found that as with the discourse marker *but* in English, the Arabic discourse marker *bass* encodes a general procedure which can mean denial of expectation, contrast, correction and cancellation. The other four discourse markers, *lakinna*,

*bainama*, *lakin* and *bal*, were analysed as lexical implementations of these functions. Concerning the discourse marker *fa*, similar to Kurdi's (2008) finding, Hussein found it functioned as sequentiality, immediacy, non-intervention and causality

In summary, the findings of all the studies reviewed above demonstrate that NNSs of English can use English discourse markers. In addition, these studies show some consensus that NNSs tend to use discourse markers less frequently than NSs. In some of these studies, there is evidence that the exposure to the target language has an influence on the mastery of the use of discourse markers. Thus, this review also highlights another important issue regarding the feasibility of teaching discourse markers.

## **2.6 The present study**

The role discourse markers play in communication has attracted considerable attention from linguists working with English discourse (for example, Blakemore, 1987; Fraser, 1990; Schiffrin, 1987; Schiffrin, 2001; Schourup, 1985). However, there have been few studies which have considered their importance in spoken discourse in an Arabic context. Gaddafi (1990), Ghobrial (1993), Al-khalil (2005) and Hussein (2009) have investigated Arabic discourse markers used by Arabic NSs in different dialects (Libyan, Egyptian and Syrian respectively). Kurdi (2008) analysed three discourse markers, *so*, *you know* and *I mean*, in the English discourse of Syrian Arabic learners but she compared them to their equivalents in Arabic not to the use by NSs of English. Although a limited number of studies have examined English discourse marker use by NNSs (Lee, 2004; Müller, 2005; Nikula, 1996; Özbek, 1995; Trillo, 2002), to the best of my knowledge, no study thus far has investigated how English discourse markers are used by Saudi speakers and for what purposes or how appropriately they use them relative to their use by NSs of English. Therefore, this study will attempt to fill this gap by making the first extensive examination of English discourse



markers in the speech of Saudi learners of English and comparing it to NS's use of discourse markers in English. In addition, this study will examine the representation of the discourse markers in Saudi textbooks to estimate what influence the teaching of English discourse markers could have on their use in authentic English.

## **CHAPTER 3: DATA DESCRIPTION AND RESEARCH METHODOLOGY**

### **3.1 Introduction**

This chapter presents the methodology of the study. Section 3.2 provides a detailed description of the corpora upon which the findings of the study are based. This section also presents the textbook database. In section 3.3 research methodologies relevant to this study are addressed. It also presents the analytical tools and the procedures for analysing the data are discussed.

### **3.2 Data description**

The data employed in this study are based on recorded interviews and learner profile questionnaires. In what follows, these tools are described. The main findings regarding the use of discourse markers are generated from the Saudi learner corpus (SLC) and LOCNEC corpora. As mentioned earlier, LOCNEC is used in this study as a reference corpus with which the results from the SLC are compared. The results concerning the functions and frequency of the discourse markers under investigation are interpreted through the examination of a database of textbooks. The reason why textbooks are examined is to compare and contrast authentic speech and teaching materials in Saudi Arabia.

#### **3.2.1 Corpora**

This study uses the methodology of corpus linguistics. In short, a corpus is ‘a collection of texts or parts of texts upon which some general linguistic analysis can be conducted’ (Meyer, 2002, p. xi). Similarly, Kennedy (1998) points out that a corpus is a body of either written text or transcribed speech which is designed to serve linguistic analysis and description (p. 1). Atkins et al (1992) and Waugh et al (2007) argue that a corpus is a body of text which is

assembled according to specific criteria for specific purposes. Because of the nature of their data, containing natural language, corpora are suitable for the analysis of discourse markers by learners (McEnery & Wilson, 1996; Oakes, 1998). The texts are carefully sampled so as to be representative of the type of language being investigated. Corpora give researchers a view of language structure which has not previously been available (Sinclair 2004a). In many cases, corpora give three types of data: empirical support, frequency information and meta-information concerning participants and source data (Lüdelling & Kytö, 2008). They argue that ‘a corpus may provide extra-linguistic information (or meta-data) on such factors as the age or gender of the speaker/writer, text genre, temporal and spatial information about the origin of the text, etc [...] [which] allows comparisons between different kinds of text or different groups of speakers.’ (2008, p.ix). Aijmer and Simon-Vandenberg (2011) point out that using corpora makes it possible to investigate the distribution of discourse markers in speech as well as in writing and in different registers. Aijmer (2002) remarks:

‘Corpora represent actual performance and provide the opportunity to study the distribution and function of particles in extensive text extracts representing different registers.’ (p. 3)

The non-native corpus employed in this study is an official component of the Louvain International Database of Spoken English Interlanguage (LINDSEI) (see Appendix 1) and as such, the design is determined by the general LINDSEI model. This corpus will be the first part of LINDSEI representing English used by Saudi learners. The LINDSEI project was launched in 1995 at the Université catholique de Louvain (UCL) and resulted from collaboration between several universities internationally. It aims to collect oral data produced by advanced learners of English (De Cock, 2004; Gilquin, 2008). The proficiency level of the participants is established by the number of years of studying English at

university. It is a criterion that all participants must be in their third or fourth year (De Cock, 2004; Gilquin, 2008; Gilquin et al, 2010). To date, 20 mother tongue backgrounds are available in the database: Arabic, Brazilian Portuguese, Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Greek, Italian, Japanese, Lithuanian, Norwegian, Polish, Spanish, Swedish, Taiwanese and Turkish. Each corpus in LINDSEI must have 50 interviews, comprising a total of approximately 100,000 words per component. The corpus has been marked up in such a way that each sentence uttered by the learner/interviewee (B-turns) can be interrogated independently of those uttered by the interviewer (A-turns). The term mark-up has been defined by Atkins et al (1992) as ‘adding into the text some conventional set of readable labels or tags, indicators of such text features [...] and sentence boundaries, headings and titles, various types of hyphenation, printers’ marks, hesitations, utterance boundaries, etc.’ (p. 9). The following example is from the SLC corpus illustrating how the interviewer’s turn and interviewee’s turn are represented in the LINDSEI database:

(1)<A> hi <first name of interviewee> how are you </A>

<B> fine how are you </B>

<A> it’s nice to me meet you </A>

<B> thank you </B>

<B> I’m fine thank you </B>

<A> (erm) so you chose topic number one </A>

<B> yes </B>

<A> you gunna talk about it </A>

<B> yes </B>

<A> ok great so you gunna talk about an experience you have had which has taught an important lesson </A>

(SLC-AR06)

All LINDSEI components must follow the same format prescribed for the database. The reason for this is to make comparisons between corpora possible for linguists and other researchers. Designing corpora on parallel principles could potentially resolve most of the difficulties arising in cross-corpora comparison (Pichler, 2010, p. 586). In principle, following the standard format of LINDSEI gives my data the advantage of using native speaker of English data, LOCNEC, compiled by De Cock (2004), which also used the same structure as LINDSEI. Native speakers in LOCNEC were British students majoring in English language and/or linguistics at the University of Lancaster. However, access to LOCNEC is restricted to the project collaborators.

The twenty sub-corpora of LINDSEI offer a wide range of possibilities for research using contrastive interlanguage analysis (CIA). Granger (1998) points out that CIA is different from contrastive analysis (CA), which works on comparing different languages, in that it compares and contrasts NNSs and NSs of a language in a comparable situation. In this study, LOCNEC is used as a control corpus. By comparing the SLC data with data of NSs, it is expected that the similarities and differences in using the discourse markers can be identified quantitatively and qualitatively.

Table 3.1 illustrates the number of interviews in each corpus of this study using the SLC and LOCNEC. It also shows the total number of words for both turns: interviewers' and interviewees'. In this study, the analysis is limited to the interviewees' turns, i.e. the B-turns.

**Table 3.1 Number of interviews and words in LOCNEC and the SLC<sup>11</sup>**

	Number of interviews	Gender		Number of words (A and B turns)	Number of words (B turns only)
		male	female		
Main LOCNEC	50	20	30	170,533	117,417
Customized LOCNEC	27	09	18	93,789	68,000
SLC	50	29	21	94,166	68,100

It can be noted from Table 3.1 that the size of LONCNEC is larger than the size of the SLC by more than 50%. Therefore – and for the purpose of comparison with LOCNEC in this study – a tailored corpus has been created (cf. Götz & Schilk, 2011). To do so, a randomized sample of interviews from LOCNEC comprising almost 68,000 words has been prepared.<sup>12</sup> This sample included 27 recordings. This customized corpus has been created because of the time limitations for my PhD. The results of frequency of use from both corpora were normalized, that is, standardized according to a consistent text length of 10,000 words. With this information, I calculated the number of each discourse marker tokens in each speaker's interview. To obtain the frequency, the number of the tokens is multiplied by 10,000 and divided by the total word count, as follows:

$$\frac{(\text{Number of discourse marker tokens}) * 10,000}{(\text{Total number of words in corpus})} = (\text{Frequency of discourse marker})$$

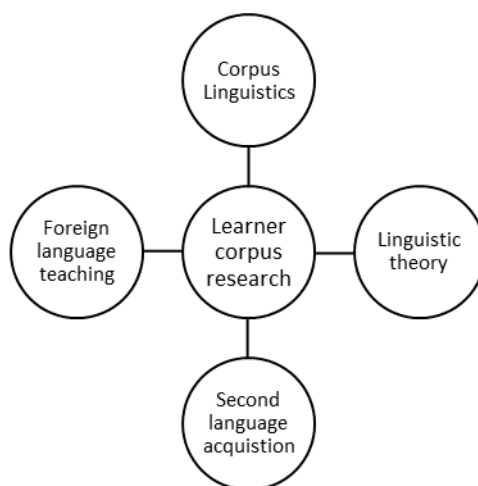
Following Pichler (2010), in indexing raw scores, the total number of tokens of discourse markers produced by participants was divided by the total number of words of the corpus. Then this number was multiplied by 10,000. Pichler (2010) believes that this method gives a normalized measure which helps to compare frequencies. She adds 'It, thus, seems to provide

<sup>11</sup> Numbers of words in the table are approximate.

<sup>12</sup> For this purpose, the website <http://www.random.org> was used.

an accountable measure of relative frequencies which is easily replicable and allows valid cross-corpora comparisons to be made' (p. 594). Adopting this process is important in this study because it aims to compare results between the two corpora: the SLC and LOCNEC.

As an indication of effectiveness, a number of previous studies have incorporated spoken learner corpora within the LINDSEI database (Aijmer 2004; Aijmer, 2009; Aijmer, 2011; Buysse, 2007; Trillo, 2002, to name but a few) which specifically investigate discourse markers. Granger (2009) defines learner corpora as 'electronic collections of foreign or second language learner texts assembled according to explicit design criteria' (p. 14). It can be argued that these studies opted to work on this collection of learner corpora because they provided what researchers were looking for. Baker (2010) believes that learner corpora can be especially useful to compare the overuse or underuse of lexis or grammar. Another benefit yielded by using learner corpora is that it gives teachers a chance to identify areas of errors at various phases of development (Baker, 2010). Thus, the reason it was chosen to use LINDSEI in this research, following Brand and Kämmerer (2006), is that these corpora enables the researcher not only to compare an NS corpus with a learner corpus but they also allow comparison of the performance of a variety of learner corpora. In addition, Granger (2009) comments that because learner corpora encompass data from language learners, they are special in terms of the range of expertise required for analysis and their inherent interdisciplinarity. Figure 3.1 depicts Granger's view of this relationship.



**Figure 3.1 Core components of learner corpus research (Granger, 2009, p. 15)**

### *Saudi sub-corpus of spoken English*

Before I started collecting my main corpus, a pilot corpus was developed to help test and refine the study; in particular, this was intended to estimate whether the full data, when complete, would be sufficient and appropriate for answering the research questions. Conducting a pilot study is a procedure recommended by Dörnyei (2007), who states:

‘[A]lways pilot your research instruments and procedures before launching your project. Just like theatre performances, a research study also needs a dress rehearsal to ensure the high quality (in terms of reliability and validity) of the outcomes in the specific context.’ (p. 75)

Because the research included human participants, I needed to obtain ethical approval. The University of Salford Ethics Committee approved my research in the first round of application (see Appendix 2). As the data in this study were to be collected for the first time with no previous knowledge of how rich they might be with discourse markers to be investigated, there were some concerns that the data in LINDSEI provided by the interviews might not be sufficiently populated with discourse markers to enable the research questions to



be answered. Therefore, I proposed to use another type of data: 30-minute paired conversations. In these conversations, there would be two participants discussing a number of general topics. Participants are offered a number of cards from which they should agree what to talk about. For each 10 minutes in the 30-minute conversation, the participants would discuss a different topic to avoid boredom. However this 'Plan B' was not needed when the results from the pilot study proved positive.

The pilot study was conducted with students based at the University of Salford. There were four participants, two males and two females. They were Arabic – but not all Saudi – non-native English language speakers and they were doing either an MA or a PhD in English. All the interviews were recorded to be analysed later. In addition, learner profile questionnaires were completed by the participants.

After all the data had been collected, the transcription process started (see one of the samples in Appendix 3). For the four interviews, the total number of words was 6,795 after the interviewers' A-turns were excluded. The average word count for each interviewee was 1,698. It was necessary also to transcribe the pilot study data to see the extent to which this type of data would be sufficient for my research purposes. The data were transcribed following the LINDSEI protocols. After transcribing the recordings, all discourse marker tokens were extracted for each interview.

**Table 3.2 Discourse markers found in the pilot data: raw and normalized frequencies**

Discourse marker	Raw frequency	discourse markers (298)	words (6,795)
<i>Yeah</i>	54	18.12	0.79
<i>I mean</i>	9	3.02	0.13
<i>Alright</i>	4	4.34	0.05
<i>Ok</i>	24	8.05	0.35
<i>Actually</i>	26	8.72	0.38
<i>Like</i>	22	7.38	0.32
<i>So</i>	36	12.08	0.52
<i>You know</i>	41	13.75	0.60
<i>Anyway</i>	1	0.33	0.01
<i>Obviously</i>	5	1.67	0.07
<i>Kind of</i>	4	1.34	0.05
<i>But</i>	45	15.10	0.66
<i>Yes</i>	19	6.37	0.94
<i>I think</i>	8	2.68	0.11
Total	298	100	4.99

Out of four recordings, there was a total count of 298 discourse markers as illustrated in Table 3.2. Among these discourse markers, there was clear variation in their use with an interesting occurrence and absence of some of them for each interviewee. To sum up, based on the results of the pilot study, I concluded that the 15-minute interviews and the methodology of collecting them should be sufficient for the research requirements.

After the pilot study, preparation for the collection of the data for the main study started. As the title of my thesis indicates, the context of the study is Saudi Arabia the official language of which is Arabic. English is officially taught in public and private schools so that students should attain a standard of proficiency which enables them to benefit from materials in English and to communicate well in spoken and written forms (Ministry of Education, 2002). In addition, it is used as the medium of training in companies and organizations such as

airline companies, oil companies, etc. (Mahboob & Elyas, 2014). However, the limited use of English is one of the reasons why collecting spoken English data is particularly difficult.

The compilation of the Saudi sub-corpus of LINDSEI began in March 2011 and was carried out over a period of one year. It consisted of 50 informally recorded interviews in English between Saudi students studying English as foreign language. The participants in this study were all third and fourth year university students (males and females, aged 20–25 years) who were studying English language at three different universities and one college. Table 3.3 shows the names of the universities and the college and the cities in which the participants of this study are based. To give an idea of the different locations, Figure 3.2 shows a map of the Kingdom of Saudi Arabia illustrating the cities mentioned in the table.

**Table 3.3 Universities and college in which data were collected**

No	Name of university/college	City	Number of participants
1	Umm alqura University	Makkah <sup>13</sup>	24
2	Taif University	Taif	16
3	Princess Nora University	Riyadh	7
4	Royal Commission Yanbu Colleges and Institutes	Yanbu	3

---

<sup>13</sup> Mecca on the map.



**Figure 3.2 Map of Saudi Arabia**

The reason for collecting data from different regions was to try to achieve a reasonable degree of demographic representativeness. Biber (1993) defines representativeness as ‘the extent to which a sample includes the full range of variability in a population’ (p. 243). However, given the constraints of time and funding for this project, it was impossible to represent every university, or even every major university, in Saudi Arabia. Thus, the representativeness of my results could be questioned. That said, despite the fact that the SLC is not a sizeable corpus, it includes spoken data from three different regions in Saudi Arabia and it is the first corpus to include spoken data by male and female English language learners. Of course, not being able to collect data from all regions, i.e. North, South, East and West, in Saudi Arabia makes it difficult to make strong claims about variations within the corpus. However, because data were collected from more than one region and had male and female participants, tentative generalizations from the results of this study can be made.

After sending letters to the Heads of Departments to ask for permission to collect data (see Appendix 4), agreement letters were successfully obtained for the study to be implemented (see Appendix 5). Once the permission letters had been received, the process of conducting the interviews started.

The interviews were conducted by members of the teaching staff according to the format of LINDSEI. This adds two significant benefits to this research:

1. As the interviewer was a member of staff, the interviewees should be familiar with him/her. This should minimize the stress that interviewees might feel and reduce the power relation<sup>14</sup> within the interview, thus helping to obtain a ‘favourable interactive position’ (Labov, 1981, p. 15).
2. It mitigates the biasing effects that could occur if the interviewer were the researcher. For instance, the researcher might consciously or unconsciously influence the interviewees’ linguistic responses (e.g. by highlighting his/her use of discourse markers or using them excessively).

Going back to the interviewers in this study, there were five Saudi interviewers (see Table 3.4). All of them were English language teachers teaching in the same department in which they collected data. Interviewers were briefed in advance on how to conduct the interviews and were fully aware of how to use Dictaphones. All steps were explained to them carefully, as was the process of how the learner profile should be completed by the participants.

---

<sup>14</sup> It could be argued that there is an unequal power relation between staff and students (see, for example, Seidman, 2013). However, in this case, the use of a task mitigates the issue, unlike situations in which opinions are sought.

**Table 3.4 Interviewers' gender, mother tongue and institute**

Interviewer	Gender	Mother tongue	Institute
1	Male	Arabic	Umm alqura
2	Female	Arabic* <sup>15</sup>	Umm alqura
3	Female	Arabic*	Umm alqura
4	Male	Arabic	Royal Commission Yanbu Colleges and Institutes
5	Female	Arabic	Princess Nora University

Procedures for how the interviews should be performed were clearly written and explained to all interviewers. Interviewers were told that the purpose of this study was to collect as much conversational material as possible as this data would be studied from many different aspects. To ensure that the procedures were clear, sample recordings from the pilot study and copies of their transcriptions were given to the interviewers to give them an idea of what each interview should look like. Interviewers were asked to submit their first recordings to me so that I could make sure they were done correctly. There were some instances of incorrect procedures in the interviews. For example, some interviews were only five minutes long. In this case, the interviewer had assumed that the participant should deliver the three tasks all in five minutes rather than five minutes each, giving a total of 15 minutes. These interviews were discarded from further analysis. Additional clarifications of how the interviews should be conducted were provided in instances in which incorrect procedures were found.

All interviewees freely agreed to participate and signed a consent form (see Appendix 6). Neither interviewers nor interviewees were informed of the main focus of this research. They were simply told that the aims of this research were:

1. To look into how English language is used by Saudi English learners.

---

<sup>15</sup> The interviewer was a native speaker of both Arabic and English having been born in the USA and having received most of her education there.

2. To construct a corpus based on these data.

Because English is spoken in very limited environments in Saudi Arabia, interviews were one of the possible options. Sociolinguistically, interviews are considered to be an effective method for collecting information and genuine data. Liao (2009) suggests that when doing interviews, several issues should be borne in mind:

1. The interviewer's presence in the discussion must be kept to a minimum to give the interviewee as much space as possible.
2. Whether the interviewer is an NS or NNS will not affect the interview (Young, 1991). However, to Liao (2009), interviewees may not feel as nervous or uneasy when speaking to an NS as when speaking to an NNS; with the latter, they might well reveal a more true reflection of their real language ability.
3. Informants/participants must not be informed of the main focus of the research (in this case discourse markers) but rather simply have a casual conversation.

All these guidelines suggested by Liao (2009) were carefully considered and applied in my data collection. For example, interviewers were told to try to keep their presence in the discussion to the minimum. However, because the interviewers and the participants shared the same L1, I noticed that participants asked the interviewers about some vocabulary or unfamiliar expressions. Similar to Liao (2009), though, I take this as a sign of feeling comfortable as opposed to when being interviewed by NSs.

On the day of the interview, every interviewee was asked to complete a learner profile based on the LINDSEI model (see Appendix 7). This learner profile contains a variety of linguistic and non-linguistic information that might potentially be correlated to the frequency of discourse marker use. For example, in terms of non-linguistic factors, the learner profile gives

the following: age, gender and the relationship between the interviewer and the interviewee. Of course, age is one of the important factors to consider when investigating the patterns and the distribution of the use of discourse markers (Blyth & Wang, 1990; Muller, 2005; Stubb & Holmes, 1995; Trillo, 2002), but because the age range of the participants in this thesis is 20–25 years, which is quite narrow, the age factor is not considered. Moreover, there is no other age group to which it can be compared. Gender was found to be significant in Östman's (1981) study as he concludes that the discourse marker *you know* tends to be used more by women than by men. Regarding the last factor in this group, the relationship between the interviewer and the interviewee, previous studies such as Redeker's (1990) and Jucker and Smith's (1998) found that this factor exerted an influence on both the frequency and selection of discourse marker use. Similar to Jucker and Smith's (1998) study, the learner profile of LINDSEI asks the interviewers about their relationship with the learner. It gives three options: familiar, vaguely familiar and unfamiliar.

Linguistic factors noted for the participants include: language spoken at home, medium of instruction in primary and secondary school, years of English at school and university, **time spent in an English-speaking country and other foreign languages**. All these factors presented by the learner profile focus on the acquisition of English (cf. Callies and Szczesniak, 2008).

The interviews in this study were made up of three tasks: set topic (see Appendix 8), free discussion and picture description (see Appendix 9). Each interview should last approximately 15 minutes following the same pattern. The interviews started with the set topic. The interviewee was asked to talk about an experience that he/she had had and that had affected him/her, a country that they had visited and liked/disliked, a film or a movie that they had watched and liked and why, etc. This led to a free discussion with the interviewer.



Finally, every interview ended with a short picture-based story-telling activity. The structure of interviews that LINDSEI provides is more like a discussion/conversation than question and answer. Underhill (1987) suggests that direct interviews are of three different types: discussion/conversation, interview and question and answer. In the discussion/conversation interview the interviewer keeps overall control but at the same time is willing and able to yield the initiative to the learner to steer the conversation or bring up a new topic. The conversations which these interviews provide are quasi-authentic. This is because they are about a short provocative topic chosen freely by interviewees which functions as a trigger for the talk. The fact that the interview is initiated by a topic selected from options predetermined by an interviewer means that the talk does not occur completely ‘naturally’, but this should soon be forgotten by the interviewees. In this study, another factor that would have helped put the interviewees at ease was that all the participants in this corpus were voluntary and most importantly the setting was outside the classroom and not an exam.

When the interviews were collected and the learner profiles had been completed, they were coded for the transcription process. Coding includes multiple tasks. Dörnyei (2007) states that the process ‘involves highlighting extracts of the transcribed data and labelling these in a way that they can be easily identified, retrieved, or grouped’ (p. 250). I transcribed the 50 interviews orthographically and marked them up myself to ensure that a high level of consistency was achieved.

All 50 recordings were transcribed and marked up following the LINDSEI protocols (see Appendix 10) using the Soundscribe tool.<sup>16</sup> The transcription included paralinguistic features such as laughter, coughing and whispering. It also included contextual comments such as

---

<sup>16</sup> <http://www-personal.umich.edu/~ebreck/code/sscriber/>

non-linguistic events. Other details such as empty pauses or filled pauses, backchannelling, unclear passages and truncated words were also transcribed. When the initial orthographic transcription was done, it was cross-checked by other linguists who are experienced in working with spoken data.

All interviewers' turns were then excluded and the revised transcripts were saved into TXT files to be imported into the Antconc program, a freeware multi-purpose corpus analysis toolkit designed by Laurence Anthony at Waseda University, Japan.<sup>17</sup> This program was used to generate concordance lines of all occurrences of each searched word. This helps to label whether the word functions as a discourse marker or not, as shown in Figure 3.3.

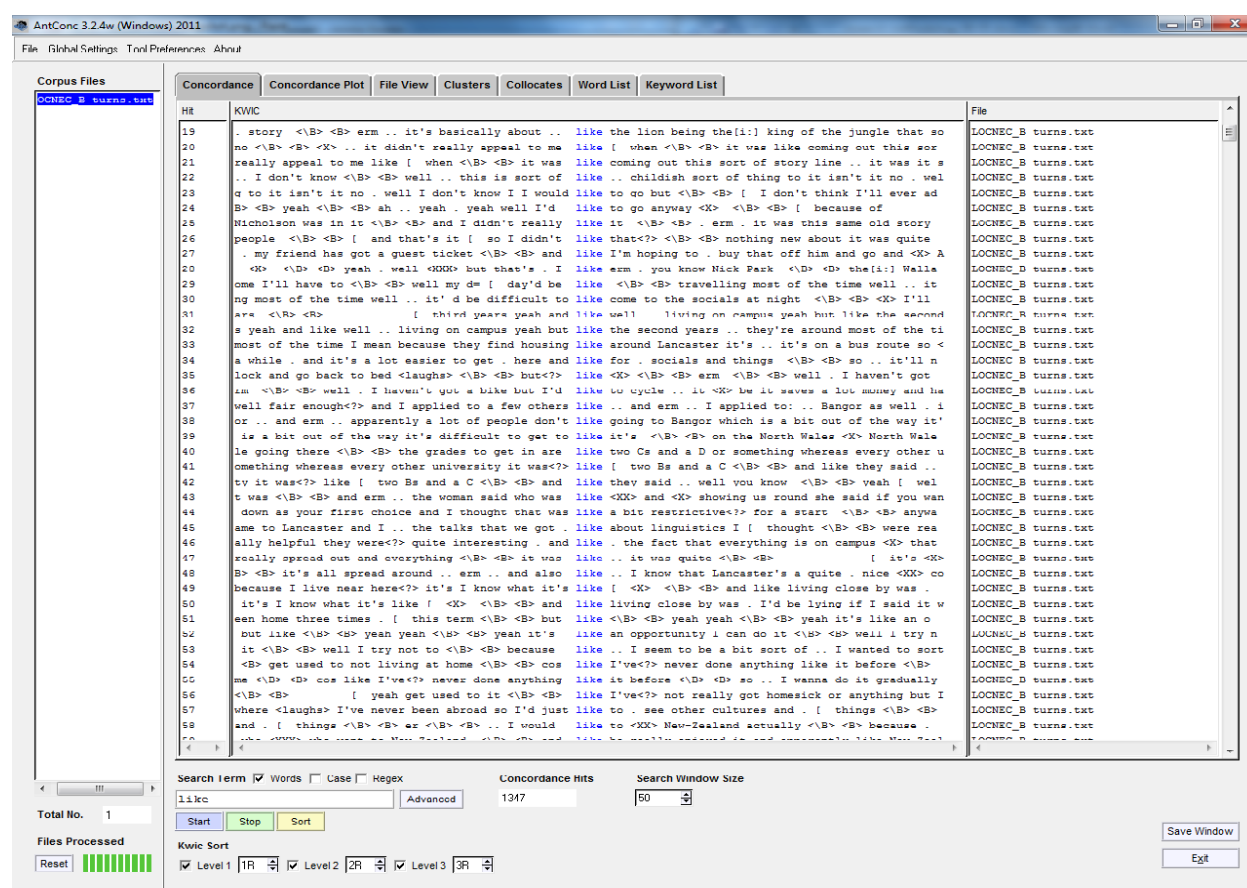


Figure 3.3 Screenshot of AntConc showing concordance lines of *like* in LOCNEC

<sup>17</sup> <http://www.antlab.sci.waseda.ac.jp/software.html#antconc>

To code the data, all the concordance lines generated from Antconc were exported into a Microsoft Excel file (which provides multiple columns to support coding). The first task in Excel was to determine whether the key word in context (KWIC) was actually functioning as a discourse marker or was merely a lexical use of the word (e.g. *A lot of people don't like going to Bangor*). This step helps to identify recurrent patterns in the discourse markers under investigation (Sinclair, 2004). Next, each example containing a genuine discourse marker was coded according to its function in the utterance. Other parameters include gender and unclear instances, for example cases in which the function could not be categorized. Additional fields were used to record various notes, such as cases of a discourse marker that it was problematic to analyse and required further attention (Smith et al, 2008). An example screenshot from Excel is shown in Figure 3.4.

	B	C	D	E	F	G	H
1	KWIC	DM/not	Function	Gender	Participant code	Date	Note
2	y a got a student because ?/B ? ?B? I actually <u>LIKE</u> teaching for my father is an English teacher ?/B	N	V	M	AR01	18/11/2012	
3	got the thing ?/B ? ?B? so I really really do <u>LIKE</u> English ?/B ? ?B? or let's say teaching Englis	N	V	M	AR01	18/11/2012	
4	my relatives so they ask me to teach them English <u>LIKE</u> basics ?/B ? ?B? is am are and so on so one of	N	PCN	M	AR01	18/11/2012	
5	on so one of these students was my friend he was <u>LIKE</u> a talkative guy you can say I mean I don't mean i	Y	SAE	M	AR01	18/11/2012	
6	?/B ? ?B? inshallah ?/B ? ?B? but that was <u>LIKE</u> XX?/B ? ?B? it's <u>LIKE</u> I can't describe it ?/B	Unclear		M	AR01	18/11/2012	
7	/B ? ?B? but that was <u>LIKE</u> XX?/B ? ?B? it's <u>LIKE</u> I can't describe it ?/B ? ?B? sorry ?/B ? ?B	Y	XML	M	AR01	18/11/2012	
8	it ?/B ? ?B? sorry ?/B ? ?B? yeah its just <u>LIKE</u> a you just said that i said it that set maybe I d	N	PCN	M	AR01	18/11/2012	
9	e youngest among all the students ?/B ? ?B? is <u>LIKE</u> i guess 16 years old ?/B ? ?B? or 17 may be an	Y	ANM	M	AR01	18/11/2012	
10	6 years old ?/B ? ?B? or 17 may be and so this <u>LIKE</u> one of the experiences in life i learn from my st	Y	SAE	M	AR01	18/11/2012	
11	so when i when my three training days came i was <u>LIKE</u> the best among of my group ?/B ? ?B? yeah beca	Y	SAE	M	AR01	18/11/2012	
12	ong of my group ?/B ? ?B? yeah because usually <u>LIKE</u> experiencing things before ?/B ? ?B? yeah lik	Y	XML	M	AR01	18/11/2012	
13	ike experiencing things before ?/B ? ?B? yeah <u>LIKE</u> teaching more than English and the proof of that	Y	XML	M	AR01	18/11/2012	
14	that sometimes some of my relatives i teach them <u>LIKE</u> say in Arabic literature ?/B ? ?B? science so	N	PCN	M	AR01	18/11/2012	
15	are ?/B ? ?B? Maybe me ?/B ? ?B? Because <u>LIKE</u> say in my second year i guess in high school ?/B	Y	ANM	M	AR01	18/11/2012	
16	fter that i started teaching and pre-experiencing <u>LIKE</u> ?/B ? ?B? new things ?/B ? ?B? so that was	Y	XML	M	AR01	18/11/2012	
17	f his classes ?/B ? ?B? which was i don't mean <u>LIKE</u> praise my father ?/B ? ?B? no but the thing is	Y	SAE	M	AR01	18/11/2012	
18	his school his own school ?/B ? ?B? they have <u>LIKE</u> summer school and they have er English courses an	Y	XML	M	AR01	18/11/2012	
19	English x he was the teacher for this course for <u>LIKE</u> from the beginning from the first time he came to	Y	ANM	M	AR01	18/11/2012	

**Figure 3.4 Screenshot of an Excel spreadsheet showing the analysis of the discourse marker *like* in the SLC**

To test the richness of the data and determine which discourse markers were used the most, initial frequency analysis was carried out. I believe that this step was significant because it enabled me to ensure early on that the data obtained were sufficient to answer the research questions. Not doing this test would have meant that there would have been a strong possibility of finding out too late that the data were insufficient.

To perform this test, the first 10 interviews were selected (20% of the total number of recordings). The results from these 10 recordings revealed a number of discourse markers which were used differently among the 10 interviewees (see Table 3.5). For example, the discourse marker *so* was used by AR5 35.2 times per 10 thousand words (5 raw counts), whereas another speaker, AR4, used the same discourse marker 105.9 times per 10 thousand words (18 raw counts). Some speakers, for example AR6 and AR10, never used *so*.

Table 3.5 indicates the overall frequency of discourse markers in the 10 recordings. Having completed a frequency table, the next step was to find the most frequent discourse markers in the data with different characteristics. These were *you know*, *like* and *so* occurring a total of 215 times per 10 thousand words (388 raw counts) (see Table 3.6).

**Table 3.5 Frequency of discourse markers per 10,000 words for 10 recordings in the SLC**

	AR01	AR02	AR03	AR04	AR05	AR06	AR07	AR08	AR09	AR10	Total
<i>Yeah</i>	187.6	33	122	112.1	133.6	153.8	9.7	-	2.4	4.5	758.7
<i>Ok</i>	10.4	25.8	71.8	-	28	12.3	4.8	-	0.4	1	154.5
<i>Actually</i>	26.1	3.7	-	93.5	77.4	24.6	9.7	-	0.2	0.3	235.5
<i>So</i>	125.1	62.5	43.1	105.9	35.2	-	53.2	53.5	3	-	481.5
<i>you know</i>	260.6	110.3	157.9	6.2	7	6.2	19.4	-	8.4	0.5	576.5
<i>I mean</i>	-	161.8	-	31.1	7	-	-	14.6	0.1	-	214.6
<i>kind of</i>	-	3.7	14.4	18.7	-	-	-	4.9	0.6	-	42.3
<i>Well</i>	0.4	0.3	-	-	-	0.6	0.2	-	0.8	0.3	2.6
<i>But</i>	135.5	106.6	71.8	74.7	70.3	37	87.1	136.4	4	1.7	725.1
<i>Like</i>	234.5	62.5	394.8	168	-	55.4	4.8	29.2	31	-	980.2
<i>Just</i>	31.3	172.8	21.5	31.1	14.1	135.3	4.8	-	1.6	0.5	413
<i>I think</i>	26.1	29.4	57.4	37.4	28.1	116.9	-	34.1	0.6	0.1	330.1
<i>Yes</i>	-	62.5	107.7	255.3	14.1	18.5	62.9	48.7	-	-	569.7
<i>Because</i>	26.1	51.5	71.8	62.3	77.4	43.1	4.8	73.1	1.2	0.6	411.9
<i>of course</i>	15.6	58.8	-	6.2	7	-	9.7	-	0.4	-	97.7
<i>Er</i>	57.3	172.8	136.4	137	225	178.4	91.9	321.5	0.8	8.5	1329.6
<i>Erm</i>	99	14.7	-	68.5	225	47.5	-	194.8	-	2.6	652.1
<i>I see</i>	-	3.7	-	-	-	-	-	-	-	-	3.7
<i>Right</i>	-	3.7	7.2	25	-	-	-	-	0.6	-	36.5
<i>Basically</i>	-	3.7	-	-	-	-	-	-	-	-	3.7
<i>Really</i>	57.3	3.7	50.3	12.5	14.1	-	-	4.8	0.4	0.6	143.7
<i>Oh</i>	-	-	-	-	21.1	24.6	9.7	-	0.4	0.3	56.1

**Table 3.6 Number of most frequent discourse markers in the 10 recordings and their frequency per 10,000 words**

	Number of tokens	Frequency per 10,000 words
<i>You know</i>	155	86
<i>Like</i>	148	82
<i>So</i>	85	47
Total	388	215

The conclusion from this pilot study guided the decision concerning what discourse markers should be investigated. It provided evidence that these selected discourse markers occurred with high frequency.

However, there are some limitations in this corpus that I would like to acknowledge. The first is that LINDSEI's learner profile does not include the social class of the speaker, despite the fact that this variable has been found to be significant in the use of discourse markers (see Fuller, 2003; Stubbe & Holmes, 1995). LINDSEI includes only students as participants. Similar to LINDSEI is the Giessen-Long Beach Chaplin Corpus (GLBCC) which Müller (2005) used for her study. Müller (2005) states 'we made sure the non-native participants would have mastered a certain level of English sufficient to conduct a conversation. Beyond this, we have not been able nor did we consider it necessary to check on the social class of the students' background' (p. 43). Although this point is certainly debatable, because my SLC data are compared to the native speaker LOCNEC data, which do not include the social class variable, there seems to be little benefit in going beyond what is already prescribed in LOCNEC. A second limitation is in terms of prosody. The LINDSEI data are all transcribed orthographically but not prosodically. However, in the case of this study, the provision of the audio files work as a reference whenever there is a need. There were many cases in which some instances were not clear at first when read from the transcriptions, but by referring to

their audio files, the picture of whether cases were discourse markers or lexical items became much clearer. At present, recordings of LINDSEI corpora are not released with transcriptions but it is being considered in the future.

Overall, the establishment of the Saudi learner corpus of spoken English is expected to make a contribution to this discipline by:

1. informing the teaching of spoken English in Saudi Arabia when analysing the results;
2. making data from Saudi learners of English available to researchers in academia internationally;
3. serving as a model for the compilation of any other corpora of spoken English in Saudi Arabia, e.g. corpora for medicine or business.

### ***Analysis of textbooks***

As it has been stated in the literature, discourse markers are not usually taught at school (Hays, 1992; Sankoff et al, 1997) although it has been argued by some scholars that English teaching textbooks do have an influence on the use of discourse markers (Müller, 2004; Trillo, 2002). De Klerk (2005) investigated the use of discourse markers (she calls them discourse particles) in Xhosa English. The results of her study showed that there were differences in the use of the discourse marker *well* between Xhosa speakers of English and native speakers of English. She argues that this is due to the fact that discourse markers are largely overlooked in the educational system. This was previously stated by Mindt (1992, as cited in Sampson & McCarthy, 2005, p. 293) who points out that the use of grammatical structures in textbooks differs from their use in authentic English (p. 186).

In Saudi Arabia, English is almost exclusively learned through formal education. Studying English language at schools used to start from first year at intermediate school (Sedgwick, 2001) but this recently changed when the Saudi Ministry of Education decided that English language teaching should start from the sixth year in the primary stage. All students – boys and girls – follow the same curriculum. Therefore, all of the participants in my study have been exposed to the same textbooks. These textbooks are the main source for English teachers and the basis of most classroom learning from year 6 of elementary education until year 12 at secondary level.<sup>18</sup> Textbooks constitute a dominant and essential part of English language teaching.

The textbook database in this study aims to provide a description of the discourse markers *so*, *like* and *you know* in teaching materials. An examination of how discourse markers are represented and used in the English language textbooks will help to see the extent of the attention given to these items in the local pedagogical setting. To my knowledge, there has been no study to date on how discourse markers are used in the classroom in the Saudi context or how discourse markers are represented in Saudi textbooks of English. Therefore, although this is not the main focus of this research, I quantitatively examine how the functions of the selected discourse markers are represented in Saudi textbooks of English as a foreign language (EFL). The reason for this examination is to explore the anticipated connection between the frequency of a discourse marker in the textbooks and its use in the data.

---

<sup>18</sup> Titles of Saudi English language textbooks:

1. Elementary level: '*English for Saudi Arabia*'
2. Intermediate level: '*Say it in English*'
3. Secondary level: '*English for Saudi Arabia*'

Note: these textbooks were those participants in this study had studied. However, they are no longer used in schools.



Unlike the two corpora which are available electronically, the textbooks in this study are in PDF formats. Some of them are easy to search using the 'Find' function but others are scanned copies which do not allow the search process. Therefore, in this case, the search for discourse markers was done manually.

The textbooks were referred to after the analysis of each discourse marker in the SLC had been done. The initial analysis of the textbooks revealed some information about only one discourse marker, *so*, but nothing was found concerning *like* and *you know*, the other major discourse markers in this thesis. Therefore, the discussion concerning textbooks will relate to the case study of *so* only.

### **3.3 Research methodology**

Although there are more than 100 discourse markers in English (Fraser, 2006, p. 196), this research focuses only on a small number of them, namely *so*, *like* and *you know*. These discourse markers were chosen for two reasons: first, they appear most frequently in the data with different characteristics; second, these three discourse markers have previously been covered in the literature on discourse markers and this provides a strong foundation on which to base and compare my own analysis.

Based on the characteristics identified in section 2.2 (Brinton, 1990; Brinton, 1996; Holker, 1991; Schiffrin, 1987), I consider the selected discourse markers to be linguistic elements which:

- are syntactically optional;
- are difficult to place within a traditional grammatical word class;
- are multifunctional;

- are predominantly used in oral discourse;
- have little or no propositional meaning;
- may occur at the beginning, middle, or end of a discourse unit or form a discourse unit of their own.

My research aims to analyse the data quantitatively and qualitatively using mixed methods research. Dörnyei (2007) describes mixed methods as a combination of qualitative and quantitative methods in a single piece of research. The term ‘mixed methods’ is that settled on in the field (Dörnyei, 2007). Schmied (1993) argues that the qualitative approach must precede any quantitative investigation. Hasko (2012) states:

‘Although Qualitative Corpus Analysis is often construed as not being concerned with frequencies and statistical classification of linguistic features identified in the data, the value of mixing qualitative and quantitative approaches in corpus research is uncontested. Rich insights that stem from Qualitative Corpus Analysis can serve as a precursor for the latter, which allows for quantification and classification of the linguistic forms.’ (p. 6 et seq.)

Similarly, Corbin and Strauss (1998, p. 34) point out that the qualitative aspect should direct the quantitative approach and that the quantitative feedback should relay back into the qualitative analysis in a circular way.

As this is the first study to explore discourse markers in the speech of Saudi speakers and because there is no previous literature, to the best of my knowledge, on discourse markers used by Saudis to which I can refer, the qualitative method may outweigh the quantitative. This will allow the provision of a fairly complete picture in terms of how the three discourse

markers under investigation were used by the Saudi participants. However, the results from the qualitative investigation should allow some quantitative results to be deduced. The quantitative side of the analysis is performed using mainly descriptive statistics. This consists of simple statistical analyses, such as establishing lexical size and frequency counts to show the occurrences and distribution of discourse markers in the text and to show whether these frequencies are significantly different across speaker groups (e.g. male/female and time spent in an English-speaking country).<sup>19</sup> For the overall frequencies of individual discourse markers, it is also appropriate to run a test of statistical significance to determine the extent to which the results of quantitative comparisons are due to chance. Hunston (2006) states that log likelihood calculations are used to measure the statistical significance of frequency differences between corpora. Dunning (1993) points out that the use of log likelihood ratios can help to achieve greatly improved statistical results; he argues that they can be applied to ‘very much smaller volumes of text [...]’ and ‘[allow] comparisons to be made between the significance of the occurrences of both rare and common phenomenon’ (pp. 62–63). The chi-square test may be less reliable than log likelihood, particularly at low frequencies, such as cases of just 5 tokens of certain discourse markers in the SLC (Rayson & Garside, 2000). Therefore, following Baker (2005), Meyer (2002) and King (2009), the log likelihood test is used in this thesis to compare the frequency of discourse markers between the SLC and LOCNEC.<sup>20</sup> To explain how the statistical test is performed in log likelihood, Rayson and Garside (2000) make use of a contingency table as shown in Table 3.7.

---

<sup>19</sup> For a holiday or a language course, but not for a stay of long duration.

<sup>20</sup> To calculate log likelihood values, use was made of the Log-likelihood calculator website at <http://ucrel.lancs.ac.uk/llwizard.html>

**Table 3.7 Contingency table for word frequency**

	<b>Corpus one</b>	<b>Corpus two</b>	<b>Total</b>
<b>Frequency of word</b>	<b>a</b>	<b>b</b>	<b>a+b</b>
<b>Frequency of other words</b>	<b>c-a</b>	<b>d-b</b>	<b>c+d-a-b</b>
<b>Total</b>	<b>c</b>	<b>d</b>	<b>c+d</b>

(Rayson & Garside, 2000, p.3)

In short, log likelihood is based on the calculation of the difference between the actual or observed frequencies of a linguistic item in two corpora and what would be the expected frequencies given the size of each corpus. The value ‘c’ refers to the number of words in corpus one and the value ‘d’ denotes the number of words in corpus two; these values are referred to as ‘N’ values. Both ‘a’ and ‘b’ are ‘O’ (observed) values (Rayson & Garside, 2000). Thus, the expected values are calculated as  $E1 = c*(a+b)/(c+d)$  and  $E2 = d*(a+b)/(c+d)$  (Rayson & Garside, 2000, p. 3). The next step is to calculate the log likelihood itself using the formula:  $LL = 2*((a*\log(a/E1)) + (b*\log(b/E2)))$  (Rayson & Garside, 2000, p. 3).

### **3.3.1 Functional analysis**

Before providing the grounds upon which the functions of the discourse markers under investigation are based in this study, the following subsection discusses why these specific discourse markers were selected.

#### ***Selection and categorization of discourse markers***

This research is focused on three discourse markers, *so*, *you know* and *like*. The reason these three discourse markers were selected was first that they occur most frequently in the SLC data. This in turn indicates the probable prominence of these items in university-level spoken Saudi learner English as a whole, although (as acknowledged in section 7.4) there are limits

to the degree of the representativeness and naturalness of the corpus. In fact, the discourse marker *so*, for instance, is not only a high frequency item in the SLC but is also one of the high frequency discourse markers in English discourse spoken by native speakers (Biber et al, 1999; O’Keeffe et al, 2007). I established high frequency as a criterion for the selection of discourse makers in this study as if these discourse markers have more than one function, there is a greater probability of all such functions being represented in the data. Another important reason was that all three discourse markers have been discussed thoroughly in the literature and are included in the inventory compiled by Brinton (1996, p. 32) (see section 2.1.3). The richness of the literature containing various previous studies covering the same discourse markers as this research provides a number of results to which I can compare the results derived from the SLC.

As discussed in the literature review (chapter 2), various functional approaches have been used in the study of discourse markers: these can be divided into two types, namely corpus-based and corpus-driven, based on whether they start from a given theoretical framework or from individual occurrences. The corpus-based approach means that the corpus is used mainly to ‘expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study’ (Tognini-Bonelli, 2001, p. 65). In contrast, in the corpus-driven approach, linguists are committed to ‘the integrity of the data as a whole’ (Tognini-Bonelli, 2001, p. 84) and aim to build theory from scratch (Xiao, 2008). Gast (2006) argues that when corpus-driven linguists want to approach data, they should be completely free from any pre-corpus theories because ‘all the relevant information is contained in the corpus itself, and the linguist’s task is to extract that information and make it visible’ (p. 114).

However, arguably the distinction between the two approaches – corpus-based vs. corpus-driven – has been overstated. McEnery et al (2006) and Xiao (2008) have argued convincingly that this is an idealized extreme.

Barlow (1996) contends that the results produced by a corpus-based investigation can be a firm basis for linguistic description and language learning. So, aiming to present a comprehensive account of the use of discourse markers at both textual and interactional levels of coherence, the approach followed in this thesis is corpus-based. There is no reason I should devise new terms or functions – ‘re-invent the wheel’ – *as long as* the data show the usage of discourse markers fits closely with the functions that have been reported. Tognini-Bonelli (2001) states that:

‘In this case [...] corpus evidence is brought in as an extra bonus rather than as a determining factor with respect to the analysis, which is still carried out according to pre-existing categories; although it is used to refine such categories, it is never really in a position to challenge them as there is no claim made that they arise directly from the data.’ (p. 66)

For example, the discourse marker *you know* has been found in the literature to have approximately 30 functions (Müller, 2005, p. 147). All the three selected discourse markers, *so*, *like* and *you know*, have been discussed thoroughly in the literature. Therefore, with the incorporation of functions from previous studies, this thesis is a corpus-based study which focuses on the coherence of relations between adjacent utterances linked by the discourse markers in question.

The qualitative aspect consists of a categorization system for the functions of each discourse marker, classifying their tokens according to the function they have been found to represent and comparing the Saudi speakers' usage to that of the NSs. The functions of each discourse marker are identified building on previous studies (e.g. Brinton, 1996; Fraser, 1990; Fraser, 1999; Fraser, 2009; Schiffrin, 1987).

To ensure the data coding in this study is consistent, certain guidelines are followed. Regarding the multi-functionality and determination of the primary function of the discourse markers, following Holmes (1984), Fraser (1999) and Aijmer (2002), the interpretation of the core or primary function can be identified based on the contextual cues available in the utterance. Fraser (1999) points out that every discourse marker has a specific core meaning. Aijmer (2002) states that 'the core function of a DM is defined in terms of what it is doing on the textual or interpersonal level' (p. 23). Even though Pichler (2010) allows two functions to be recognized in a single instance of a discourse marker, this approach is not adopted here because focusing on the core function makes it easier to achieve consistency and comparability. For illustration, let us consider two examples from the SLC and LOCNEC regarding the discourse marker *like*:

- (2) <B> or these things so i have to be very like very er very not er like i have to choose my acts<\B>
- <A> yeah I know what you mean <\A>
- <B> carefully yeah and this is almost all things (er) yeah that's all I think (er) <\B>
- <A> did you did you continue like after the institution did you like like college there or <\A>
- <B> no after that after *like* about four months <\B>
- <A> yeah <\A>

<B> I came back to Saudi Arabia because got married <\B>

<A> oh congratulation <\A>

(SLC-AR04)

(3) <A> oh how interesting <starts laughing> you must have a lot of imagination

<stops laughing> <\A>

<B> <laughs>.. yeah but (erm) .. I don't know I enjoy it but I'm not I don't think

I'm particularly good . so I might do that as a hobby <\B>

<A> uhu <\A>

<B> but probably think about going into advertising <\B>

<A> oh yeah <\A>

<B> or say *like* <X> article writing <\B>

<A> (mhm) <\A>

(LOCNEC-E03)

The two discourse markers *like* in these examples are both multifunctional. In example (2), *like* has two functions, one core function which is marking an approximate number, but it could also function as a hedge. Similarly, the second *like* in example (3) has a function introducing an example and could also function as a hedge. In this study, the core functions – marking an approximate number and introducing an example respectively – are those considered by virtue of their dominant role in each utterance.

The previous steps were adopted to ensure the analysis was done consistently. The last but not the least step was to ensure the reliability of the analysis was as high as possible. The following measures were taken to strive for consistent interpretation of the pragmatic



meanings of *so*, *you know* and *like*. A number of procedures were employed to set up the categories of the pragmatic functions before the full analysis. First, the main functions for each discourse marker were determined. Second, clear examples from the SLC and LOCNEC were used to illustrate each of these functions. Finally, the rest of the data were coded at least twice and there were two to three weeks between each coding. The agreement rate between the two codings was approximately 85%. According to Müller (2005, p. 28) when recording reliability achieves 70%, it is considered to be sufficient to deliver reliable results.

Regarding the quantitative part of the study and the frequency of discourse markers, the tokens were counted manually and then double-checked via a computer search. In what follows, the process of how occurrences were counted is explained. When a repetition of the same discourse marker was found, it was counted as a single occurrence, as illustrated by the double occurrence of *you know* in example (4).

(4) <A> so is it . a film play <\A>

<B> it was a film that I thought was particularly good <X> <\B>

<A> which is <\A>

<B> eh Much Ado About Nothing <\B>

<A> oh yes <\A>

<B> I loved that film so much I think (erm) .. I liked the fact that I I suppose like Shakespeare is so regarded as a real theatrical medium but now it's been <X> like (erm) with Zefirelli's version of Hamlet and then <\B>

<A> (mhm) <\A>

<B> you've got Kenneth Branagh with Much Ado and .. I to watch . the[i:] actors that you've got involved in it it's got to be one of the best things I've seen you had big Hollywood actors with <\B>

<A> (mhm) </A>

<B> the big stage actors of England all together in the same film </B>

<A> (mhm) </A>

<B> and they were all .. <X> </B>

<A> brilliant <laughs> </A>

<B> brilliant it was just I mean it had *you know you know* Denzel Washington and Keanu Reeves and who would have thought you'd put Keanu Reeves </B>

<A> yes </A>

<B> in that sort of film I mean he was so .. he was such a baddie in the film but he came across so well </B>

<A> yes </A>

(LOCNEC-E22)

In addition, *like* in example (5) was counted only once even though the two instances did not occur next to each other as with *you know* in the previous example.

(5) <B> and the one kilo they they charged it for like ninety four </B>

<A> yeah I know very expensive </A>

<B> it was *like* yes it was *like* thousand five hundred </B>

<A> so extra weight or </A>

<B> yes and he said I don't have money I bou= I I I pay the whole thing when we came here he forgot his (ur) iron with me

<A> right </A>

(SLC-AR3)

Similarly, in example (6) the *so* is repeated three times by speaker E05 in <B> turns but is only counted as a single occurrence.

- (6) <B> so yeah . that was <X> so .. that that that that . got me pretty annoyed <\B>  
<A> yeah <\A>  
<B> because <X> I'd just met this really nice girl and . I wasn't able to <\B>  
<A> (mhm) <\A>  
<B> start anything .. *so* .. *so* then . in .. *so* it was like in the[i:] end it . it it turned out  
that this girl that I met at the camp .. she had .. you know <X> continued writing to  
each other and phoned cos cos we lived in different parts of the country <\B>  
<A> oh yes (mhm) <\A>  
<B> <X> miles apart from each other and (er) . now she's my girlfriend <\B>  
<A> oh great <\A>

(LOCNEC-E05)

Finally, there were some instances classified as unclear. In this study unclear instances refer to instances which were, for instance, acoustically unintelligible as in example (7), or incomplete as in (8), or due to the interruption by another speaker as in example (9), or with all words available but still difficult to classify functionally as in example (10),

- (7) <A> so you want to continue <\A>  
<B> (mhm) . I took a year out between when I did that and then here *so* <XX> sort  
of <XXX> <\B>  
<A> oh what did you do <\A>

(LOCNEC-E10)

(8) <A> oh yeah (mhm) </A>

<B> something that is interesting but doesn't take so long and . *you know* </B>

<A> (mhm) </A>

<B> yeah <laughs> </B>

(LOCNEC-E03)

(9) <B> <laugh> in picture number three when he d= (er) he drew her as another I mean  
he he made her more beautiful and she was happy </B>

<A> but is she her </A>

<B> although it isn't really her *like* </B>

<A> but because she is </A>

<B> pretending that </B>

<A> because she is sitting there y= this made you say </A>

<B> yes that's (er) this is me </B>

<A> he painted me </A>

<B> yes </B>

<A> ok </A>

(SLC-AR05)

(10) <A> (mhm) <begin\_laughter> and you said oh yes <end\_laughter> </A>

<B> which sort of like they weren't s= they weren't supposed to do . never mind but  
I got I got on quite well with them (erm) .. but the Spanish family I mean the father  
was quite strict and the mother would sort of cook for you and put the food on the  
table and expect you to eat it all and *so* it you felt very much like another daughter

rather than a visitor <\B>

<A> (uhu) <\A>

<B> really .. *so* I mean sort of like if if your friend was getting punished you'd get punished with her if you'd done anything wrong .. I mean that that didn't actually happen but that was the way it felt sort of like if if you didn't abide by their rules then they'd be far more strict than the French <\B>

(LOCNEC-E18)

To conclude, this chapter has presented the databases used in this empirical study. It has introduced the Saudi learner corpus – SLC – and the process of its compilation, as well as the native speaker corpus, LOCNEC, which will be used as a reference data. The textbooks included in the study have been presented and the purpose of analysing them discussed. Finally, the methods used in establishing the discourse functions of discourse markers have been discussed.

## CHAPTER 4: *SO* IN THE DISCOURSE OF THE SLC AND LOCNEC

### 4.1 Introduction

In this chapter, the focus is on the discourse marker *so* and how it is used by Saudi English language speakers. This chapter consists of three sections. The first, section 4.2, looks at how *so* has been reported in the literature. Section 4.3 examines qualitatively how *so* is employed through an analysis of instances from the data of this study. In section 4.4, the quantitative results for *so* are presented. In section 4.5, the comparison of findings from the corpus and the textbooks is discussed. A summary of the whole chapter is provided in section 4.6.

### 4.2 Functions of *so* in the literature

As a discourse marker, *so* has received attention from a number of researchers (e.g. Aijmer, 2002; Blakemore, 1988; Bolden, 2008; Bolden, 2009; Müller, 2005; Östman, 1981; Redeker, 1990; Schourup, 1985). One of the first studies to focus on *so* was that conducted by Van Dijk (1979). In his paper *Pragmatic Connectives*, he points out that *so* ‘links two speech acts of which the second functions as a “conclusion” with respect to the first speech act’ (1979, p. 453). Schiffrin (1987) provides a detailed treatment of *because* (with a causal meaning) and *so* (with a result meaning). She argues that both of them ‘have semantic meanings realised on sentence and discourse levels’ (1987, p. 201). Schiffrin (1987) suggests that these meanings occur on three planes of discourse (out of the five planes discussed in her book), primarily functioning in relation to the ideational structure (Redeker, 1990; Redeker, 1991) and secondarily the information state and action structure. She gives a special term to each plane as Table 4.1 illustrates.

**Table 4.1 Semantic realization of *so* and *because* in discourse**

	Discourse plane		
	Ideational structure	Information state	Action structure
	Fact-based	Knowledge-based	Action-based
<i>because</i>	Cause	Warrant	Motive
<i>So</i>	Result	Inference	Action

(Schiffrin, 1987, p. 202)

For a fact-based relationship in the ideational structure, Schiffrin (1987) claims that *so* holds between idea units ‘the events, states, and so on, which they encode’ (p. 202). A knowledge-based causal relationship holds when some pieces of information are used as a warrant by a speaker to be inferred by the speaker him/herself (a speaker inference) or by the hearer (a hearer inference). In the action-based causal relation – the action structure plane – a motive for an action is presented by a speaker. This action could be carried out by the speaker or the interlocutor (Schiffrin, 1987). Schiffrin (1987) provides her own examples to illustrate the functions of *so* on each plane. Example 1 shows the fact-based result in the ideational structure:

(1)n. I said, ‘When we get home, Joe,’ I says, ‘you’ll carry me in the house!’

o. I say, ‘I ain’t got shoe:s, or no stockin’s on.’

p. He said, ‘Carry yourself in the house! I ain’t carryin’ y’in!’

q. **So** I went BOOM

(Schiffrin, 1987, p. 203)

The *so* expressed in (q) is a local fact-based use of *so*. Schiffrin (1987) states that *so* is used locally when only one event is causally related to it, whereas ‘*so* functions globally when it

covers a wide range of talk' (p. 203) as in example (2). Example (1) is an account of Henry and his brother walking in the snow and his brother refusing to carry him. The result in (q) is that Henry punched his brother on the nose. In the coda given in example (2), Henry recounts a story about throwing his brother in the snow and evaluates it as funny. Schifffrin (1987) believes that *so* in this example has a global function because it 'reports the brother's laughter as a reaction to the entire experience' (p. 204).

(2)w. *So*, he laughed about it for years,

x. he says it was worth it!

(Schifffrin, 1987, p. 204)

Example (3) illustrates the knowledge-based inference (warrant inference) on the information state plane.

(3)Zelda: And she says, 'We have lots of rooms,' she said, 'even for- Gary had a big dog,' she said, 'even for the dog!' *So*, it's really nice eh: y'know t' renew friendship.

(Schifffrin, 1987, p. 206)

In this example, Zelda was narrating her son's reunion with his friends. Schifffrin (1987) argues here that the events in this story do not factually cause the concluded result at the end of the story. In fact, they work as background evidence which warrants the speaker's general conclusions (p. 206).

The third type of *so* exemplifies the action-based use which functions on the action structure plane. The following is an example of this use of *so*:

(4) Irene: d. you take somebody like A1 says, I-he don't know anything about him.

e. Why should he vote for him?



Zelda: f. He'll-

Irene: g. There are alot of people that fell that way=

Zelda: h. Wait a minute=

Irene: =Henry!

Zelda: *so* who would A1 vote for?

(Schiffrin, 1987, p. 208)

At the beginning of this example, Irene was trying to support her position that Catholics will not vote for the candidate. She was describing her Catholic neighbour A1's reaction, saying that he would say 'no' to the candidate. Then, Zelda wanted to know who Irene's neighbour would vote for. Thus, Zelda's use of *so* shows that her request for this information was motivated by Irene's description of 'A1's indecision' (Schiffrin, 1987, p. 208).

In addition, *so* functions on other planes of talk. For example, it appears in the participation structure to organize transitions. Transitions occur when speaker and hearer make some adjustments to what each one is responsible for in the achievement of a particular task in the conversation. Examples of such tasks are turn-taking, linking parts of an adjacency pair and organizing discourse topics (Schiffrin, 1987).

Similar to Schiffrin's treatment of *so* is Redeker's (1990; 1991). She points out that *so* was one of the most frequent connectives in her data, together with *and*. She states that *so* was found in her data to occur between successive elements of events to mark 'summing-up or conclusion' as examples (5) and (6) illustrate.

(5) He talks to the girl and says that she has uhm her father has money due, uhm and so she gives him the sixty dollars asking if that would cover it. And so he leaves.  
(CTfmj)

(6) and he says you're gonna have to leave here. So he/he- kind of uhm kicks the guy out.  
(TLfdj)

(Redeker, 1990, p. 373)

Similar functions of *so* are discussed by Fraser (1990; 1999; 2009). He explains that the functions of the discourse marker *so* are based primarily on the interpretation of the prior context (1990, p. 393). To explain this, Fraser lists several examples:

(7a) Susan is married. *So* she is no longer single.

(7b) John was tired. *So* he left early.

(7c) Attorney: And how long were you part of the clean-up crew?

Witness: Five years.

Arty: *So* you were employed by G for roughly 5 years?

(7d) Son: My clothes are still wet.

Mother: *So* put the drier on for 30 minutes more.

(7e) Teenage son: The Celtics have an important game today.

Disinterested parent: *So*?

(7f) [A student upon encountering her professor for the first time in two weeks] Hi. *So* when are you leaving for Hawaii?

(7g) [Grandmother to granddaughter] *So* tell me about this wonderful young man you're seeing.

(Redeker, 1990, p. 393, examples 11a–g).

Fraser (1990) argues that *so* in the first example is a logical conclusion which could be compared to Schiffrin's warrant inference. *So* in the next example (7b) could be interpreted as a reasonable consequence (cf. Schiffrin's cause result). In (7c), *so* is interpreted slightly differently from (7a) as this one is considered to represent 'I conclude'. Similar to Schiffrin's motive action, Fraser interprets the use of the discourse marker *so* in the mother's turn (7d) as 'you should'. Fraser's (7e) has *so* with a questioning intonation as the father was asking his son why anything should follow from this, which could be interpreted as 'so what?' In example (7f), *so* could be interpreted as indicating that the student's question was based on something known earlier to both the student and the professor, i.e. the speaker/hearer shared knowledge. In the last example, *so* was interpreted as 'to continue' (Fraser, 1990, p. 394). I would like to argue that *so* in (7d) and (7g) in a sense have the same interpretations/functions. I agree with Müller (2005, p. 81) that they preface directive speech acts. In fact, (7f) is probably no different from the previous two examples. In (7f), *so* also prefaces a directive speech act.

Discussing the transitional or what he calls 'stand-alone' *so*, Raymond (2004) argues that participants use *so* in conversations 'to prompt action by a recipient' (p. 185). One of the ways in which this function is implemented is by 'invoking an upshot that is claimed to be available to a recipient' (Raymond, 2004, p. 211). In this study, the analysis of *so* suggests that it helps invite hearers to make connections between the preceding turn and the course(s) of action in which they participate. Raymond shows that the stand-alone *so* is deployed to manage overlapping contingencies imposed by the organization of turn-taking, sequencing, and the overall structural organization of conversation as a unit (cf. Raymond, 2004, p. 212).

Bolden (2006) analysed actual recorded social interactions between close familiars. Focusing on the deployment of the discourse markers *so* and *oh*, her goal in this paper was to describe discursive practices involved in showing engagement with the other party, or other-attentiveness. Bolden (2006) suggests that *so* was mainly used to preface other-attentive topics, i.e. topics related to other conversational partners/addressees as in the following example:

(8)

1 Leni: [Yeah that's (what I'm worrying about)

2 Leni: (thinking about).

3 (2.3)

4 ( ): ((sniff)) |

5 ( ): ((grunt)) ——

6 Leni: ***So yuh haven't been out tuh Palm Springs fer awhile.***

7 (.)

8        ↓Have you, I c'n tell you lost yer ta: n.

9 Sam: Not fer three weeks (now).

10 Leni: Yeah, ( ),

11 Sam: ( [ )

12 Joe: [Gee it's nice down in Escondido.

13 Sam: I wanna go when [it's convenient fer me.

14 Jim: *Yeh?*

15 Leni: Yeah, You [( )

16 Sam: [The weather aint right [I don' go.

((Leni, Sam, and Edith continue talking about Sam's tan.))

(2006, p. 670, example 5)

Bolden (2006) points out that *so* which prefates a topic as in line 6 ‘is initially formulated in the form of a statement about the addressee [...] that requires a confirmation or a disconfirmation’ (p. 671).

Arguing that the indexing inferential and causal connections is not the only function of *so*, Bolden (2009) applies conversation analysis methodology to examine a large corpus of recorded conversations collected from everyday talk to demonstrate another function of *so*, i.e. achieving incipient actions. The analysis primarily deals with *so* as ‘prefacing sequence-initiating actions’. It shows that *so* is used by speakers to signal an action which is about to take place, calling this status ‘emerging from incipency’ and is not defined by the preceding context. She concludes that *so* can also be used by speakers to establish certain actions to shape their ‘interactional agenda’. This use of *so* as a discourse marker is meant to create discourse coherence to achieve understanding (Bolden, 2009, p. 996).

Some previous studies have compared *so* as used by NSs of English and by NNSs. Anping (2002) investigated the use and the frequency of *so* in the English of Chinese learners of English. Her motivation for conducting this study was the overuse of *so* in Chinese EFL learners’ written English. The data in this study were based on a number of English corpora including the discourse of both native English speakers and learners of English in China. As anticipated, Anping found *so* was used more frequently by Chinese learners than by the native speakers of English. Anping (2002) believes that the likely reasons for this overuse are the Chinese learners’ lack of awareness of the stylistic differences between written and spoken English, the limited exposure the Chinese have to English, the style of instruction they have at school and L1 negative transfer.

Regarding the acquisition of the discourse marker *so* by NNSs, Hays (1992) examined the usage of English discourse markers by Japanese university students. He followed Schifffrin's (1987) model of discourse planes and concludes that Japanese learners of English acquire ideational markers such as *and*, *so*, and *or* earlier than markers on other planes such as *you know* and *well*, which require exposure to a speech community (Hays, 1992).

To my mind, one of the most comprehensive studies of *so* was Müller's (2005). Her qualitative and quantitative analysis of *so* as used by English NSs and German NNSs suggests that learners of English use the functions of *so* as do native speakers but with noticeably lower frequency (Müller, 2005). The functions of *so* in her study were organized on two levels, textual and interactional. On the textual level, *so* is a marker of result or consequence, a marker of a main idea unit, summarizing/rewording/giving an example, a sequential marker (coherently transitioning from one event to another in narrative), and finally *so* is a marker of a boundary between instructions and the beginning of the narrative (Müller, 2005, pp. 80–81). On the interactional level, *so* is as a discourse marker that indicates a speech act of questioning or request, a speech act of opinion, a marker of an implied result (i.e. directly addressing the hearer and challenging her/him to establish what the speaker is implying) (Müller, 2005, p. 86). The last function on the interactional level is *so* as a marker of a transition relevant place. Interestingly, qualitatively, all nine functions occur in the native and in the non-native data. However, the NSs of English use *so* almost twice as often as German NNSs.

Another comparative study which investigated *so* in non-native and native speakers of English was conducted by Buysse (2012). The data in this study consisted of informal interviews with 40 undergraduate Belgian native speakers of Dutch, 20 of them majoring in

Commercial Sciences and 20 majoring in English Linguistics. His data were compared to native English speaker data.<sup>21</sup> There were 10 functions of *so* in this study distributed over three domains: ideational, interpersonal and textual. All the 10 functions were used by both the language learners and the NSs. The results of this study showed that the language learners used *so* significantly more than the English NSs and that the students in English Linguistics used *so* slightly more often than those in Commercial Sciences.

In academic advice sessions conducted in an English as a lingua franca (ELF) context, House (2010) found that the occurrence of the discourse marker *so* was extraordinarily frequent. *So* was found to perform different functions, e.g. signalling causal and inferential connections between clauses and introducing a new topic. *So* is here employed as a deictic element that speakers use both to support the planning of their upcoming moves and to help them ‘look backwards’, summing up previous stretches of discourse. In another paper, based on the same data, House (2013) states that:

‘ELF speakers have acquired *so* when they need a self-monitoring filler to prevent conversational breakdown. *So* thus follows hesitation markers “erm”, “hmm”, or pauses of varying lengths. [...] These co-occurrences can be taken as evidence for the use of *so* as a self- prompting strategy to monitor a speaker’s output, marking his/her resumption of speech after hesitations and pauses both turn-initially and in mid-turn.’ (p. 62)

In sum, the review of the previous literature convincingly shows that the discourse marker *so* is associated with inference or consequence. In addition, the studies show how NNSs of English use it (Anping, 2002; Buysse, 2012; Hays, 1992; Müller, 2005) and when they

---

<sup>21</sup> This study used LINDSEI.

acquire it (Heys, 1992). In the following sections, the use of *so* by the Saudi English language learners is explored.

### 4.3 Functions of *so* in the SLC and LOCNEC

This section predominantly discusses examples of the discourse marker *so* in the data of Saudi English language speakers and NSs of English. But first, I provide some examples of the propositional or non-discourse marker use of *so*.

#### 4.3.1 Non-discourse marker functions of *so*

*So* in this category functions as an adverb of degree or manner (Müller, 2005), or as an intensifier (Tagliamonte, 2005). Let us consider the following examples:

- (9) <B> it was totally changed and also whatever noticed there is that and there are a lot of foreign people okay <\B>  
<A> yeah <\A>  
<B> there are a lot of Americans British people I do know a lot of foreigner people so and I didn't expect that I'm gonna use my English so I was *so* surp= surprised I didn't expect at all but when I went of course I expected to use it in in the hotel in restaurants <\B>  
<A> yeah <\A>  
<B> but not in the malls no= never think so I was *so* surprised and also my family did not know anything about English they don't speak English at all so I'm the only <\B>  
<A> so you did the whole job <\A>

(SLC-AR19)



(10) <A> oh yes (mhm) so you can see it from the[i:] . the plane </A>

<B> from the[i:] air yeah </B>

<A> oh yes </A>

<B> as we were sort of circling round to land </B>

<A> (mhm) </A>

<B> I thought it was fantastic it was a massive bridge </B>

<A> (mhm) </A>

<B> I couldn't believe it .. and the first thing that struck me was just how the whole place was </B>

<A> (mhm) </A>

<B> it was just *so* much bigger than England .. (erm) .. and then I was met by my cousin . and she drove me to where I was staying that was about a two hour drive away and I was just not used to the distances </B>

<A> (mhm) for them it's just </A>

<B> yeah </B>

<A> it's very small I guess </A>

(LOCNEC-E17)

The non-discourse marker use of *so* also occurs when a speaker uses it in a fixed expression as below:

(11) <A> (uhu) </A>

<B> and somehow they are they are treated they are kept in .. (ur) in a safe environment they always have (ur) health checks and *so* on </B>

<A> (uhu) </A>

(SLC-AR07)

Other examples from my data and from LOCNEC were *so* used as a preform:

(12) <A> but he might be better than you only maybe in the spoken </A>

<B> in the speaking yes in the speaking <foreign> yaghni </foreign> <\B>

<A> but in other skills </A>

<B> no I think I'm better than him I think *so* <\B>

(SLC-AR12)

(13) <B> so .. it may just be me dreaming again but <\B>

<A> oh come on <laughs> <\A>

<B> it would be lovely <\B>

<A> well you you need to have dreams <\A>

<B> (mm) <\B>

<A> even though well sometimes they're not you know <\A>

<B> yeah <\B>

<A> they're never going to come true but <\A>

<B> but <X> yeah . it gives you something to cling on to <\B>

<A> yeah <\A>

<B> I guess <\B>

<A> and I mean your dreams are quite realistic <laughs> <\A>

<B> I hope *so* <\B>

<A> well I think they are <\A>

<B> yeah I think I'd have to earn earn a bit of money before I . before I go over there

<\B>

<A> yeah <\A>

(LOCNEC-E11)

The last function of *so* in this group is when it occurs as a conjunction to express purpose, as below:

(14) <A> so your mom is doing this role </A >

<B> yeah even my: money you know ur to ur like when I got my s= my salary (mukafaty) I= have to you know like withdraw like three hundred or two hundred fifty and give to my mom *so* that she gives it to me x fifteen ten fifteen </B >

<A> yeah </A >

(SLC-AR09)

(15) <B> you know you just if if you want to work in an evening and you don't want to go out till about nine or half nine it doesn't leave you an awful lot of <begin\_laughter> time <end\_laughter> to <X> and relax </B>

<A> no it's true sometimes you just want to work until . say ten or eleven and then go and [ have a drink but </A>

<B> yeah .. you can't </B>

<A> obviously you can't </A>

<B> apparently somebody was saying that's because . erm it's deliberate . on campus because they think that maybe students want to go back and work afterwards . so it's to try to force them back to their rooms earlier *so* that they'll <begin\_laughter> go back and study <end\_laughter> but </B>

<A> do you think that's . what's happening </A>

(LOCNEC-E42)

Some of the most prominent functions of *so* as a discourse marker are discussed in the subsequent subsections.

#### 4.3.2 *So* indicating a result

The use of the discourse marker *so* to indicate result is one of its most well-established functions (Anping, 2002; Blakemore, 1988; Buysse, 2012; Fraser, 1990; Fraser, 1999; Müller, 2005; Schiffrin, 1987). In line with Müller (2005, p. 72), this function of *so* is not only syntactically but also semantically optional and is exemplified using an extract from the SLC data:

- (16) <A> ok but can you mention any special experience that really you felt that you are learning English </A>
- <B> yes there was actually . there was some classes I have (er) I had that there were many student who were better than me <\B>
- <A> right </A>
- <B> *so* I tried to be better than them or learn from them <\B>
- <A> (uhu) </A>

(SLC-AR12)

In this example, speaker AR12 is recounting his experience of English language learning. In his experience, there was a cause – *many student who were better than me* – and a result – *I tried to be better than them or learn from them*. The use of *so* here functions as a discourse marker of a local fact-based result in the ideational structure. To provide further argumentation that this *so* functions as a discourse marker, it could be said that the chronological sequence of the cause-result in this sentence could be inferred as a resultive relationship between ‘many student who were better than me’ and ‘I tried to be better than

them or learn from them' even without *so*. In fact, Biber et al (1999) state that '[M]ost resultive relationships, however, are not signalled with linking adverbials. Rather, cause and result are inferred from a chronological sequence' (p. 883).

The following example from the native speaker data also illustrates *so* indicating result:

- (17) <A> but would you like to stay in Britain or would you like to travel <\A>  
<B> no no I don't wanna stay in Britain <\B>  
<A> oh where would you like to go then <\A>  
<B> (erm) .. all over the place (erm) . I'd like to I went to Amsterdam in the summer  
<\B>  
<A> oh yeah <\A>  
<B> and I loved it there I really really loved it that was brilliant brilliant fun . (erm) .  
*so* I really liked that <X> somewhere like a city that's more cosmopolitan <\B>  
<A> oh yes <\A>  
<B> and there's more going on <\B>  
<A> (uhu) <\A>

(LOCNEC-E10)

In example (17), speaker E10 is speaking about her visit to Amsterdam. Because she had fun there, '*brilliant fun*', she liked it. Therefore, the same line of argumentation in the previous example works here as well. The discourse marker *so* marks '*I really liked that*' as a result of having fun in Amsterdam. Another example is provided by the SLC data:

- (18) <A> what state sorry <\A>  
<B> yeah California I went to California (er) and Las Vegas also and Nevada (erm)  
yeah . so yeah one day I was at the: at the mall (er) my with my brother we were at

the department store (er) a lady came up to me and offered me a package of perfumes I didn't want to buy anything . *so* I told my brother in Arabic as if I was (erm) asking asking him about the perfume I told him that I don't wanna buy anything just . do something I don't wanna but anything he said what am I supposed to do I said (er) do anything I just don't wanna buy (er) (er) all that was in Arabic so she stopped she stopped us and said (er) wait do you speak Arabic I said yes (er) why do you . she said (er) she said in Arabic yes I do speak in Arabic she turned around and (er) we: she didn't say anything she just turned around (er) we ran away <laughs> (er) <\B>

<A> embarrassing <\A>

(SLC-AR22)

In this example, speaker AR22 wished to avoid buying perfume but she did not want to speak in English. Here, *so* makes *I told my brother in Arabic* a consequence of the lady not understanding Arabic. Speaker AR22 assumed that the lady did not speak Arabic but she did.

The previous three examples illustrate resultive relations between propositions. In the following examples, there are instances of *so* marking a knowledge-based inference:

(19) <B> as usual I always do philosophical things and x <\B>

<A> yeah I I <\A>

<B> I don't wanna talk about this <laughs> <\B>

<A> I like this okay <\A>

<B> so that's it (er) when the artist draw or drew drew her: her portrait <\B>

<A> (uhu) <A>

<B> or whatever okay as as it is as she seems in the reality she: doesn't (laughs) she

didn't like it <\B>

<A> (uhu) <A>

<B> because it shows her her real picture <\B>

<A> (uhu) <A>

<B> *so* maybe she asked him to to drew to draw it again again or to make her beautiful something that then actually she she she shows she is showing pictures to fr= friends and that's it but this w= what I <foreign> yaghni </foreign> I actually I I suffered from this (er) <\B>

(SLC-AR48)

In this excerpt, speaker AR48 is telling a short story based on four pictures. He sees a lady looking at a portrait of herself, who is apparently dissatisfied with the image. She seems to want a more beautiful picture, presumably so that she can show it to her friends. In fact, the sequence in this example does not factually lead to the conclusion AR48 reaches. Rather, they work as background evidence which warrants his general conclusions (cf. Schiffrin, 1987, p. 205f). Another example from the native speaker data which illustrates *so* marking an inference is as follows:

(20) <A> if you want to know specific words <\A>

<B> (mm) <\B>

<A> like eating <begin\_laughter> I know you like eating <\A>

<B> I'll just go like that with sort of like knives and forks to my mouth <\B>

<A> yes I know but I mean <end\_laughter> <\A>

<B> it's it's very inadequate but it it does work doesn't it <\B>

<A> yes I know but I mean you you can't do that for: [ for twelve months

<begin\_laughter> <\A>

<B> <begin\_laughter> I could <X> I could I could go to a different restaurant every day <end\_laughter> and like pretend give me food ... <\B>

<A> they're going to say oh <\A>

<B> give me drink .. <XXX> I could sort of like go give me drink and like look drunk as well as I do it <\B>

<A> oh they might give you anything then <end\_laughter> <\A>

<B> <laughs> <X> I <X> probably end up with sort of like toilet cleaner or something like that <\B>

<A> <laughs> <\A>

<B> *so* maybe I should learn to speak French it would be a good idea <\B>

<A> yes I mean a few words just things you say when you meet people or <\A>

<B> (mm) yeah <\B>

(LOCNEC-E31)

Example (20) is another case in which *so* marks an inference drawn on the basis of the previous discourse. The woman in this extract was speaking about her travel experience to Belgium. She had some difficulties because she was unable to speak Flemish or French. She warrants her conclusion that she *should learn French* with how important she thinks to learn it. Another excerpt in this category is example (21):

(21)<B> first of all the way that they speak English which is great for me cos I speak no Dutch at all <\B>

<A> (mhm) <\A>

<B> although I must learn .. (erm) and .. there's <X> there there's a difference in . in attitude I find <X> (erm) the way my dad described it to me was .. that they look at .



somebody who can build a boat is as good as somebody who can be a famous pop star so it's like the[i:] idea of <X> sort of value of on . that they put on what people <X> <\B>

<A> (mhm) <\A>

<B> do <\B>

<A> (mhm) <\A>

<B> (erm) .. *so* I find that really interesting .. and there's there's the sort of .. I don't know cultural difference I mean that people sort of <X> Amsterdam as a sort of sex and drugs and rock and roll sort of thing <\B>

<A> yes (mhm) . (mhm) <\A>

(LOCNEC-E16)

In this interview, E16 chooses to speak about a country she visits and loves – Holland. She goes to visit her father who moved out and married there. She explains the Dutch attitude of egalitarianism. She uses that knowledge as background provided to her by her father as a ‘warrants’ to infer that this is *really interesting*. The discourse marker *so* in this example prefaces an inference from facts.

The last example of *so* illustrates an action-based causal relation on the action structure plane. On this plane, a motive for an action is presented by a speaker and the action is either performed by the speaker or the hearer. In (22), *so* prefaces an action that has just been accounted for by prior discussion between the interviewee AR14 and the interviewer. AR14 chooses to speak about her latest experience of looking for a job over her summer holiday.

(22)<B> I thought about doing something that is good for me that will help me in my future because as you know translation student <\B>

<A> yeah <\A>

<B> so I asked about a summer programme <\B>

<A> oh that's that's great <\A>

<B> yeah yeah and I looked on the internet I asked some friends and so on so I found a summer program that is specialised in tr= in training students <\B>

<A> yeah in which field <\A>

<B> in (er) it was in (er) <name of a hospital> in Riyadh <\B>

<A> oh that's nice<\A>

<B> yeah (er) it wasn't only for translators but for medical students also <\B>

<A> oh <\A>

<B> *so* (er) I applied and (er) I took an (er) exam (er) after about a month after applying <\B>

<A> yeah <\A>

(SLC-AR20)

Speaker AR20 is motivated to do something useful during her holiday, something which equips her with more experience as a translation student. Therefore, she decides to take action. Her action is prefaced by the discourse marker *so* the inclusion of which is justified by prior talk.

To sum up, Schiffrin (1987) provides more than one category under the discourse marker *so* indicating a result. I agree with Biber et al (1999) that '*so* does not always have such a clear role' (1999, p. 877). This is the case when analysing different sub-functions of *so* and the borderlines between these functions become indistinct. Müller (2005) points out that the boundaries between the categories of *so*, i.e. fact-based result and knowledge-based inference,

are not always clear. Therefore, to be consistent in the analysis of the functions of *so*, all of these different categories are here grouped under ‘*so* indicating a result’.

#### 4.3.3 *So* marking a main idea unit

The second category is *so* as a marker of main idea unit (Müller, 2005; Schiffrin, 1987). To set out how this form of *so* functions, Schiffrin (1987) believes that ‘... stories can be told not just to report a specific experience, but to warrant a general point being made by the speaker. The entry to such a story can be marked with *because*, and the return to the main point marked with *so*’ (p. 195). In the same vein, Müller (2005) categorizes *so* as the marker of a main idea unit when the speaker returns to the main idea/topic after a digression. The following example illustrates *so* referring to a main idea unit:

(23) <B> I actually like teaching for my father who is an English teacher <\B>

<A> okay <\A>

<B> and I got the thing from him <\B>

<A> yeah <\A>

<B> *so* I really really do like English <\B>

<A> (uhu) <\A>

<B> or let’s say teaching English <\B>

(SLC-AR11)

In this example, AR11 opts to speak about his experience of English language teaching. He says he likes teaching. Then he elaborates on this by providing the reason why he does so – essentially, because his father is an English language teacher and he tries to emulate him. AR11 then comes back to say that he really does like teaching English. It could be argued that *so* in this example indicates a result. However, at the beginning of the extract, the speaker

used the adverb *actually*, which is used to admit something or emphasize it (cf. Aijmer, 1996). The discourse marker *so* is followed by the adverb *really*, used in a similar manner to *actually* to emphasize what is going to be said. Therefore, I argue that *so* in this example marks the main idea unit. Another similar example is from the native speaker data given in extract (24). Earlier in the discussion, speaker E02 speaks about the disadvantage of studying at a university which is close to his home. He admits that he goes home at the weekend which he believes is unusual. However, he thinks that visiting every now and then could help him to move out from home.

(24) <B> <XX> I wanna do it gradually <\B>

<A> uhu <\A>

<B> get used to not living at home <\B>

<A> (ah) yes (mhm) <\A>

<B> cos like I've<?> never done anything like it before <\B>

<A> (mhm) <\A>

<B> *so* .. I wanna do it gradually <\B>

<A> you have to: to get used to it <\A>

<B> yeah get used to it <\B>

<A> yeah I understand <\A>

(LOCNEC-E02)

Speaker E02 has explained how it would be helpful to study at a university which would make it possible for someone to visit home over the weekend. He says that he wants to move out but he wants to do it gradually. Then he carries on saying that he is not used to living away from home and then repeats *so .. I wanna do it gradually*. A contextual clue that *so* is a discourse marker is that here it is followed by an unfilled two second pause and thus

functions as a main idea unit marker. Another example in this category is (25). Following the thread of the discussion here, it could be noted that the interviewee E03 mentions that he is the only child at home living with his mother. He describes the nature of the life he had. Then he brings in more life details. However, he rolls back to the point at which he began – *brought up like an only child* – prefacing that with *so* which functions as a marker of returning to the main idea unit.

(25) <B> yeah but I'm an only child so <\B>

<A> ah yeah <\A>

<B> it's like getting fussed over all the time <\B>

<A> <XX> more protected <\A>

<B> yeah <laughs> <\B>

<A> do you do you .. I mean would you have liked to have a brother or a sister <\A>

<B> erm well . the situation my dad was married before he met my mom <\B>

<A> oh yes <\A>

<B> so I've got two half sisters and a half brother <\B>

<A> uhu <\A>

<B> but they've never lived with me <\B>

<A> (mhm) <\A>

<B> so I've been I've got those which is nice cos they're not . they are there <\B>

<A> (mhm) <\A>

<B> but I've never lived with them <\B>

<A> (mhm) <\A>

<B> *so* I've been brought up like an only child but I've still got these in the background <\B>

<A> (uhu) </A>

<B> so that's nice </B>

(LOCNEC-E03)

#### 4.3.4 *So* introducing a new sequence

The function of *so* introducing a new sequence basically enables the speaker to accomplish coherence ‘between successive elements in a chain of events’ (Redeker, 1990, pp. 373–374). Müller (2005) acknowledges this function of *so* as it occurs 140 times in her data. Dictionaries define this function of *so* as introducing the next event/part in a story (Cobuild English Dictionary, 1995; Oxford Advanced Learner’s Dictionary, 2000). Buysse (2012) argues that the sequential *so* can be distinguished from the resultive *so* as the ““sequential” *so* starts a new sequence within the turn, whereas “resultative” *so* brings a sequence or a turn to a closure’ (p. 1773). Example (26) relates to an interviewee who speaks about one of his life changing experiences. He tells his story of applying to two universities. He mentions that he was accepted at both. He uses *so* in the last turn in this excerpt, which in fact has neither any clear ideational nor rhetorical relation (Redeker, 1990, p. 396).

(26)<B> it’s no problem I I an experience you have had which has taught you an important lesson </B>

<A> ok </A>

<B> describe the experience and say what you have learnt from it I have an experience a previous experience with the studying at university </B>

<A> ok </A>

<B> and that’s a very a a point that has changed my course in life </B>

<A> all right </A>

<B> I applied for <name of a university> when I finished high school I applied for

<name of a university> I got accepted </B >

<A > (uhu) </A>

<B> in engineering and then I applied for (er) <name of a university> and I got directly a= accepted </B >

<A> (uhu) </A>

<B> *so* I chose <name of a university> I didn't do very well there study for four years ... and after maybe years my GPA was dropping down </B >

<A> (uhu) </A>

(SLC-AR04)

The same kind of sequence takes place in extract (27). This interviewee is a native speaker of English who speaks about her year of teaching in France. The discourse marker *so* serves to introduce new sequences.

(27) <B> cos they're gonna know I'm English and <\B>

<A> oh come on <laughs> <\A>

<B> and they're gonna say oh you're just making mistakes which sounds really bad and <laughs> <X> <\B>

<A> they wouldn't say that <\A>

<B> no they wouldn't I think it's just . you have nerv= you're nervous and <\B>

<A> yeah <\A>

<B> it makes you have more complexes I suppose .. so (erm) . but . like it really sort of worked out well I had a flood in my flat over Christmas <X> pipe and it could have been a disaster but it actually worked out to be the best thing because it meant that people friends in the building like who<?> weren't friends but people my neighbours who I'd not really much contact with they came out to help and I made

my friend <\B>

<A> oh yes <\A>

<B> Olivia she was brilliant and she was . so it sort of helped and it made me like find people meet people and then they knew who I was and they'd come round and ask for help with their English . essays they happened to be doing English <X> and business <\B>

<A> oh yes <\A>

<B> so it worked out quite well <\B>

<A> (mhm) <\A>

<B> *so* (er) and then I met some people who worked at my school the[i:] surveillant the <\B>

<A> (mhm) <\A>

(LOCNEC-E08)

The final example below is from the picture-based story. At first, speaker AR32 finds it hard to describe the picture and make a story out of it. The interviewer tries to help by asking her some questions. Then AR32 starts her story with a scene. She introduces the second scene with *so*.

(28) <A> okay (erm) okay let's move to the you know other paper (er) here you have four pictures (erm) can you you know can you make up a story of your own you know describe these pictures and make up a story perhaps <\A>

<B> you know it's very difficult <\B>

<A> <XX> look at it <\A>

<B> very difficult to make up one (er) <\B>

<A> what do you see you know <\A>



<B> (erm) there are there was a women walking down the street and she saw an artist <\B>

<A> okay <\A>

<B> *so* she came up to him and asked him (er) if he if he could draw her then he said yes <\B>

<A> (uhu) <\A>

(SLC-AR32)

#### 4.3.5 *So* marking transition

In her participation framework, Schiffrin (1987) asserts that the discourse marker *so* can function in the organization of transitions. She believes that transitions in participation have two characteristics: the speaker shifts responsibility to hearer and/or there is a shift from the speaker to hearer, either of which is dedicated to the fulfilment of a certain interactional task. Sacks et al (1974) state that '[o]nce a state of talk has been ratified, cues must be available for requesting the floor and giving it up, for informing the speaker as to the stability of the focus of attention he is receiving' (1974, p. 697). This is one of the functions *so* fulfils in the discourse. *So* can indicate the speaker is willing, or more directly, encouraging the addressee to take the floor (Lam, 2010, p. 670). This function has different labels in the literature, for example in Müller (2005) it is 'marking implied result' and in Bysee (2012) it is termed a 'prompt'. This non-prefatory *so* usually is illustrated by the following examples from my data:

(29)<B> but my my friend's got a room twice the size of mine and it is freezing <\B>

<A> and and the heating is not working at night *so* <\A>

<B> is it not <\B>

<A> no if you wake or if you go to bed say at two o'clock .. freezing . if you wake up

at four five it's freezing .. it's terrible <XX> <\A>

<B> our heating our heating is on all night .. we we've got controls on our radiators  
but it doesn't actually get shut down at any central point <\B>

<A> that's not fair <\A>

<B> <begin\_laughter> so that's lovely <end\_laughter> <\B>

<A> we're all complaining because it's really cold because of the wind and <\A>

<B> I can't believe ... it's the last . you know middle of winter the last thing you  
wanna be doing is turning the heating off <laughs> <\B>

<A> I had no idea that it was not everywhere I thought it was like that everywhere  
<\A>

<B> no .. our heating is <X> <\B>

<A> are you privileged or something <laughs> <\A>

<B> must be <laughs> must be I think it's to make up for all the gaps in the windows  
where the windows don't quite meet when they shut all the wind comes whistling  
through <laughs> <\B>

<A> oh right perhaps . perhaps <\A>

<B> my my friend is on the side of the tower <X> that's being really buffeted by the  
wind at the moment because of the direction of the winds coming in <\B>

<A> (mhm) <\A>

<B> and her windows don't shut properly at all and her curtains are just billowing  
the whole time with all this wind it's managed to [ managing to <X> through her  
windows <laughs> <\B>

<A> oh no .. yes I can imagine that the heating is certainly .. very important <\A>

<B> yeah yeah <\B>

<A> if it's that bad <\A>

<B> I mean it I mean it it you can't counter it at all you know <\B>

<A> (mm) <\A>

<B> the the heat just goes out completely because it's such a big room as well *so*

<\B>

<A> can't you mend it <\A>

<B> I don't know I don't know I don't think they'd bother they think Bowland tower is just generally too old <laughs> for them to worry about <\B>

<A> well there's students living there they could do something about it <\A>

(LOCNEC-E15)

In excerpt (29), there are two instances of *so* marking transition. The first instance is provided by the interviewer at the beginning of the excerpt. Of course the analysis of discourse markers in this study focuses on the participant/interviewees' tokens but there is a reason for discussing the interviewer's use of *so*. I would argue that the use of *so* by the interviewer is probably different from that of *so* employed by the interviewee in this example. The interviewer uses *so* not only to encourage the interviewer to take the turn but also to infer what he/she means with his utterance prior to *so*. The second instance of *so* is by the interviewee, speaker E15. Her use of *so* occurs to let the hearer recognize the conclusion presented in her discussion. This *so* is a sign of willingness on the part of the speaker for the hearer to participate/take part in the discussion.

The same line of argumentation fits the following example. *So* at the end of the E12's turn cannot be interpreted as resultive. Rather, it is another instance of *so* which requests the hearer's attention in participating in the discussion.

(30) <A> oh so you already have to teach <\A>

<B> oh you take over the class from the teacher and you have to plan everything

<\B>

<A> oh right <\A>

<B> you have to do schemes of work and lesson plans and <\B>

<A> (uhu) <\A>

<B> and everything do everything properly *so* <\B>

<A> and so you you just teach English and religion <\A>

<B> no that's you pick two subjects [to specialise in <\B>

<A> uhu <\A>

(LOCNEC-E12)

The next example from the Saudi data illustrates *so* marking a transition:

(31) <B> yeah I agree with you and also the prices </B >

<A> yeah </A >

<B> prices were so so high </B >

<A> it is </A >

<B> I donno why even in taxi ev= in everything so </B >

<A> yeah </A >

<B> so </B >

<A> yeah </A >

<B> I've noticed that *so:*<sup>22</sup> </B >

<A> yeah that's you know bad result </A >

<B> (mhm) (mhm) </B >

---

<sup>22</sup> The use of a colon at the end of a word indicates that the last syllable is lengthened.

<A> of big cities </A >

<B> the problem the problem is the prices I think that why the Emirate people do do  
not live in Dubai </B >

(SLC-AR19)

Speaker AR19 tells the interviewer about her visit to Dubai. She mentions a couple of issues related to Dubai. However, she not only uses the discourse marker *so* to indicate transition but also lengthens it. I would argue that this *so* demonstrates the speaker's desire to relinquish her turn and hand the floor to the hearer. It is a point at which the speaker wants to exchange the turn with the hearer.

Most of the examples within this function have *so* in utterance-final position. However, some interesting instances in my data show that this is not always the case. There are some instances where the interviewee uses the discourse marker *so* 'deliberately' to give up the turn/floor. I say 'deliberately' here because the speaker provides an additional clue to the context, for example using the word *yeah* after *so*. Some of examples from LOCNEC and the SLC are given here:

(32) <B> I mean <X> I picked up a bit of Swahili <X> always keen to teach you <\B>

<A> oh great <laughs> <\A>

<B> yeah they want to teach you new words <laughs> and things so <laughs> <\B>

<A> (uhu) ah oh that's interesting <\A>

<B> *so*: yeah <\B>

<A> would you like to go and live there <\A>

(LOCNEC-E12)

(33) <B> the standard of living is excellent it's just the prices <X> </B>

<A> very high </A>

<B> very high indeed </B>

<A> (mm) especially for alcohol I think </A>

<B> nine pounds . for a pint of beer </B>

<A> . nine pounds </A>

<B> yeah about five pounds for a hot chocolate or a coffee . so wherever you went  
you knew you <X> be spending quite a lot of money *so* .. yeah <laughs> </B>

<A> were you travelling on your own or </A>

<B> (er) yeah but I met some people on the way so <X> </B>

(LOCNEC-E38)

The previous two examples (32) and (33) from the native speaker data illustrate the case in which the discourse marker *so* is used to indicate the speakers' willingness to give the floor to the interviewers. In extract (33) it can be noted that there is an empty pause for two seconds; however, when the interviewer does not take over, the interviewee carries on with *yeah* and laughs. In this case, it possibly appeared to the interviewee that the short pause after *so* was not enough of a clue to the interviewer to take over, although the occurrence of *yeah* after the discourse marker *so* is a contextual clue which shows the interviewee's attempt to give up the floor.

The following excerpt is a similar example from the native speaker data in which *yeah* occurs after the discourse marker *so*, marking transition:

(34) <B> <laughs> yeah it is (er) I think she: she doesn't have confidence in herself </B>

<A> I think so </A>

<B> because you know yeah (erm) </B>

<A> so she looks different </A>

<B> yeah she they look different (er) (er) everything is different the facial expressions and the hair the body everything is different *so*: yeah (erm) what else (erm) that's pretty much it </B>

<A> try try harder <laughs> </A>

<B> she went home and hung it up at the wall </B>

<A> so so she is pleased now </A>

<B> yeah she is pleased </B>

(SLC-AR23)

In example (34), speaker AR23 describes the lady in the picture description, which is the last part of the interview. It seems that the interviewee has not got much to say. She uses the lengthened *so* at the end of her turn to inform the interviewer that she is willing to exit the turn and give the interviewer the floor. What is interesting about this example is that speaker AR23 not only uses this function of *so* to notify the hearer that she is now awaiting her to take the turn, but also says clearly that she has nothing more to say. The interviewer's turn '*try try harder*' comes as verification that *so* has done its job of marking transition. I would argue that what comes after *so*, '*yeah (erm) what else (erm) that's pretty much it*', is another clue in this utterance that *so* marks transition.

#### 4.3.6 *So* introducing a turn/discourse

Another function of *so* is to start a turn. The reason this function comes last in this section is because it appears with the least frequency in the data; a total of two tokens/times in B turns<sup>23</sup>

---

<sup>23</sup> The <B> turns are the interviewee's whereas the <A> turns are the interviewer's.

in the SLC and LOCNEC. However, this function of *so* is used in A turns at least once by almost all interviewers. I argue that this has to do with the nature of the interview or, as Buysse (2012) puts it, the fact that ‘the specificity of the environment in which *so* takes on this function heavily restricts the frequency of “discourse initial” *so*’ (p. 1772). Along the same line, Buysse’s (2012) study found that this function occurred 15 times in the whole corpus, 13 in his non-native data and only two in the native speaker data. He took this frequency in his non-native data to be ‘relatively high’ and explained it in terms of the way in which the interviewers prompted the interviewees to start the conversation (p. 1772).

A similar function of *so* is mentioned in another study by Johnson (2002). She argues that when *so* prefaces questions, it acts as a topic developer: ‘We could describe the function of the *so*-prefaced questions as *topic developers* or *topic sequencers* whilst at the same time *marking* the discoursal act and topic boundary’ (Johnson, 2002, p. 103).

Using a different label for the same function, Müller (2005) calls this function a ‘boundary marker’ (p. 80) which signals the boundary between instructions and the beginning of the narrative. She claims that this function of *so* has not received any attention from Schifffrin, Blakemore or Fraser. Defining this function, Müller (2005) states that ‘*so* does not relate propositional ideas [...] [r]ather, it structures the spoken material into types of speech [...] and thus can also be seen as functioning at the textual level’ (p. 81).

In the following examples from the native speaker data in LOCNEC (35) and in the SLC (36), the speakers started their turn with *so* as their first word. In both excerpts, the preceding two lines by the interviewers act as a short prompt to start talking:



(35) <A> oh yes .. <X> . I'd like you to have a look at that . so four pictures it's one story  
 and I would like to tell <X> well you to tell me that story <\A>  
 <B> *so* the story <\B>  
 <A> yeah <\A>  
 <B> all right you have an artist drawing . a model or a lady <\B>  
 <A> (mhm) <\A>

(LOCNEC-E34)

(36) <A> I'll have to say thank you very much for you and for participating in this (er)  
 data . spoken data collection yes carry on <first name of interviewee> go ahead <\A>  
 <B> *so* what I'm gonna talk about is (er) .. a movie die hard <\B>  
 <A> okay <\A>

(SLC-AR50)

#### 4.3.7 Unclear instances

There were some cases in my data which could not be classified. They amounted to 62 tokens in both corpora. The majority were not classified because the context was not sufficiently clear. This could be due to unfinished utterances, as in example (37), or all the words were available but were still hard to classify, as in (38), or there were unintelligible passages, as in example (39).

(37) <B> yeah yeah and the character pretending to: . (erm) . to be crippled was a very  
 good cov= cover cos nobody could possibly suspect him of being capable <\B>  
 <A> (mhm) <\A>  
 <B> of all those things <\B>  
 <A> (mhm) <\A>

<B> cos you know physical disability *so* it was <\B>

<A> and the way he described himself .. (erm) . you know as being somebody: a bit  
.. I wouldn't say stupid <\A>

<B> (mm) <\B>

<A> but <\A>

<B> yeah he was quite ready to put himself down and and help generate that picture  
of himself <\B>

<A> (mhm) <\A>

<B> yeah that it is a very clever film <\B>

(LOCNEC-E15)

(38) <B> they say you're been out almost all of the time and your home just like a hotel  
you know </B>

<A> right </A>

<B> you know you just came to come to sleep and go out </B>

<A> right </A>

<B> so what I think that I'm I'm right cos (ur) you I don't smoke and I go out for  
<XX> or hubble-bubble *so* or such </B>

<A> (uhu) </A>

<B> things that way most of the time I play football and be with my friends and that  
bother them I donno why and I don't see that I'm wro= in a wrong side I don't see  
that </B>

(SLC-AR09)

(39) <A> and (er) what did you do there <\A>

<B> I was working as an assistant so I was teaching . nine hours in two schools in

Germany . a Gymnasium <\B>

<A> secondary le= level or just university <\A>

<B> yeah a a grammar school and a: a comprehensive . school they were just across  
the road from each other <\B>

<A> (uhu) <\A>

<B> so they shared me <\B>

<A> (mhm) <\A>

<B> *so* <XX> <\B>

<A> and did you enjoy that <\A>

<B> yeah I did actu= because er actually it was the it was the best year . because you  
I was getting paid and I actually wasn't doing an awful lot at all <\B>

(LOCNEC-E42)

#### 4.3.8 Summary of the functions of *so*

In the previous sections, some previous studies of the discourse marker *so* were reviewed and explained with examples from the literature to show how the discourse use and the propositional use of *so* are distinguished and as a reference points for the examples from my data. The main distinction between the discourse use and the propositional use of *so* is that *so* in the latter cannot be omitted without leaving the utterance ungrammatical/incomplete. As for the non-discourse marker *so*, there were examples where *so* occurred as an adverb of degree or manner as in ‘I was *so* surprised’, a fixed expression as in ‘and so on’, a preform as in ‘I hope so’, and finally it was used as a conjunction to express purpose as in ‘so he does it *so* that she looks beautiful’.

As a discourse marker, *so* is commonly found to have five functions. The first function is *so* indicating a result. It may also be used to introduce main idea unit as a recap of something that has been said or discussed earlier. Another function for *so* is to help the speaker to introduce successive elements in a chain of events. This function is different from *so* introducing a turn. In the latter, *so* introduces a turn which is new/does not relate to a previous discussion and mostly occur turn initial as a discourse opener. As a transition organiser, *so* helps the speaker to give the floor to the listener.

#### 4.4 Quantitative results for functions of *so*

In this quantitative section, I first provide an overview of *so* as a linguistic item. Table 4.2 shows an overview of the main functions of *so* and their corresponding abbreviations in this study. All the native speakers were found to use the discourse marker *so* whereas only 42 participants out of 50 did so. This means that there were eight participants in the SLC corpus who did not use a single *so* as a discourse marker (for more information about use of discourse marker *so* per each participant in SLC, see Appendix 11).

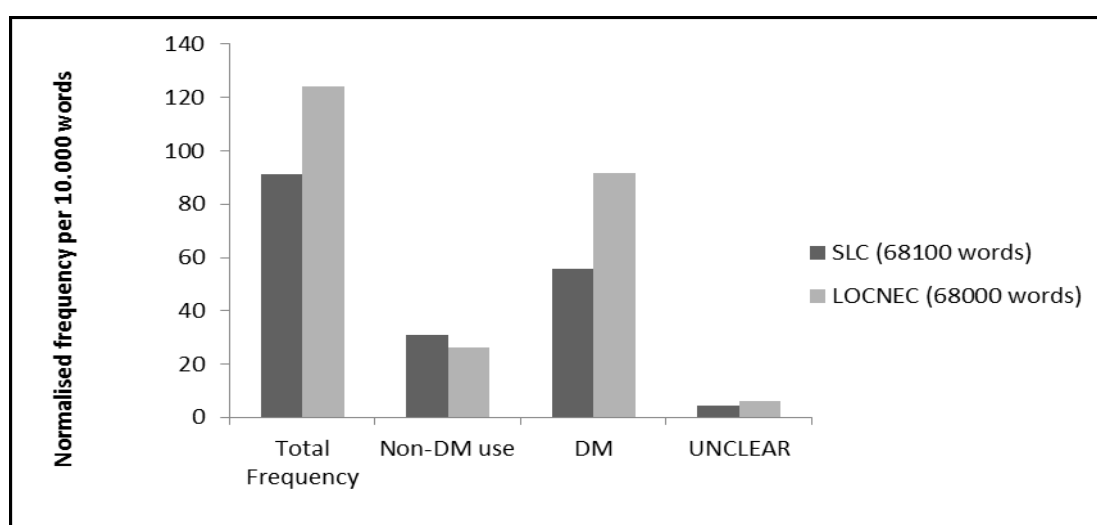
**Table 4.2 Functions of *so* and their abbreviations**

FUNCTION	ABBREVIATION
<i>So</i> indicating a result	Result
<i>So</i> marking main idea unit	Main idea
<i>So</i> introducing a new sequence	New seq.
<i>So</i> marking transition	Transition
<i>So</i> introducing a turn	Introductory

Table 4.3 and Figure 4.1 show the distribution of the raw frequency of *so*, the non-discourse marker (non-DM) *so*, the discourse marker (DM) *so* and unclear instances in the two corpora. The non-discourse marker *so* is used more by the Saudi speakers than by NSs. However, when the instances of *so* as a discourse marker in the two corpora were analysed, it was found that there were 624 instances of use by NSs of English and 380 by Saudis. To test the significance of the frequency of *so* between the SLC and LOCNEC, log likelihood test has been performed. The test showed that the difference of the use of discourse marker *so* in the two corpora was statistically significant at  $p < 0.0001$ . The discourse marker *so* was used almost twice as often by the NSs as by the Saudi NNSs. There were 42 instances of unclear instances in the native speaker data compared to 31 instances in the Saudi data.

**Table 4.3 Numbers of tokens of *so* in the SLC and LOCNEC (frequencies per 10,000 words in parentheses)**

	Total frequency	Non-DM use	DM use	Unclear
SLC (68,100 words)	662 (91.3)	211 (31)	380 (55.8)	31 (4.5)
LOCNEC (68,000 words)	845 (124.2)	180 (26.4)	624 (91.7)	42 (6)

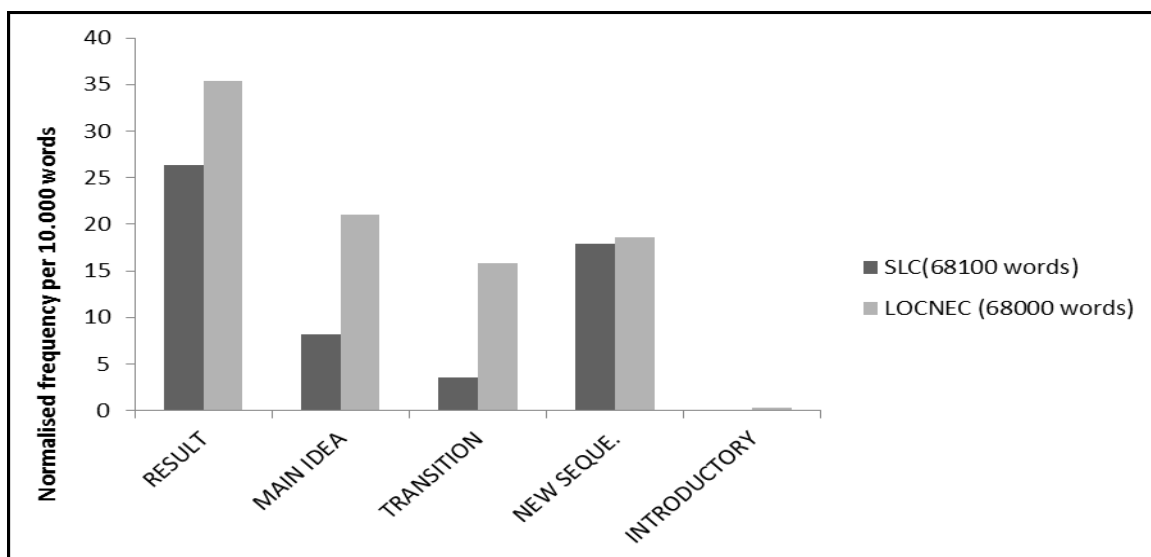


**Figure 4.1 Distribution of the linguistic item *so* – SLC versus LOCNEC**

Regarding the discourse marker functions of *so*, from Table 4.4 and Figure 4.2, it can be observed that the most frequent discourse marker function used by both NSs and Saudis was *so* indicating a result, with the NSs employing it more frequently than the Saudis. Saudi speakers used *so* indicating result 180 times or 25 times per 10,000 words, whereas the NSs used it 241 times or 35 times per 10,000 words. Such normalised frequency was used to provide an accountable measure of relative frequencies which can be replicated and therefore allows valid cross-corpora comparisons to be carried out (Pichler, 2010). Frequency in this study was obtained by dividing the total number of tokens of discourse markers by the total number of words of the corpus then this number was multiplied by 10,000. *So* marking the main idea unit was the second most frequent function employed by the NSs although not by the Saudis. The former used it almost two and a half times more than Saudi speakers (143 tokens, or 21 times per 10,000 words versus 56 tokens, eight times per 10,000 words). NSs of English used *so* marking transition 108 times, or 16 times per 10,000 words while Saudis only used *so* with this function 26.4 times, or four times per 10,000 words. These last three functions achieved high statistical significance ( $p < 0.001$ ). NSs used *so* to introduce a new sequence five times more than Saudis with 127 tokens versus 122. The function which occurred the least was *so* introducing a turn; there were only three tokens in both corpora.

**Table 4.4 Functions of *so* in the SLC and LOCNEC (frequencies per 10,000 words in parentheses)**

	Result	Main Idea	Transition	New Seq.	Introductory
SLC (68,100 words)	180 (26.4)	56 (8.2)	25 (4)	122 (17.9)	1 (0.14)
LOCNEC(68,000 words)	241 (35.4)	143 (21)	108 (15.8)	127 (18.6)	2 (0.29)

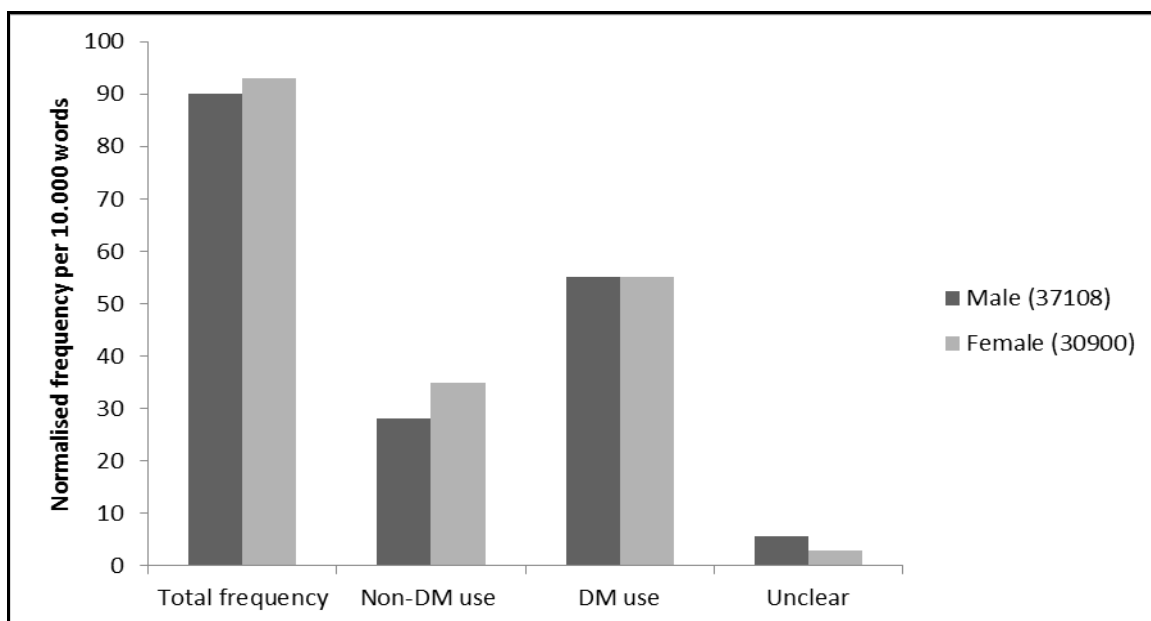


**Figure 4.2 Distribution of discourse marker functions of *so* – SLC versus LOCNEC**

Regarding gender in the SLC corpus, as Table 4.5 shows, the males were found to use *so* more than the females, with 334 versus 288 tokens. In addition, the males used the discourse marker *so* more than the females. However, no function was found to reach statistical significance. Figure 4.3 represents this graphically.

**Table 4.5 Number of tokens of *so* in the SLC according to gender (frequencies per 10,000 words in parentheses)**

	Total frequency	Non-DM use	DM use	Unclear
Male (37,108 words)	334 (90)	104 (28)	204 (55)	21 (5.6)
Female (30,900 words)	288 (93)	107 (34.6)	171 (55)	10 (3)



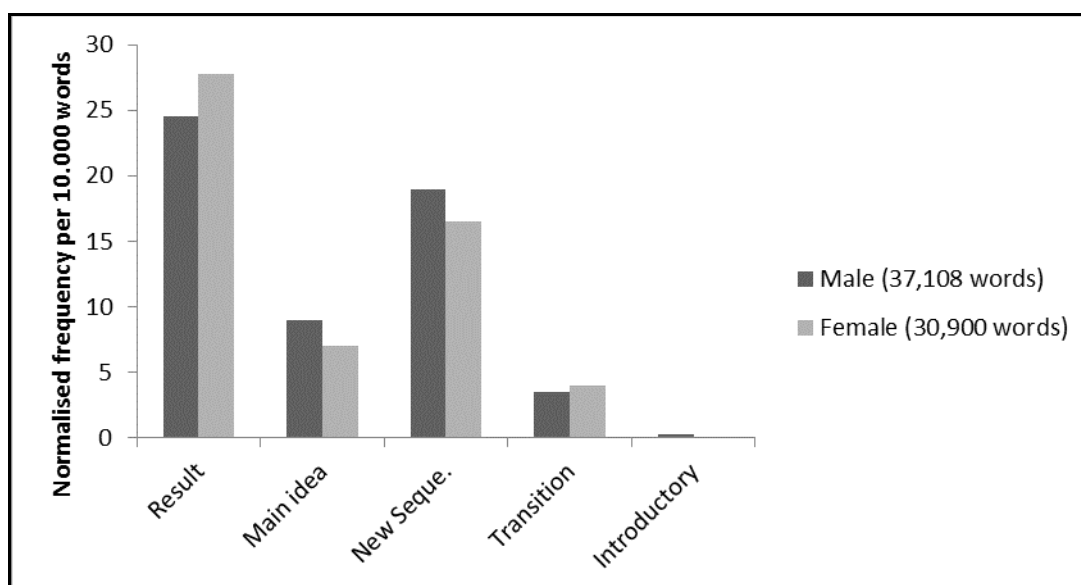
**Figure 4.3 Distribution of the linguistic item *so* in the SLC according to gender**

In terms of the functions of the discourse marker *so*, we can see that the use by the two genders is almost exactly the same for each function (Table 4.6). None of the differences is statistically significant. However, because this is the first study ever to include spoken data from Saudi females and males are segregated from females in the Saudi education system, this comparison needs to be better established.

**Table 4.6 Functions of *so* in SLC according to gender (frequencies per 10,000 words in parentheses)**

	Result	Main Idea	New Seq.	Transition	Introductory
Male (37,108 words)	91 (24.5)	34 (9)	70 (19)	13 (3.5)	1 (0.26)
Female (30,900 words)	86 (27.8)	22 (7)	51 (16.5)	12 (4)	0





**Figure 4.4 Distribution of discourse marker functions of *so* in the SLC according to gender**

Further analysis of the quantitative results points to another conclusion. By counting down the number of discourse markers uses of *so* per speaker in both corpora, it can be observed that there are some results which require questioning. In LOCNEC, it was found that the use of the discourse marker *so* ranged from the maximum of 42 tokens by one speaker to a minimum of eight tokens by another speaker. On the other hand, in the SLC data, the analysis revealed that there was one speaker who used *so* as a discourse marker 43 times, which is almost similar to the NS data in LOCNEC. However, there were two speakers who used *so* only once, five speakers who used *so* twice and four speakers who used it three times. This gives a total of 10 speakers out of 42 from the SLC corpus who used *so* with very low frequency.

What led to this observation was the distribution of frequency of using *so* in the native speaker data as a reference. NSs were all found to use *so* as a non-discourse marker and as a discourse marker. In fact, eight Saudi participants did not use any discourse marker function

of *so* at all. Therefore, the discrepancy between the highest and the lowest frequency in the SLC data raises some concerns regarding the mastery of the use of *so* and familiarity with its functions among Saudi speakers (six Saudi speakers used *so* as a linguistic item). This suggests the need to investigate the textbooks which all these participants studied to compare what is found in their authentic discourse – the spoken corpus – to what they have been exposed to/taught.

#### **4.5 Comparison of findings in the corpus and textbooks**

This section examines the discourse marker use of *so* in the textbook database and compares that to the findings from the SLC. In the textbooks, there were occurrences of *so* functioning as a discourse marker. When analysing the textbooks year by year individually, it was found that *so* was not represented in the first year of introducing Saudi students to English, i.e. the sixth year of primary school. A similar result was found for the textbook in the first year of intermediate education. The representation of *so* was introduced from the second year of intermediate school onwards until the third year of secondary, which is the last year before college. The total number of the discourse marker uses of *so* was 75 tokens performing different functions. Following are some examples from the textbooks which represent all functions introduced to students at school.

There were occurrences of *so* as a discourse marker. Example of *so* indicating a result is given in the following excerpt:

(40)Noura: Guess what we did last week, Sara?

Sara: What?

Noura: We rented a beach house and stayed there for the weekend.

Sara: Great! What did you do?

Noura: Well, I walked on the beach, watched the sunset and went sailing. Oh, and I collected many nice shells, as well.

Sara: How nice!

Noura: Yes, and my brother took part in a jet ski race.

Sara: *So*, you really had fun?

Noura: Yes, I did. What about you? I'm sure you had a good time last weekend, as well.

(Second year intermediate, p. 22)

Another discourse marker function found in the textbooks was *so* marking a main idea unit.

The following example occurred in the third year intermediate textbook.

(41) Student 1: I wish I had Bill Gates' brains?

Student 2: Why?

Student 1: If I had his brain, I'd go to an institute of technology.

Student 2: There're so many. Which one do you want to go to?

Student 1: I'd love to go to the one in India. It's famous for information technology.

Student 2: Then you'd become someone special in the future.

Student 1: It sounds fantastic, but it seems impossible.

Student 2: No, it's not. I just read about a Saudi boy who invented a computer programme in mathematics. *So* you can make your dream come true if you study hard.

(Third year intermediate, p. 41)

Regarding *so* introducing a new sequence and *so* marking transition, there was no occurrence of them in all the seven textbooks. *So* introducing a turn was used once in the spoken data and interestingly was found only once in all the textbooks. It occurred in the third year secondary textbook:

(42) John: Ahmad, are you any good at fractions?

Ahmed: Not bad, Why?

John: Well, what's half a half?

Ahmed: A quarter.

John: O.K. *So* what do you get if you add a half and a quarter?

(Third year secondary, p. 2)

Thus, it can be said that the ratio of the occurrence of the discourse marker *so* in the textbooks is far lower than in the SLC. Given the popularity of *so* in spoken discourse, it is rather surprising that the writers of textbooks seem to make little effort to explain how it is used in ways that will help students comprehend it and become more familiar with its use. There was no separate lesson dedicated to the function of *so*, i.e. how it can be used. Even though the representation of *so* as a discourse marker in the textbooks is fairly scarce, Saudis still use them to fulfil all the functions used by NNs. This might relate to the availability of this discourse marker in Arabic.

Saudi speakers use the discourse marker *so* in its various functions regardless of the under-representation of this marker in the textbooks to which they have been exposed. I argue that this has to do with the similarities between *so* in English and *fa* in Arabic. For example, functions of *fa* as a marker indicating a result were found in Saeed and Fareh (2006). They argue that the following example of the Arabic connective *fa* signals a result:

أحب أحمد المسرح فأبدع فيه (43)

‘ahabba ‘ahmadu l-masraha *fa* ‘abdaca fih.

Ahmad loved theatre and *so* he excelled in it.

(Saeed & Fareh, 2006, p. 20, example 1)

Therefore, I argue that the Saudi participants are able to use the functions of the English discourse marker *so* in their speech even though they have not explicitly been taught the functions at school. I also argue that this does not undermine the importance of teaching the functions of this discourse marker. In fact, teaching the functions of *so* would help in raising the awareness of how to use *so* appropriately and enable language learners to employ it in their speech. It is worth bearing in mind that in my data, there were eight participants who did not use *so* whatsoever.

#### 4.6 Summary and conclusion

In this chapter, I have presented the analysis of *so* both as a discourse marker and as a non-discourse marker. The analysis of the SLC and LOCNEC corpora reveals that four non-discourse marker functions of *so* and five discourse marker functions are to be found in both corpora. These functions are grouped in the Table 4.7.

**Table 4.7 Summary of *so* in the SLC and LOCNEC**

Non-DM functions of <i>so</i>	DM functions of <i>so</i>
Adverb of degree or manner In a fixed expression A preform As a conjunction, is expressing a purpose	So indicating a result So marking main idea unit So introducing a new sequence So marking transition So introducing a turn

According to Brinton's (1996) classification of functions (cf. 2.3), *so* can be used at the textual level to indicate result. In addition, *so* is used to lead to the main idea of an argument. It is also used to introduce a turn and introduce a new sequence in the narrative, moving from one event or scene to the next. At the interactional level, *so* functions to mark a transition, indicating a potential transition between speakers' turns.

Regarding the frequency of *so*, the analysis shows some discrepancy in the use of some of the discourse marker functions of *so* between the NSs and Saudi speakers. If we take one example such as *so* marking transition, we can see that NSs used this function almost four times more than Saudis. This considerable difference also applies to *so* marking the main idea unit.

In terms of the non-linguistic variables (e.g. gender, time spent in an English-speaking country), the analysis of the SLC data shows no influence on the use of the discourse marker *so* at all. The results show no statistical significance between Saudi males and females and no statistical significance between those who have spent some time in an English speaking country or a foreign country like France and between those who have never been abroad.

## CHAPTER 5: *YOU KNOW* IN THE DISCOURSE OF THE SLC AND LOCNEC

### 5.1 Introduction

This chapter focuses on the use of the discourse marker *you know* in Saudi learner English. First, it describes the previous related literature of *you know* in section 5.2. Section 5.3 discusses the discourse marker functions of *you know* in the SLC and LOCNEC from a qualitative perspective. The quantitative results for *you know* are presented in section 5.4. Section 5.5 provides a comparison of findings in the corpus and textbooks. The final section, 5.6, comprises a summary of the chapter.

### 5.2 Functions of *you know* in the literature

According to Schiffrin (1987), *you know*, as its literal meaning suggests, functions within the information state of talk to mark transitions (cf. Watts, 1989). In particular, it serves to gain the hearer's attention as the speaker presents information (cf. Watts, 1989). By breaking down this discourse marker, it can be observed that it consists of a) a second person pronoun 'you' and b) a verb 'know', which means to have information about something. The main function of *you know*, as He and Lindsey (1998) point out, is to mark information which is new, important and salient for the benefit of the hearer. Similarly, Östman (1981) points out that the primary function of *you know* is to get the hearer to cooperate and/or to 'accept the propositional content of his utterance as mutual background knowledge' (1981, p. 17) Van Bogaert (2009) explains the main function of *you know* in the same vein, stating that:

'By using *you know* the speaker makes a request for cooperation and benevolence on the part of the hearer. The speaker wants the hearer to accept their choice of wording and appeals to the hearer and the hearer's metalinguistic awareness to accept their lexical choice.' (p. 144)

According to Halliday (1973; 1980), Leech (1983) and Östman (1981), *you know* functions at interactional as well as the textual level. At the interactional level, for example, when *you know* is accompanied by an interrogative intonation contour, it expects a response from the hearer (cf. Schourup, 1985). In contrast, when it appears with falling intonation, it carries a meaning that the speaker will say no more (Östman, 1981, p. 26). In addition, when *you know* appears in narrative, it serves to introduce background information or a digression. At the textual level, on the other hand, there are different functions of *you know*. In relation to pauses, *you know* may indicate lexical search or content search (Erman, 1987; Müller, 2005; Östman, 1981). Östman (1981) points out that the content search function of *you know* gives an indication of ‘hedginess and indirectness’ because the speaker’s knowledge of the content or topic is fuzzy (p. 30). Edmondson (1981), similar to Östman (1981), points out that one of the functions of *you know* is to allow the speaker time for linguistic planning or plugging.

In her study, Holmes (1986) found that the discourse marker *you know* fulfilled two broad functions: expressing speaker confidence or certainty and reflecting uncertainty of various kinds (p. 7). Under each main function, she found several sub-functions for *you know*. For instance, under the first main function, *you know* may be used to express ‘conjoint knowledge’ which she explains as literally to introduce what the speaker regards as ‘incontestable mutual knowledge’ (p. 8). In example (1), Holmes (1986) states that this *you know* indicates that what follows is relevant background information:

(1) [Woman to husband introducing a narrative at dinner party]

well: *you know* we went to Sally's that night

(Holmes, 1986, p. 8, example 3)



The second function of *you know* expressing certainty is the emphatic *you know*. This *you know*, as in example (2), is simply used to ‘stress the speaker's confidence and hence reassure the addressee concerning the validity of the proposition asserted’ (Holmes, 1986, p. 8). Holmes paraphrases *you know* in the following example as ‘let me assure you’:

- (2) [Young woman joking to neighbour in presence of flatmates]

I'm the boss around here *you know*

(Holmes, 1986, p. 8, example 5)

The last function here is the attributive use of *you know*. Holmes explains this function as ‘I'm confident you know the kind of thing I mean’ (1986, p. 9) as exemplified below:

- (3) [Young woman in conversation to flatmates over dinner]

they obviously thought he was a bit stupid *you know*

(Holmes, 1986, p. 8, example 10)

On the other hand, there two main functions of *you know* when expressing uncertainty. It serves as an ‘appeal to reassurance’ (Holmes, 1986, p. 10) from the interlocutor, for example when recounting an embarrassing experience, or sharing very personal information or feelings, as in the following example:

- (4) [Young man to friends at dinner party]

people can now see how narrow my eyes are set together *you know*

(Holmes, 1986, p. 8, example 13)

The second function of *you know* regarding uncertainty is to mark that the speaker is uncertain about some aspects of the linguistic expression of the proposition. Holmes (1986) points out that linguistic imprecision is manifested in three forms. The first form is when the

speaker is uncertain of the word or a phrase choice so uses *you know* to signal lexical imprecision, as in example (5):

(5) [Young man requesting clarification of previous speaker, his flatmate]

better / entertainment product or better / ***you know*** / music musicians<sup>24</sup>

(Holmes, 1986, p. 10, example 16)

The second form is when the speaker feels the needs to provide further information to clarify the point being made, so *you know* here introduces a qualifying structure as in excerpt (6):

(6) [Young man to friends]

we've got quite a big track / ***you know*** relatively speaking over here

(Holmes, 1986, p. 11, example 18)

The last form of *you know* can be identified by ‘the change in syntactic direction which follows...’ (Holmes, 1986, p. 11). This *you know* serves a false start as in example (7):

(7) [Young man in discussion with flatmates]

I mean look what Travolta as a as ***you know*** / he's not a pretty face or anything

(Holmes, 1986, p. 11, example 21)

Similar to Holmes (1986), Erman (2001) points out that one of the important functions of *you know* is concerned with the encoding of the text. She believes that the reason a speaker needs encoding and the ensuing editing is because of a preoccupation with selecting the right word or structure in relation to what he or she is trying to say (Erman, 2001). She states that ‘for this purpose the speaker may signal either repair of previous discourse or a new direction of it, or s/he may stall for time for the planning of the continuation of the utterance’ (2001, p.

---

<sup>24</sup> The sign / indicates a pause (Holmes, 1986, p. 5).

1340). To her, as in Holmes (1986, p. 11), repair is where the speaker ‘stops in mid-structure to make a restart’ (Erman, 2001, p. 1342), as in example (8):

- (8) The question is are you actually interested, attracted to her enough to want to, *you know*,  
what are you really interested in doing /.../

(Erman, 2001, p. 1345, example 13)

However, when comparing Holmes’s (1986) false start *you know* and Erman’s (2001) repair *you know*, it can be observed that the structure of utterances provided in both their examples are similar, i.e. the speaker starts a sentence, then breaks off with *you know* and carries on with a different syntactic structure. I agree with Müller (2005) that both Holmes (1986) and Erman (2001) are essentially talking about the same function, but using different terms. Müller (2005) comes up with her own classifications of repair and false start *you know* which I will later argue against.

With respect to its literal meaning, Schiffrin (1987) points out that the discourse marker *you know* expresses two possible composite meanings. The first suggested meaning is that information X is available to recipient(s) of talk. The second meaning is that information X is generally available. Therefore, these meanings suggest two discourse functions of the discourse marker *you know*. The first function is as a marker of meta-knowledge of what is shared by speaker and hearer. The second function is as a marker of meta-knowledge of what is generally known (Schiffrin, 1987, p. 268). In the first function, *you know* occurs when a speaker speaks about something it is assumed that the hearer shares. However, one of the basic problems of talk is when the speaker does not know whether the hearer shares knowledge of the topic or not. Holme (1986) and Stubbe and Holmes (1995) point out that

the discourse marker *you know* is considered to be an addressee-orientated device which may express the speaker's certainty or uncertainty in the addressee's sharing of relevant knowledge. Certainty is when the speaker is confident that the hearer shares the relevant knowledge or reassures the hearer that the proposition in question is valid. On the other hand, uncertainty is when the speaker is unsure either of the hearer's attitude or the linguistic precision of the message (Stubbe & Holmes, 1995, p. 69).

Schiffrin (1987) suggests four different possibilities for the distribution of speaker/hearer knowledge as Figure 5.1 below illustrates.

		Does speaker know of hearer's knowledge?	
		Yes	No
Does hearer know of X?	Yes	(a)	(b)
	No	(c)	(d)

**Figure 5.1 Meta-knowledge concerning speaker/hearer shared knowledge**

The first case is (a), when the hearer knows the background information and the speaker knows that the hearer does. In the second case (b), the hearer knows the background information but the speaker does not know that the hearer knows. Third is (c), where the hearer does not know the background information and the speaker knows this. Finally comes (d), where the hearer does not know the background information but the speaker does not know this.

To summarize Figure 5.1, it could be said that (a) and (b) represent a state of shared knowledge by the speaker and the hearer, even though the speaker in (b) does not know that the hearer shares that knowledge. In cases (a) and (c), the speaker is able accurately to assess

the hearer's knowledge, even though in (c) the hearer shows his/her lack of knowledge. In contrast, in the two situations (b) and (d) there is an inaccurate assessment on the part of the speaker. This requires, as suggested by Schiffrin (1987), the need for 'remedial work' (p. 268). As an example of remedial work, in (b) and (d), Schiffrin states:

'(... remedial work for b is likely to come from the hearer, e.g. *I knew that already*, and ... for is likely to come from the speaker, *I thought you knew that already*).'  
(1987, p. 269)

The ultimate function of *you know* is to make the conversation reach (a) in the meta-knowledge matrix, where the speaker knows about the knowledge which is shared with the hearer (Schiffrin, 1987). Let us consider the following example from Schiffrin (1987):

(9) Debby: a. And what d' you do?

Irene: b. I'm an N.T.A. In school.

Debby: c. What's that?

Irene: d. Uh:: really a cop. See that the- Yeh. See that=

Debby: e. Oh real: ly?! NO kidding?

Irene: f. =the kids behave a little bit. ***You know***. And-

Debby: g. I didn't know they had those: uh discipline

Irene: h. Yeh in the Philadelphia schools they do.

(p. 273, example 7)

However, most of the examples which Schiffrin (1987) provides to illustrate the cases in Figure 5.1 relate to *you know* as a non-discourse marker, i.e. *do you know* in questions.

The second function is *you know* as meta-knowledge of generally shared knowledge. Under this function, Schiffrin (1987) argues that *you know* can mark general consensual truths. The speaker here assumes the hearer shares this information through different channels, e.g. co-membership in the same culture, society, or group (p. 274). One of the examples of this function, as Schiffrin argues, related to ‘formulaic expressions of such truths’ (1987, p. 275) as in the following example:

(10) Henry: a. A mitzwah a day will keep the doctor away.

b. So if i can do this mitzwah today, maybe i don’t have t’ go  
t’ the dentist tomorrow?

Zelda: c. No y’s still have t’ go Henry, tomorrow!

hhhhhhhhhhh That’s a thought  
though, isn’t it hhhh

e. ***You know*** they say an *apple* a day keeps the  
doctor away?

(1987, p. 275, example 9)

In this example (10) Henry and Zelda have been discussing Henry’s visit to the dentist. Henry was not looking forward to it. So, Henry was playfully suggesting in (a) and (b) that since he was taking part in the interview, he should not have to see the dentist. He refers to the interview as a mitzah (a good deed) and metaphorically used instead of the apple. Schiffrin (1987) states that in (e) Henry ‘prefaces this expression not only with *you know*, but with *they say*, a quotative expression which itself conveys general consensus’ (p. 275). In addition, there is another example in which the discourse marker *you know* not only occurs with formulaic expressions, but also with general descriptions of situations, states, or events

(Schiffrin, 1987, p. 276). To demonstrate why a speaker would use this function of *you know*, the following example is given:

(11) Henry: I'm not a-...we're all not perfect, *you know*.

I'm not perfect Zelda, after all.

(1987, p. 276, example 11)

The scenario in which this utterance occurred was when Zelda, Henry and Schiffrin were talking about issues with husbands' dependence on their wives' daily presence in the home. So, Zelda states that Henry did not want her to continue to vacation at their summer home when he had to return to the city. Then Henry interrupts the discussion to defend himself. He begins with a consensual truth (nobody's perfect). As part of this discussion, Schiffrin comments on Henry's utterance by saying that Henry implicitly says that if no one is perfect, certainly Henry cannot be so either (Schiffrin, 1987, p. 276).

To conclude, it could be said that *you know* is a marker of meta-knowledge about generally shared knowledge. This function is fulfilled when speakers assume their hearers are likely to share consensual truths or facts because they share the same culture, society, or group.

In addition to the functions of *you know* in the information state, there is another interactional function of *you know* performing in the participation framework, namely *you know* in narratives. For the narrative *you know*, and as the name of the function suggests, it appears to occur when the speaker invites the hearer to share in the information being accomplished through narrative discourse. The discourse marker *you know* in narrative works on engaging the hearer by transforming his/her position in the discussion from just a recipient of information to an audience member.

Furthermore, Schiffrin points out that the function of *you know* in narratives is suggested by the two primary locations of this discourse marker in narratives. The first location is when *you know* occurs with internally evaluative events and the second one is with external evaluation (1987: 284). The internal evaluation which *you know* performs is achieved within the story frame and the external evaluation is outside the story frame. However, it is worth mentioning that there is one prime function of *you know* in these two locations “to draw the hearer’s attention to material which is important for reaching an understanding of why the story is being told” (1987: 282). Following is an example for illustration:

- (12) Henry: a. And she sat down there,  
                   b. and she says, ‘**You know** I got a problem Zelda’.  
                   c. she says, ‘I really got a problem’.  
                   d. So Zelda says, ‘What’s your problem? ’  
                   e. I was sittin’ right there.  
                   f. She says, ‘**You know** I gave you a tomato.  
                   Your tomato’s not as big as the one I gave you.  
                   What shall I do with it?  
                   g. what d’ y’ think I wanted t’ tell her t’ do with it?!

(1987, p. 282, example 18)

In example (12) the discourse marker *you know* occurs twice. Schiffrin (1987) believes that both occurrences work on a deeper level of internal evaluation for the reason that they preface two instances of reported speech comprising the central events of the story. Henry has been describing his neighbour’s stinginess and dissatisfaction after Zelda had given her a tomato in return for one previously loaned to her. *You know* prefacing the quotes in (b) and (f) shows that the neighbour presumes that Zelda should have been aware of the unfairness of



the unequal trade of tomatoes. In this example and by using two instances of reported speech which are prefaced with *you know*, Schiffrin argues that Henry in this example transforms his hearers into an audience enabling them to hear for themselves the way the neighbour presented her character.

However, the designation of this last function of *you know* in narratives has been criticized by Müller (2005). Müller (2005) states that when the discourse marker *you know* occurs at the beginning of reported speech, *you know* functions as a quotative *you know*. She states that ‘Schiffrin does not see its [the discourse marker *you know*] primary function in marking quoted material; rather, she explains what other functions *you know* has in the examples’ (2005, p. 154). Müller believes that because there are only two instances of *you know* prefacing reported speech occurring in Schiffrin’s data (on p. 282), this is the reason why she only mentions it without devoting any further space to it. However, to argue against Müller, there is in fact a third example of *you know* prefacing reported speech. To illustrate this quotative *you know*, Müller (2005) uses three examples from her own data as follows:

(13) 67 so Charlie Chaplin’s like *you know* &

68 & <Q hey waiter what’s the matter Q>,

69 and (H) the waiter says he was &

69a & ten cents short.

(6B, NS)

(2005, p. 169, example 151)

(14) 239 B: he didn’t have his money.

240 .. and then this beggar comes in &

240a & off the street,

241 A: .. yeah,  
 242 B: (H) and shining this nice coin up you know,  
 243 and polishing it all up nice and he's like,  
 244 ... ***you know*** .. I know that's my coin.  
 (47B, NS)

(2005, p. 170, example 127)

(15) 92 .. he's like <Q I'll pay for you Q>,  
 93 ***you know*** <Q don't worry about it Q>,  
 94 an' Charlie Chaplin is sayin' &  
 94a & <Q no no I have the money Q>,  
 (5B, NS)

(2005, p. 170, example 125)

The discourse marker *you know* in examples (13) and (14) occurs between the quotative *like* – which Müller also classifies as a quotative *like* in her treatment of *like* – and the reported speech. Her third example (15) shows *you know* embedded between stretches of reported speech. In fact, even though quotative *like* was not included in Müller's functions of the discourse marker *like* in terms of her own justifications (see Müller, 2005, p. 226), it can be noted that in all her three examples *like* was preceded by forms of verb 'be'. Thus, I would suggest, in line with Schiffrin (1987), that when *you know* prefaces reported speech, it clearly shows that the speaker is striving for the hearer's attention to what succeeds it.

However, even though Schiffrin's (1987) treatment of this discourse marker is one of the most detailed, it is still problematic. Müller (2005) agrees with Schiffrin (1987) that the literal

meaning of *you know* influences its discourse use but rightly challenges her when it comes to differentiating between discourse marker uses and non-discourse markers uses. In fact, when analysing Schiffrin's (1987) examples in her chapter on *you know*, it can be observed that more than half of her examples include *you know* as a sentence constituent which is not syntactically optional. Here are some examples from Schiffrin (1987):

- (16) Jack: a. And when you're a cripple, you're a prej-  
b. in other words...they're cripples because they're so religious is what-is  
the point I'm trying to make.  
c. In other words they're sick. Religiously.  
d. Like the:...***you know*** what Hasidic is?
- Debby: e. Umhmm.
- Jack: f. The Hasidic Jew is a cripple in my eyes, a mental cripple.

(1987, p. 269, example 1)

- (17) Jack: a. And I betcha you couldn't picture... America losing in Vietnam.  
b. In fact you said that.  
c. That we were somehow gonna win.  
d. **You know** why?  
e. Because you're American.  
f. and you were already built up in your mind,  
g. they built it up for you, whether you know it or not, that we never lose a  
war!

(1987, p. 280, example 16)

Müller (2005) points out that in these two examples, the auxiliary ‘do’ has been omitted. She continues that whether it is omitted or re-introduced, this will not leave the utterance ungrammatical or meaningless (2005, p. 153). Redeker (1991) criticizes Schiffrin’s examples for the reason that ‘(Do) *you know* questions are not discourse markers but rather a particular type of conversational move’ (p. 1166).

I agree with Redeker (1991), Stubbe and Holmes (1995) and Müller (2005) in their criticism as one of my main criteria for considering a linguistic item a discourse marker is that it is syntactically optional (cf. Chapter 2). Therefore, unavoidably, most of Schiffrin’s (1987) conclusions about the functions of *you know* are not relevant for the discourse marker *you know* in this thesis.

Not far from Schiffrin’s (1987) *you know* in narrative, commenting on one of her examples (example 54), Müller (2005) points out that this position of *you know* ‘carries an element of appeal to the hearer to imagine the scene which has been presented by the speaker as lively as possible through oral means’ (p. 171). For this *you know*, Müller (2005) dedicates a function called ‘imagining the scene’.

Müller (2005) dedicates a chapter to *you know* in her book *Discourse Markers in Native and Non-native English Discourse*. Müller’s (2005) *you know* functions on two levels: textual and interactional. On the textual level, *you know* has been found to function as a search for a lexical item or for some content a speaker is trying to use. It has been found to be used as a restart/false start and a repair. Another function on the textual level is *you know* to mark approximation of the linguistic expression. Müller (2005) states that ‘the linguistic expression is only an approximation of what would have been the most appropriate or correct

expression’ (p. 162). The last two functions on the textual level of Müller’s (2005) treatment of *you know* are *you know* to introduce an explanation of a previous idea or opinion (p. 166) and *you know* to introduce quoted speech (p. 168).

On the interactional level, *you know* also performs many functions. Because Müller’s (2005) data derived from the GLBCC<sup>25</sup> were describing a silent movie, *you know* was used to appeal to the hearer to ‘imagine the scene’ (p. 171) and to ‘see the implication’ (p. 175) of what a speaker was describing in the narrative. *You know* was also used to focus the hearer on information which it was believed was shared, i.e. ‘both participants knew the character or scene, either because they had both seen it, or because partner B had told partner A about it’ (2005, p. 178). The last two functions in Müller’s (2005) treatment were appealing for understanding which often concerns lexical choices and *you know* to acknowledge that the speaker is right, or as Müller puts it ‘Did you like it? Why?’ (p. 183).

The last two studies considered here investigated the use of the discourse marker *you know* in Arabic. Following Schiffrin’s (1987) discourse model, Gaddafi (1990) analysed the functions of the three Libyan markers *Garafit* (you know), *Garafit Keif* (you know how), and *taGrif* (you know). He argued that his results showed that there is a relationship between the discourse features of the three markers and their functional role in conversational interaction. His participants used all three markers to check the knowledge of the hearers and draw their attention to particular information in the discourse.

Kurdi (2008) explored three discourse markers used in Syrian Arabic: *ṣraḥṣraḥ* (did you know), *btaṣṣraḥ* (you know) and *ṣraḥṣraḥ kif* (you know how). She pointed out that all these markers were

---

<sup>25</sup> The Giessen-Long Beach Chaplin Corpus.

used in the Arabic interviews. The functions of these markers were ‘to check on the understanding of the hearer of what has been presented and to appeal to her/him to accept it. In this sense, it is used as a solidarity marker’ (2008, p. 125).

After reviewing the literature on *you know*, it seems clear that *you know* plays a significant role in verbal interaction on both the textual and interactional levels. In what follows, I describe the functions of *you know* as they occur in the SLC and native English speaker corpus, LOCNEC.

### **5.3 Functions of *you know* in the SLC and LOCNEC**

This section discusses examples of the discourse marker *you know* in the data of Saudi English language speakers and NSs of English. .

#### **5.3.1 Non-discourse marker functions of *you know***

As noted in section 5.1, there can be instances in which *you know* does not occur as a discourse marker, but rather as the propositional or non-discourse marker use. One type of these instances is when *you know* is part of a question as in extract (18). Another instance is when *you know* takes a complement as in extract (19) below. *You know* in these examples is not syntactically optional.

(18) <B> but I mean I used to get into conversations about about philosophies and about

<X> <\B>

<A> you like that don't you <laughs> <\A>

<B> I like that <X> deep things about politics and which you don't really get into in England in England you talk much more about if [ you talk about politics <\B>

<A> the weather <\A>

<B> yeah you talk about the weather but <X> politics <XX> talk political  
personalities do *you know* what I mean or </B>  
<A> (mhm) .. yes </A>

(LOCNEC-E36)

(19) <B> I donno maybe he is he found him his inner strength maybe </B>

<A> uhu </A>

<B> and and his transformation when he first got was in the island and then that  
moment he was found it was amazing because he lost a lot of weight and th= they  
really showed in the movie how his his appearance and his like hair and his beard  
and everything it not just only just like *you know* that it's just ok this is er him cast  
away in an island and that's set without finding that's a three years and nothing  
change in his appearance </B>

<A> uhu </A>

(SLC-AR10)

### 5.3.2 *You know* signalling lexical or content search

According to Holmes (1986), one of the functions of *you know* is to signal lexical or content search. In this function, she states that 'is generally preceded by a pause or a verbal hesitation such as *erm* or *er*' (1986, p. 11). To distinguish between lexical search and content search, Östman (1981) points out that for lexical search *you know* is often followed by a pause and for content search *you know* is potentially preceded by a pause. These can be seen in some extracts from the SLC and LOCNEC.

Extracts (20), (21) and (22) seem to confirm Holmes's (1986) and Östman's (1981) ideas:

(20) <A> so your mom is doing this role </A>

<B> yeah even my: money *you know* (er) to (er) like when I got my s= my salary

<foreign> mukafaty </foreign> I= have to you know like withdraw like three

hundred or two hundred fifty and give to my mom so that he give it to me x fifteen  
ten fifteen</B>

<A> yeah</A>

(SLC-AR09)

In example (20), we can see *you know* is followed by a filled pause (*er*) and the preposition *to* and then again a filled pause (*er*). Essentially, it seems that speaker AR09 is trying to find an equivalent in English to refer to the sum he gets paid every month by the university, his salary. In the same utterance, there is another piece of evidence which supports my categorization of this *you know* as signalling a lexical search. It can reasonably be assumed that the use of the Arabic word *mukafaty* is intended so that the speaker makes sure that the hearer understands the word he is trying to find. The following is another example similar to (20):

(21) <A> and so you .. it was difficult for you to say well it's it's him </A>

<B> (mm) </B>

<A> he's telling lies and </A>

<B> he was a very good story-teller </B>

<A> yeah </A>

<B> I mean when he told that story about . that he'd heard from the[i:] others about  
this Kaiser Soze and *you know* .. very .. sort of middle-eastern setting you know

</B>



<A> (mhm) <\A>

<B> obviously Turkish setting which was completely fictional <\B>

(LOCNEC-E15)

In extract (21), taken from native speaker data, the discussion between the interviewer and the interviewee is about a film called ‘*The Usual Suspects*’. They are speaking about a character called Keyser Söze. In this extract, the interviewee talks about Keyser Söze and then says *you know*, followed by a pause then the word *very* then another pause then *sort of*. Stubbe and Holmes (1995) point out that a form such as *sort of* may mark ‘propositional uncertainty’ (p. 83), i.e. uncertainty as to the truth of a statement. Therefore, it could be said that these contextual clues are evidence that this *you know* is another case of lexical search. Comparing (20) and (21) to the following two examples, however, it seems more like a content search. Let us consider an example from the SLC:

(22) <B> yes then he promised me to pay for the last month or something

like that and he didn't (er) really I was very young to to know how  
to deal with these people <\B>

<A> yeah<\A>

<B> and <\B>

<A> maybe too young to<\A>

<B> too young to to deal like this situation I've learned really how to  
deal with this people coz wa= my parents told me to be patient  
and just talk with him slowly and don't be angry don't do anything  
that most of (er) young people do <\B>

<A> (uhu) <\A>

<B> as (er) *you know* some mistakes how can I say it <\B>

<A> like bullying with those people or fighting them <\A>

(SLC-AR41)

In example (22), it appears that AR41 is narrating what his parents told him about dealing with his boss who refused to pay him. AR41 wants to give the interviewer an example of his own about what most young people would do in a similar situation. However, the occurrence of a filled pause (*er*) preceding *you know* suggests that the speaker is not struggling with wording but rather with the content of what he is going to say. I suggest this can be inferred from the use of *you know* + *some mistakes* without any pause to break the flow of the speech. The same explanation could be applied to the next example (23). It can be observed that *you know* is preceded with a filled pause (*erm*) and an unfilled pause (*..*), suggesting that the speaker is rather searching for content and not a lexical item.

(23) <A> so what else do you recommend since you: you've seen so many films and

<\A>

<B> <laughs> (*erm*) have you seen Seven <\B>

<A> no but it's not . on this term but <\A>

<B> no it will be on next term with any luck <\B>

<A> (*mhm*) <\A>

<B> very very good <\B>

<A> (*mhm*) <\A>

<B> very good (*erm*) . I sort of just I was spell bound from beginning to end . and

<\B>

<A> oh right <\A>

<B> you sit there thinking . it's unbelievable and the the techniques are very good  
it's very shot very dark . and (*erm*) .. *you know* the city is very very dark very ..

unfriendly and it's raining and there's storms and .. there's traffic and noise and commuters and everything's disjointed <\B>

(LOCNEC-E15)

Similar to *like* searching for the appropriate expression (see chapter 3.2.1), I consider the repetition of a word before or after *you know* as a lexical search.

### 5.3.3 *You know* for restart and repair

As opposed to Holmes (1986) and Erman (2001), Müller (2005) suggests that there are different uses of *you know* as a false start and as a repair. She believes that a ‘false start’ is when there is repetition without correction, that is, when there is a word followed by *you know* and the same word repeated, as in the following example from Müller’s data:

(24) 488 he .. he was sort of dressed like--

489 *you know*,

490 like you knew he didn’t have no .. &

490a & not too much money,

(5B, NS)

(2005, p. 161, example 106)

In terms of repair, Müller (2005) states that she does not include cases where the speaker ‘cuts off in mid-sentence and continues with a different syntactic structure after *you know* [...] to refer to cases in which a truncated or completed word is substituted by another word after *you know* within the same syntactic structure’ (p. 161). She gives the following example:

(25) 34 it was rocky.

35 .. rocky so they were kind of roll--

36 *you know* rocking back and forth.

(142A, NS)

(2005, p. 161, example 108)

She comments on this example that the speaker's first choice was *kind of rolling back* but for some reason she decides that *rocking* is more appropriate word for the meaning she wants to express.

In fact, I would agree with the notion of *you know* used for repair, but would argue against the false start. I would rather call it a 'restart'. Both the Saudi learners and the British native speakers appear to use it, as the following examples from SLC and LOCNEC show:

(24) <A> but whom is she going to: to check I mean you can't<\A>

<B> (huh) <\B>

<A> you can't tell people that this is me and <\A>

<B> n= sh= she cannot cheat other people because they will see the difference between to both both of them Because in.. in *you know* in picture she is more beautiful than her. herself <\B>

<A> yeah <\A>

(SLC-AR28)

(25) <A> you don't have that in Belgium well not that I know anyway so I mean in

Belgium for Christmas it's more like a quiet thing in family <\A>

<B> (mm) <\B>

<A> and here it's just party and noise [and <\A>

<B> publicity I hate it in France it was nice as well it's *you know* it's tradition and

Christmas decorations come out in December in England the Christmas decoration's

[sic] October </B>

<A> yes </A>

(LOCNEC-E34)

Extract (24) shows a case in which just one word, *in*, is repeated while (25) demonstrates the repetition of more than one word, *it's*. In both cases, the speaker cuts off in mid-sentence, inserts *you know* and starts again with the repetition. So, this *you know* signals a restart.

In the following set of examples, *you know* marks a repair. In these examples, we see that the word or words before *you know* is or are substituted by another word or words. Extract (26) shows a case of one word change and (27) shows a case of more than one word change:

(26) <B> yeah but I didn't really understand it but now I'm in that position </B>

<A> yeah yeah </A>

<B> e= even my brother is like </B>

<A> everything falls on you</A>

<B> yeah (eh) my brother eldest ***you know*** older than me with three years but  
<starts laughing> I'm the one helping him out <stops laughing> and you know  
taking care of him </B>

<A> yeah </A>

<B> yeah because I'm I'm the girl </B>

(SLC-AR01)

In extract (26), the speaker describes her relationship with her brother. She says that he is the eldest although she means older. Thus, she uses *you know* as a repair to follow it with *older* as it is the case with her brother.

In extract (27), nothing is repeated. The speaker says that the poor *are having* but has something different in his mind, *like they are living*. The discourse marker *you know* is used by the speaker to repair the incorrect utterance with what he wants to say:

- (27) <B> poor people will learn how to respect and teach people good things <\B>  
<A> because of the type of life they have <\A>  
<B> yes <\B>  
<A> which is (er) they have weak personalities do they not necessary <\A>  
<B> maybe, but I think because they are having *you know* like they are living in a rural life so they act to be like good people and respect <\B>  
<A> not necessarily some people don't live in rural life they live in urban lives like city life <\A>  
<B> (uhu) <\B>  
<A> but they're still poor <\A>

(SLC-AR37)

#### 5.3.4 *You know* introducing an explanation

Holmes (1986) points out that the function of *you know* when preceding an optional clause constituent is to 'introduce[s] more specific or clarifying information' (p. 11). Similar to Holmes (1986), Crystal (1988) argues that there are particular situations in which the discourse marker *you know* can be used. One of these is in the middle of a sentence, often to 'clarify or amplify the meaning of what one has just said. It warns the listener that the next words are particularly important. *He's just got a new BMX – you know, one of those tough little bikes...*' (Crystal, 1988, p. 47). The following are some examples from my data concerning *you know* introducing an explanation:

- (28) <A> so you you want to: to work in the cinema industry or is it just <\A>  
<B> I don't know <\B>

<A> you don't know .. do you want to: to write articles </A>

<B> <X> yeah </B>

<A> reviews of films or </A>

<B> I think if I did work in the cinema industry that I'd be <XX> sort of production side .. or the sort of . the critical side of it *you know* looking at films and writing about them </B>

<A> (mhm) </A>

<B> (erm) .. I don't really know <laughs> </B>

<A> well you still have some to think about it </A>

<B> see what happens .. yeah </B>

(LOCNEC-E20)

Speaker E20 tells the interviewer about her interest in media but she has nothing in particular in mind in terms of what job she might choose. When the interviewer asks about the option of writing reviews for films, she accepts that as one of the possible options she might pursue. She specifies her interest, i.e. the critical production side. She tries to clarify her answer, prefacing her explanation with *you know looking at films and writing about them*. It can reasonably be argued that *you know* here functions as an explanation because E20 pauses for a while after saying *the critical side of it* and says *or the sort of* and pauses again. These clues show that, at some point, she wants to make herself clear and provide a further explanation. A similar function is illustrated here in another example from the Saudi corpus:

(29) <A> (um) how many times did you visit UAE</A>

<B> twice </B>

<A> twice </A>

<B> yup </B>

<A> ok (erm) is there any other country that you've visited that you've liked </A>

<B> UAE Bahrain I went to: (erm) to Turkey but I was a little girl that back then I wasn't interesting in such organizing things (laughs) so maybe I didn't noticed (erm) we're not a big favour of (laughs) travellers because of my father's job ***you know*** bec= he usually has a lot of job in the summer so we're not free to travel and were not able to travel without dad </B>

<A> (uhu) </A>

(SLC-AR16)

In this extract, speaker AR16 recounts the places she had visited which are few. The reason, as she says, is because of her father's job. However, she realizes that simply stating her father's job as the reason is somewhat vague. Therefore, she explains this with *you know bec= he usually has a lot of job in the summer*.

### 5.3.5 *You know* for speaker/hearer shared knowledge

There were a number of instances in my data where *you know* appeared to function as a marker of shared knowledge between the speaker and the hearer. Extract (30) illustrates this:

(30) <B> (er) a movie (er) it has to be a movie or </B>

<A> it could be a show that you like </A>

<B> yeah I loved (er) sitcom called friends </B>

<A> ok yeah </A>

<B> for (er) I lo= he was it was the first show that I watched that really was English and I loved it and I I was able to understand first I was I= just lgh= laughing because they were laughing I couldn't understand anything </B>

<A> yeah </A>



<B> but but then I realized that now I I understand what they are saying and (er) I really loved the show because not only about friendship and about that about life in general </B>

<A> (uhu) </A>

<B> they each episode they had a topic and they were discussing it and how do they (er) how do they deal with it with their friends </B>

<A> right </A>

<B> and I loved their relationship although it is complicated and it is really messed up but they </B>

<A> (uhu) </A>

<B> they they s= they they were good then </B>

<A> (uhu) </A>

<B> and I loved that er in he w= it was ten seasons and I was really sad when it finished but I think that they ended in a good time because they know if they cont= if continue to produce more seasons it would it would be a little bit boring </B>

<A> right </A>

<B> because there is not topics to talk about and everything </B>

<A> (uhu) </A>

<B> they covered it really well I loved the cast I loved the the the plot the story their jokes everything </B>

<A> right I agree I agree it was ten years </A>

<B> yeah </B>

<A> yeah I think that's long enough </A>

<B> yeah </B>

<A> that's true how old were you when you saw it </A>

<B> (er) mid of school </B>  
 <A> really </A>  
 <B> yeah </B>  
 <A> (uhu) </A>  
 <B> I watched it </B>  
 <A> you saw it on TV or </A>  
 <B> yeah in TV because my only source TV </B>  
 <A> (uhu) </A>  
 <B> I could watch </B>  
 <A> all ten seasons from TV </A>  
 <B> yeah </B>  
 <A> ok so you're you're </A>  
 <B> yeah </B>  
 <A> on schedule </A>  
 <B> yeah </B>  
 <A> all the time </A>  
 <B> yeah sometimes they would they don't *you know* air new episode </B>  
 <A> yeah </A>

(SLC-AR11)

In this example, the interviewer speaks about a sitcom she loves. She chooses to ensure that the interviewer understands this detail by employing the discourse marker *you know*. When speaker AR11 says that they do not air new episodes, the interviewer responds with *yeah*. Similarly, there is the following example from LOCNEC:

- (31) <A> it's quite a big city isn't it </A>  
 <B> it is quite big yeah that's why I came here cos I wanted to come to somewhere

smaller <\B>

<A> oh yes <\A>

<B> yeah <\B>

<A> (mm) .. well it's not very far away from here is it <\A>

<B> no it's only a couple of hours away <\B>

<A> oh yes but you're staying on the campus <\A>

<B> no I'm living in town now <\B>

<A> oh yes but you stayed on campus last year <\A>

<B> yeah I did <\B>

<A> okay so what do you prefer staying in town staying on campus <\A>

<B> (erm) .. well I enjoyed it being on campus last year <\B>

<A> (mhm) <\A>

<B> because I ended up in a really good block where we got on with everybody  
<\B>

<A> (mhm) <\A>

<B> that was good <\B>

<A> that's very important <\A>

<B> yeah (erm) and it was really convenient for lectures *you know* you could get  
up five minutes before and still not be late <\B>

<A> (mhm) <\A>

<B> so . I'll probably try and get back on campus next year <\B>

(LOCNEC-E17)

Extract (31) begins with the interviewer's question about the interviewee's home town, Sheffield. The interviewee answers the question and explains why she chose Lancaster as the

place to go to university, namely that it was smaller than Sheffield. The discussion is about life in the campus and then in the town, Lancaster. Because both the interviewer and the interviewee live in Lancaster, this means that they share the same knowledge. The interviewee explains the advantages of living in Lancaster. She uses *you know* because both of them share the same knowledge: Lancaster and attending lectures. Extract (32) likewise shows *you know* used as speaker/hearer shared knowledge:

- (32) <A> (erm) but are you living on camp= in in Lancaster <\A>  
<B> no I live in Carnforth which is north of Lancaster <\B>  
<A> have you been living here well there <\A>  
<B> I've always lived here I was born in Lancaster <\B>  
<A> oh yes so you you it's not new for you <\A>  
<B> no no well one of the factors was obviously when you're married and .. my  
husband's working and we've got a house and <\B>  
<A> yeah <\A>  
<B> location was really as much as anything the deciding factor <\B>  
<A> (mhm) <\A>  
<B> although Lancaster did offer the course which I wanted <\B>  
<A> (mhm) <\A>  
<B> (erm) but I wouldn't have been prepared to travel very far *you know* I  
wouldn't have <X> much further than Manchester say <\B>  
<A> no I understand .. yes .. (uhu) <\A>

(LOCNEC-E50)

Speaker E50 speaks about her life as a wife, a mother to a little girl and a student. She tells her story about her previous job as a training instructor in British Telecom. She then decided

to leave her job and study linguistics at Lancaster University. This is where this extract comes in. In the flow of speech, the interviewer asks the interviewee if she is living in Lancaster. She says that she lives in Carnforth which is quite close to Lancaster, approximately seven miles away. She wants to comment on the importance of the location in her case. She says that the furthest point she could go is Manchester. She uses *you know* in her utterance to elicit the hearer's display of information reception. The interviewer provides feedback, '*no I understand .. yes .. (uhu)*', which shows the interviewee that he shares knowledge with her , and empathizes with her point of view.

### 5.3.6 *You know* for generally shared knowledge

This *you know* is similar to the previous point in terms of sharing knowledge but is different in terms of the scope of the shared knowledge. The previous *you know* deals with shared knowledge between speaker and hearer, i.e. as colleagues in the same field, whereas this *you know* is about what is generally shared through co-membership in the same culture or society.

Let us have a look at the following example:

- (33) <A> (mm) are you staying on campus <\A>  
<B> no not this year no I was on campus last year <\B>  
<A> mhm <\A>  
<B> but erm I'm living in town <\B>  
<A> .. Lancaster <\A>  
<B> yeah Lancaster yeah <\B>  
<A> what do you prefer .. staying on campus or in town since you've <\A>  
<B> erm <\B>  
<A> experienced both <\A>  
<B> yeah <\B>

<A> so <\A>

<B> erm .. it's diff= difficult to say really I mean campus is brilliant

cos you're right near everything and .. but I really enjoy living

in town cos I'm living *you know* a great bunch of people <\B>

<A> mhm <\A>

<B> and I got on really well with them .. and (er) <X> a nice house

<X> it's always a bit of a tip but . but you know I'm I'm really enjoying it <\B>

<A> (mhm) <\A>

<B> but *you know* there's hassles with getting on buses and <\B>

<A> yeah <\A>

<B> I've gotta walk across that bridge over the river .. I don't know if you know by

Sainsbury's <\B>

<A> (mhm) <\A>

<B> and it's so windy and cold <\B>

<A> oh I I can imagine <\A>

(LOCNEC-E14)

In example (33), speaker E14 is answering the interviewer's question about where he is living while studying in Lancaster. Speaker E14 says that he was living on campus last year but not anymore, as he is now living in town. Then the interviewer asks him which experience he prefers more. E14 starts with his first experience of living on campus and praises it by saying 'brilliant'. However, when he moves on to describe his second experience of living in town, E14 switches from a description of his own situation to a general statement about how trying it could be to get on buses when living in town. He uses the discourse marker *you know* to seek the hearer's endorsement of the generalization '*you know* there's hassles with getting on

buses’. As can be seen in the following turn, the interviewer agrees with what E14 says ‘yeah’ as they both live in the same society and share the same culture. Another example from the SLC shows *you know* employed for generally shared knowledge:

- (34) <B> two thousand yeah in two thousand I visited Syria and I <foreign> yaghni  
</foreign> and I’m still visiting it <laughs> <\B>  
<A> every year <\A>  
<B> yeah every year with my family of course <foreign> yaghni </foreign>  
<\B>  
<A> <foreign> mashalaah </foreign> <\A>  
<B> (er) (er) really I don’t know what is it ok but when I when I talk about my  
feelings about Syria you know sometimes when you <foreign> yaghni  
</foreign> sometimes when you feel that I feel that this is just like you know  
second home for me yeah <\B>  
<A> it’s good <\A>  
<B> sometimes yeah and if you think about it may be or of course that’s sure  
<foreign> yaghni </foreign> in Islam we know that <foreign> yaghni  
</foreign> <foreign> (al sham) </foreign> in in general <foreign> yaghni  
</foreign> <\B>  
<A> yeah <\A>  
<B> *you know* Palestine Syria <\B>  
<A> (eh) <\A>  
<B> Jordan Lebanon is I don’t know how to say it in English <laughing><\B>  
<A> is it you mean a blessed place <\A>  
<B> yeah a blessed place that’s it <\B>

(SLC-AR41)

This is another example of *you know* for generally shared knowledge. Speaker AR41 chooses to speak about a country he has visited and liked. He says that he visits this country every year. He explains how he feels towards this country as it is like a second home to him. It is a special place not only to him, but also Islam treats it as such. His use of *you know* prefaces generally shared knowledge, i.e. that the word ‘al sham’ includes ‘Palestine Syria Jordan Lebanon’.

The final example in this category is excerpt (35). E36 speaks about her experience of teaching English in France. She expresses her joy concerning what she believes to be an achievement. The discussion then changes from teaching experience to her experience of living in France. She says that France was not as she had imagined. She does not like it so much. However, she wants to generalize her opinion. Thus, she prefaces her justification with general shared knowledge *you know*. In my opinion, the use of this *you know* helps speaker E36 to impersonalize her tone of judgment, by relating it to a broad class of people, rather than anyone in particular:

(35) <A> (mhm) and do you enjoy teaching <\A>

<B> oh . I did I used to get up on Thursday morning cos Thursdays <XXX> I can't cope with this it's just too much (er) I just <X> it got worse and worse but at the[i:] end of it when it looking back on it I really feel like I achieved a lot <\B>

<A> (mhm) <\A>

<B> (erm) because these kids had had this dreadful teacher they all hated English <XX> speak English and I got them speaking and I felt like I really achieved something and <\B>

<A> it was a very . (er) enriching experience for you <\A>

<B> yes (mm) <\B>



<A> but (er) otherwise erm .. did you enjoy living in France <\A>

<B> yeah <\B>

<A> I mean it's a different mentality <\A>

<B> yes it is (er) <\B>

<A> how did you experience that <\A>

<B> well I'd always had this image of France being this wonderful place because you know the[i:] only times I had ever been independent when I was young was when I'd gone on French exchange <XX> really happy time for me . and then I arr= arrived and it's not quite so wonderful after all *you know* I think it's like anywhere you get nice people and you get less nice people but <\B>

<A> (mhm) <\A>

(LOCNEC-E36)

### 5.3.6 *You know* in narratives

The narrative function occurs when a hearer is invited to share in information transfer. It only happens through narrative discourse (Schiffrin, 1987). As I discussed in the introduction to this chapter, this type of *you know* occurs within reported speech to transfer the hearer from just being an information recipient into a participant in the storytelling. Müller (2005) calls this function of *you know* 'imagine the scene' (p. 171). However, in this study I classify *you know* under this function as 'narratives *you know*'. Let us consider the following example from the LOCNEC data:

(36) <A> <laughs> <X> so you also went to France on an exchange <\A>

<B> yeah I went to France <\B>

<A> where was it was it <\A>

<B> (erm) .. that was in the north (erm) it's a village called <name of the village>

and I can't think where it was near now but it it was sort of like very north <\B>

<A> was it not far from Lille or <\A>

<B> (mm) I can't <\B>

<A> Turcoing or <\A>

<B> I'm not sure .. I'm not sure .. we went somewhere <starts laughing> but I can't think where I've a terrible memory <stops laughing> <\B>

<A> but you you didn't really like that . did you <\A>

<B> (erm) .. that that was very much more sort of like going out and seeing France seeing bits of France I mean (erm) we saw the: Bayeux tapestry .. and (erm) <\B>

<A> (mhm) <\A>

<B> we saw (er) Le Mont Saint Michel <\B>

<A> (mhm) <\A>

<B> and all<?> things like that . so and so that this is when came home with photos of the place <\B>

<A> (mhm) and your dad was happy <laughs> <\A>

<B> but .. <laughs> yes .. but I mean <X> we didn't I mean . we found out later that our French teacher had actually said *you know* . don't let them go out at night with their friends don't let them get together and speak English together *you know* they should be with the family speaking French and you know me being quite shy I was sort of like I didn't know what to say to them so I spent a very quiet week there <\B>

<A> so it was just sightseeing and nothing else <\A>

<B> yeah really I mean it it was good in a way but it wasn't as much fun you know <X> it was more of a learning experience <laughs> <\B>

(LOCNEC-E18)

E18 tells the interviewer one of her experiences when she went to France on an exchange. The discourse marker *you know* prefaces two instances of reported speech in which E18 introduces and describes what her French teacher has asked them to do and not to do. Thus, rather than just telling her audience – the interviewer – about the teacher’s instructions, E18 transforms her hearer into an audience to the telling. This *you know* helps to allow the hearer to hear for himself the teacher’s orders. Schiffrin (1987, p. 282) points out that when *you know* occurs within reported speech, it draws the interviewer’s attention to material which she believes important for reaching an understanding of why this part of the story is being told.

The same function of *you know* occurs in the following example. *You know* in (37) is employed with an evaluation which is deeply embedded in the narrative structure. AR1 seems unhappy with what her younger sister is saying about her. Her sister is not convinced that she has a lot to do, e.g. looking after the house as well as studying. AR1 uses *you know* in this narration to draw the interviewer’s attention to material which she believes important for reaching an understanding of why this part of the story is being told.

(37) <B> (er) but thank god my sisters were English department too </B>

<A> ok </A>

<B> but at college so it didn’t really help </B>

<A> (uhu) </A>

<B> alot (erm) </B>

<A> did it help you in studying </A>

<B> yeah of course you know writing <starts laughing> they have to correct everything <stops laughing> but my younger sister (erm) in the high school she completely don’t understand anything she was like <first name of interviewee> you are *you know* making a story out of it just cool down it’s like nothing she

understand a= she doesn't understand anything specially this semester it was like  
the the most difficult one I have semantic phon= (er) phonology novel and drama  
like ru= time was running from the first moment <X> the semester so she was like  
<first name of interviewee> <X> <foreign> dafora </foreign> <laugh> </B>  
<A> (uhu) yeah </A>

(SLC-AR1)

In the final example, speaker 32 speaks about the film 50/50 that she chose to discuss with interviewer. She used the discourse marker *you know* to check on the interviewer's comprehension. Müller (2005) interprets the function of *you know* in an example similar to this in her study as 'can you imagine the scene?' (p. 172), which seems highly appropriate.

(38) <B> (erm) (er) I think I will choose the last topic </B>

<A> a[ei] movie </A>

<B> yeah (erm) (er) about a year ago I saw (er) a movie (er) (erm) called fifty fifty  
(er) it's really great great movie (er) (er) it is about (er) man who was who who  
lived his live normally and all of a sudden everything changed (er) he discovered  
that he has (er) cancer </B>

<A> oh that's bad </A>

<B> yeah it is and (er) (er) everything seemed very possible and *you know* he gave  
up all he gave up his beliefs and and he </B>

<A> oh </A>

<B> he started to to lose to lose hope and hope is a very dangerous thing to lose as  
they would say </B>

<A> yeah </A>

(SLC-AR32)

### 5.3.7 Unclear instances

The final category of *you know* consists of instances which are hard to classify. Let us consider the following example:

(39) <A> more fun <laughs> <\A>

<B> well at one point they said this to her and she ran out crying and I was left with all thirty of them in the classroom <\B>

<A> thirty <\A>

<B> well I had cos we were both with the with the: group at the time <\B>

<A> (mhm) <\A>

<B> <X> *you know* <XX> rather one each half and they were saying you know oh why can't we have Karin she's a lot more fun you know and she ran out in tears and I was like oh jeeze you know so I had to deal with thirty overexcited fifteen year olds .. (erm). and and and <X> I couldn't go <XX> head because she was already so . fragile <\B>

(LOCNEC-E36)

In extract (39), for example, the function of *you know* is hard to classify because it is preceded by one intelligible word and followed by two intelligible words, as <XX> illustrates. This instance of *you know* is hard to recognize because there is an insufficient amount of context. In the last example (40), *you know* is followed by an empty pause for two seconds. Therefore, I classified it under the category 'unclear'.

(40) <B> for me (er) my job was a photographer, if I did not do that I would be a good bad photographer they would actually spread some kind of you know things bad things about me they wouldn't say that I'm a good photographer or something like that <\B>

<A> your being honest, your being honest to yourself and people <\A>

<B> yeah <\B>

<A> but this painter I don't think his honest you know <\A>

<B> I don't know If someone told me that they wanted to look like a way that is not him I will tell him that there's nothing in common you and this picture <\B>

<A> but if you want it <\A>

<B> but if I want it I will do it I'm a photographer not *you know* .. <\B>

<A> how about her friends <\A>

<B> I don't know her friends are actually look like they are actually not happy with that they seems kind of disappointed in her <\B>

(SLC-AR44)

### 5.3.8 Summary of the discourse marker functions of *you know*

In this chapter, we have seen that *you know* functions at both textual and interactional levels. The first function is *you know* for speaker/hearer shared knowledge. The function of this *you know* is to create a situation in which a speaker knows that a hearer shares knowledge about what they are discussing. In addition, *you know* here helps to represent the speaker as an information provider who depends upon hearer reception of information. The second *you know* is as a marker of meta-knowledge about generally shared knowledge. Here *you know* marks consensual truths which speakers assume that hearers are likely to share as they belong to the same culture, society, or group of people. The last function is *you know* in narrative. It occurs with both internal and external evaluation. Both locations assist to draw the hearer's attention to materials which are central for reaching common ground and understanding why the story is being told.

## 5.4 Quantitative results for *you know*

In what follows, I present the quantitative results which illustrate the frequency of each of function of *you know*. In addition, I compare the frequency of *you know* functions in the SLC to that in LOCNEC. Table 5.1 below shows an overview of the main functions of *you know* and their corresponding abbreviations. In the table, the first three functions denote *you know* at the textual level whereas the next three functions operate at the interactional level.

**Table 5.1 Functions of *you know* and their abbreviations**

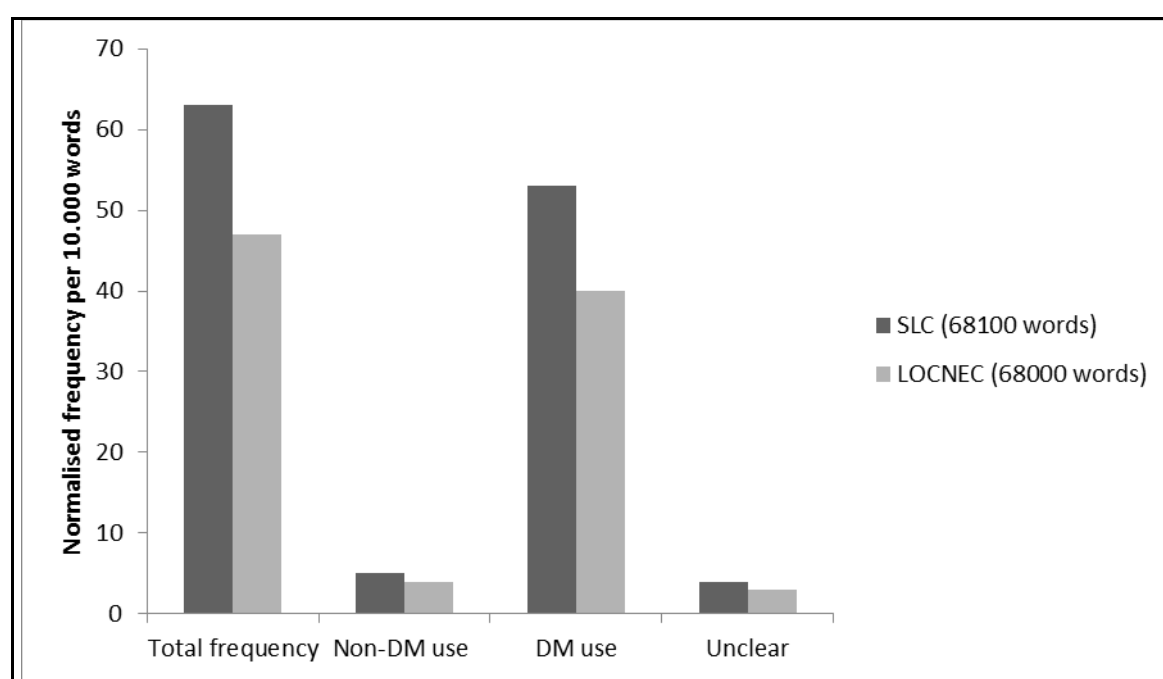
<b>FUNCTION</b>	<b>ABBREVIATION</b>
<i>You know</i> for lexical or content search	Search
<i>You know</i> for restart and repair	Restart and repair
<i>You know</i> introducing explanation	Explain
<i>You know</i> for speaker/hearer shared knowledge	Shared knowledge
<i>You know</i> for generally shared knowledge	Gen. knowledge
<i>You know</i> in narratives	Narrative

I begin this section by presenting the overall occurrences of *you know* in the SLC and in LOCNEC. As Table 5.2 and Figure 5.2 below illustrate, the raw occurrence of *you know* amounted to 428 in the SLC versus 318 tokens including non-discourse marker use, discourse markers and tokens which were hard to classify. It can be noted that there was a high frequency of the use of *you know* as a discourse marker in the SLC, with 358 tokens, or 53 per 10,000 words. These tokens covered different functions which I explain below. In LOCNEC, the frequency of *you know* as a discourse marker was 273 tokens, or 40 per 10,000 words. Statistical significance was achieved between the SLC and LOCNEC in the discourse

marker use of *you know* ( $p < 0.001$ ). For non-discourse marker use and unclear tokens, there were some slight differences but none achieved statistical significance.

**Table 5.2 Numbers of tokens of *you know* in the SLC and LOCNEC (frequencies per 10,000 words in parentheses)**

	Total frequency	Non-DM use	DM use	Unclear
SLC (68,100 words)	428 (63)	36 (5)	358 (53)	30 (4)
LOCNEC (68,000 words)	318 (47)	25 (4)	273 (40)	20 (3)



**Figure 5.2 Distribution of the linguistic item *you know* – SLC versus LOCNEC**

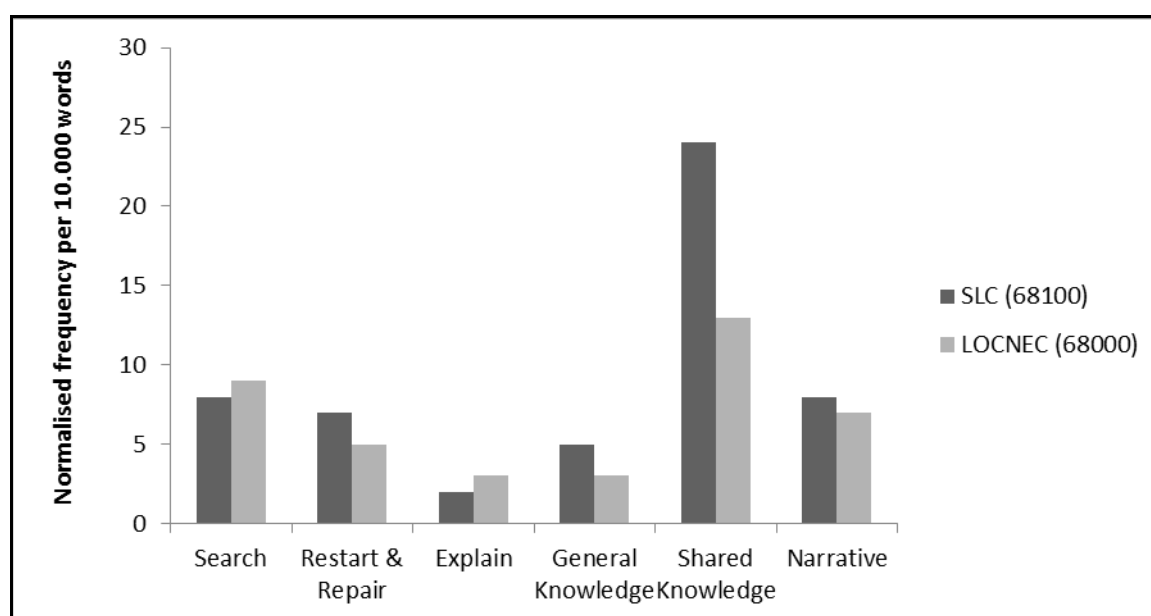
Regarding the occurrence of the discourse marker functions of *you know*, as can be seen in Table 5.3 and Figure 5.3, all functions were used by Saudi speakers and native speakers. Most of the functions were used by Saudi speakers more than native speakers except *you know* for lexical or content search. The most frequent function in both corpora was *you know*



for speaker/hearer shared knowledge. The difference between the two corpora for this function achieved statistical significance  $p < 0.0001$ .

**Table 5.3 Distribution of *you know* in the SLC and LOCNEC (frequencies per 10,000 words in parentheses)**

	Search	Restart & Repair	Explain	General Knowledge	Shared Knowledge	Narrative
SLC (68,100 words)	52 (8)	49 (7)	11 (2)	33 (5)	161 (24)	52 (8)
LOCNEC (68,000 words)	63 (9)	32 (5)	20 (3)	20 (3)	91 (13)	47 (7)



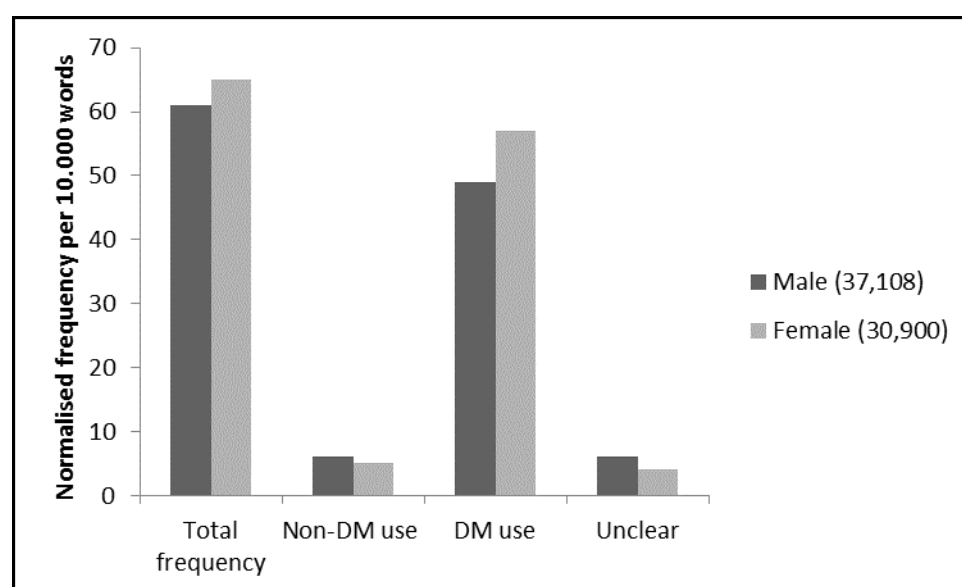
**Figure 5.3 Distribution of *you know* as discourse marker – SLC versus LOCNEC**

Regarding gender in the SLC, after analysing the overall occurrence of *you know* it was found that the female speakers used a higher rate ‘raw frequency’ of *you know* with 200 tokens or 65 per 10,000 words compared to the male speakers who used it 228 times, or 61 per 10,000 words. Indeed, the female speakers used *you know* as a discourse marker more frequently than the male speakers. The frequency of the non-discourse marker and the unclear instances

was higher for the male speakers than the females. However, neither frequency difference was statistically significant.

**Table 5.4 Number of tokens of *you know* in the SLC according to gender (frequencies per 10,000 words in parentheses)**

	Total Frequency	Non-DM use	DM use	Unclear
Male (37,108 words)	228 (61)	23 (6)	183 (49)	21 (6)
Female (30,900 words)	200 (65)	14 (5)	175 (57)	11 (4)



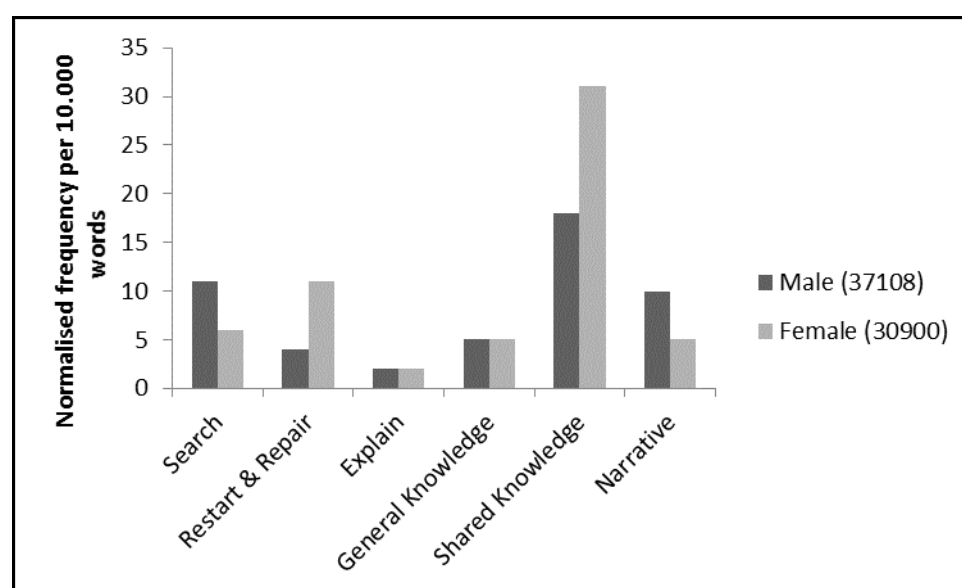
**Figure 5.4 Distribution of the linguistic *you know* in the SLC according to gender**

The analysis of the discourse marker use of *you know* shows that the gender has some effect. Even though the number of tokens is quite small, it is worth discussing. Table 5.4 and Figure 5.4 illustrate in detail the distributional pattern between male and female speakers. For both gender groups, *you know* is used most frequently as a speaker/hearer shared knowledge marker. The female speakers used 95 tokens, or 31 per 10,000 words in their speech, whereas the males used 66 tokens, or 18 per 10,000 words. This difference is highly significant ( $p < 0.0001$ ). Table 5.5 below shows that the female speakers used *you know* for speaker/hearer shared knowledge more than the total of all other functions (80 versus 95). My results are

similar to those of Macaulay (2002) who found that women are more likely to use *you know* than men. In another study by Stubbe and Holmes (1995), *you know* was found to be more frequent in the speech of the young male than in the speech of the young female, both from a working class background. However, because LINDSEI does not provide information about the social background of the participants, the results of this study are not compared to those of Stubbe and Holmes (1995).

**Table 5.5 Functions of *you know* in the SLC according to gender (frequencies per 10,000 words in parentheses)**

	Search	Restart & Repair	Explain	Gen. Knowledge	Shared Knowledge	Narrative
Male (37,108 words)	41 (11)	15 (4)	6 (2)	17 (5)	66 (18)	38 (10)
Female (30,900 words)	11 (6)	34 (11)	5 (2)	16 (5)	95 (31)	14 (5)



**Figure 5.5 Distribution of DM functions of *you know* in SLC according to gender**

In fact, there are considerable individual differences in the SLC in terms of the rate of occurrence of the discourse marker *you know*. There were some speakers in this corpus who used *you know* at a markedly higher rate (one speaker 30 tokens and another speaker 34

tokens) in an average of 1,400 words per recording, which is a clear contrast to the other speakers' use which ranged from two to 20 tokens (see Appendix 12). On the other hand, the analysis of the native speakers data showed that the highest frequency of *you know* use was 35 tokens in one interview, interviewee E11. The average total number of words in that interview (B turns) was 2,800 words.

However, in this situation, I argue that the high use of *you know* which occurred in the SLC is not necessarily due to the lack of fluency of speakers but rather occurred as 'speech habit' (Östman, 1981, p. 27) or what Macaulay calls 'part of the speaker's discourse style' (2002, p. 765). Macaulay (2002) points out that the discourse marker *you know* does not appear to mark 'assumptions of shared knowledge, but rather to form part of the speaker's discourse style and the rhythmic organization of utterances' (2002, p. 765). Even if Macaulay's argument is valid, it is still necessary to classify tokens of *you know* which conform to shared knowledge as a discourse marker use and this is what I did with my data.

### **5.5 Comparison of findings in the corpus and textbooks**

I examined the textbooks to search for the discourse marker *you know*. However, similar to the discourse marker *like*, no tokens of *you know* as a discourse marker were found. Müller (2005) searched for the discourse marker *you know* in German textbooks. Her participants studied three textbook series at school, Green Line, English G and Notting Hill Gate. Müller (2005) found that *you know* occurred only once in each book.

In the case of my participants, I suggest that they acquired *you know* through exposure to the media and their interaction in English with their peers. I also suggest that there is an influence of the L1 on the use of *you know*. As discussed in the literature in this chapter, studies such as

those of Gaddafi (1990) and Kurdi (2008) have demonstrated that there is an equivalent marker in Arabic to the English *you know*. The review of the textbooks revealed no occurrence of *you know* as a discourse marker. Similar to *so*, the ability of Saudi speakers to use the discourse marker *you know* has to do with the similar marker they have in Arabic. I provide a few examples from Gaddafi (1990) and Kurdi (2008) for illustration. I start with an example from Gaddafi (1990), who points out that one of the functions of *you know* he found in his data was *Garaft* (you know) to check the hearer's knowledge as in the following example:

(40) **Khalid**

1.fih v usandu:qhiki **Garaft** ... fih ZI atmurabaG hiki Garaft...

there+is like the+box like know+you there I likke the+square Like knew you

there is like a box like this **y'know** ... There is Like a square Like this y'know

2. nji nagaGmiz Galeih ana wa Gabdullah

come sit on+it I and Abdullab

we come and alt on it Abdullah and I

3. wa nadarbik Galeih ha ha ha

and beat on+it ha ha ha

and beat it ha ha ha

(Gaddafi, 1990, p. 104f, example 1)

Another example is provided by Kurdi (2008). She points out that the marker *btaʕfri* (you know) is used to express certainty on the part of the speaker that what she or he is talking about is knowledge shared with the hearer. Kurdi (2008) states that 'The learners use *btaʕfri* to signal certainty about what they are talking about ... because they think the interviewer

shares views with them by virtue of being Syrian' (2005: 120). The following is one example from her data in relation to this function:<sup>26</sup>

(41) H1 when did you start learning English first first

when you start?

Gs2 it is we start normal in the seventh grade

H3 hmm

Gs 4 but *btaʕfri* I mean it is I mean- yeah

5 I mean we can say that we started learning from the seventh grade

H6 hmm

Gs7 but honestly seriously seriously we didn't start until I can say until

the sixth year after the sixth year seven

(2008, p. 121, example 32)

These two examples are used to show that the marker *you know* is similar to one in Arabic. Therefore, this may indicate an interference of L1 in L2. However, I still emphasize the importance of teaching the discourse marker *you know* at schools. This will help make Saudi learners more aware of the different functions of *you know* in English discourse.

## 5.6 Summary and conclusion

In this chapter, I have discussed the functions of the discourse marker *you know* in the SLC and LOCNEC. In my data, *you know* demonstrates two functions as a non-discourse marker and six functions as a discourse marker as in Table 5.6.

---

<sup>26</sup> Kurdi's examples were translated literally into English while the marker (*btaʕfri*) was kept as is for clarity.

**Table 5.6 Summary of *you know* in the SLC and LOCNEC**

Non-DM functions of <i>you know</i>	DM functions of <i>you know</i>
<i>You know</i> as a part of question <i>You know</i> with a complement	<i>You know</i> for lexical or content search <i>You know</i> for restart and repair <i>You know</i> introducing explanation <i>You know</i> for speaker/hearer shared knowledge <i>You know</i> for generally shared knowledge <i>You know</i> in narratives

Regarding the discourse marker functions, following Brinton's (1996) classification, there were six functions of *you know*, three textual and three interactional. On the textual level, *you know* was found to signal lexical/content search. It was also found as a restart and repair. The last function was *you know* introducing explanation. On the interactional level, *you know* was used for speaker/hearer shared knowledge and for general shared knowledge. It was also found to be used in narratives.

## CHAPTER 6: *LIKE* IN THE DISCOURSE OF THE SLC AND LOCNEC

### 6.1 Introduction

This chapter consists of two parts. The first part, section 6.2, looks at how *like* has been discussed in the literature. In the second part, section 6.3 discusses how *like* is used by Saudi learners and native speakers. In 6.4, the results from the Saudi corpus are then compared quantitatively to data from NSs of English (LOCNEC). Section 6.5 presents a comparison of findings from the corpus and textbooks.

### 6.2 Functions of *like* in the literature

Since research on *like* first began appearing in the 1980s, its quotative and discourse marker functions have been identified (e.g. Schourup, 1985; Underhill, 1988). Schourup (1985) describes *like* as what he calls ‘evincive’. He claims that *like* in most cases is used ‘to express a possible unspecified minor nonequivalence of what is said and what is meant’ (Schourup, 1985, p. 42). Siegal (2001) points out that the discourse marker *like* has no defined grammatical role and seems to convey something about the speaker’s relation to what is asserted in the utterance.

Schourup (1985) points out that all the functions attributed to *like* in the literature can be subsumed under the overall function of approximation or indicating loose talk. He says:

‘*like* allows the speaker to call attention to current thought in the private world and to specify ... the tenor of what is in mind, without placing the details of the speaker’s thoughts in the shared world.’ (Schourup, 1985, pp. 35–36)



According to Schourup (1985), the discourse marker *like*, as derived from a corpus of radio talk shows and informal conversations between friends, has five uses all of which he analyses as evincive. The following are his five uses of *like*:

- 1) Before numeral expressions, as in:

I'm *like* six feet tall.

- 2) Before direct discourse, as in:

and she's *like* 'Come in here and have a beer' you know.

- 3) After questions, as in:

Eva: When I'm down here I listen to Dayton/When I'm at home I listen to Akron.

Sue: ( ) Yeah but which one *like*

Eva: W. Oh! WNXQ ...

- 4) Equivalent to 'for example', as in:

You know um - besides taking care of groups of people or - um you know uh I'm speaking in *like* a secretarial situation – where you're working for - you know you're you're h- having to - set up your time ...

- 5) *LIKE* as an interjection:

but I found *like* that helped me a lot, and this *like* – This movie takes place in 1968.

(Schourup 1985, pp. 37–63)

However, regardless of the straightforward treatment of *like* by Schourup (1985), I agree with Fuller (2003b) that the meaning of the discourse marker *like* is more elusive. Following are some arguments in this respect. In contrast to Schourup's argument (1985, p. 42), Underhill (1988, pp. 241–245) offers instances in which *like* does not mean 'approximately', 'sort of' or 'similar to'. He claims that '*like* functions as an approximator much less often than one

might think'; instead, *like* functions as a 'focuser' to mark new and significant information (1988, p. 238). For instance:

(6) Man, get in that car, *like* now.

(7) One friend to another (talking about the beach): The waves were *like* really big.

(8) Student coming in for help on a homework assignment:

I had problems *like* on the second question.

(1988, pp. 238–239, examples 29, 18 and 16)

Underhill (1988) points out that *like* in examples (6) and (7) are very direct examples in which the focus is clearly marked. Similarly, *like* in (8) focuses on the specific area of problems because it is already clear that the student had problems, otherwise he would not have been coming to seek help. However, it has been argued that there are some occasions when *like* is found to introduce numerical expressions and is clearly not an approximator but a 'focuser'. To illustrate this, Underhill provides this example:

(9) (The teacher was scheduling in-class essay exams):

Teacher (confirming): Friday at one.

Student: Change mine to Wednesday. I'll do it *like* twelve to two.

(1988, p. 245, example 71)

He states that because the exam takes two hours and is scheduled at 12, there is nothing approximate about the time; 12 to 2 is rather focus-marked (Underhill, 1988, p. 245).

Like Underhill (1988), Fuller (2003b) argues that 'many instances of *like* which indicate approximation can also be explained as focus' (p. 367). She argues that in the following

example the speaker does not want to be held to the fact that her friend is exactly eight years older than she is, but wants to focus on the age difference:

- (10) She was, she's *like*, eight years older than me ... so there was actually kind of a big age difference, but we just hit it off.

(Fuller, 2003b, p. 367, example 2)

One of the meanings of *like* as a preposition is 'for example' (Jucker & Smith, 1998, p. 184, p. 188; Schourup, 1985, p. 48). Meehan (1991) examines the major uses of *like* in chronological order according to the *OED* (Oxford English Dictionary). She claims that the discourse marker *like* is still undergoing a process of grammaticalization. Following are some of the examples she provides in her paper *It's like, what's happening in the evolution of like?*:

- (11) You go through a text *like* a Shakespearean play  
(12) ... it might be difficult if you start *like* in the fall.

(1991, p. 42, examples 11a & 11c)

However, unlike Meehan's (1991) account, I argue that there are differences between the two usages of *like* in the previous two examples. According to the criteria I established in terms of what should be classified as a discourse marker and what should not, it can be noted that *like* in example (11) bears a syntactical role and must not be omitted because its omission will render the sentence ungrammatical, whereas *like* in (12) is syntactically optional.

Commenting on examples in their article, Miller and Weinert (1995) point out that in some instances where *like* introduces examples, it also serves to give salience to the statement following them as example (13) illustrates:

- (13) A1: at the dance ... he was surrounded by the sixth-year girls we were just

tormenting him and she was sitting there going ‘tsk tsk’

L1: disapproval - I mean and if we go up and say something to to him she -

you know, not really cheeky - but she considers it cheeky and I can see

her looking and glaring at you and

A2: In disapproval and Terry's quite happily standing there taking it all

L2: *like* you keep telling to stop smoking

A3: *like* I walk up to him and say ‘How many have you smoked today’ and

he looks very guilty and says ‘too many’

(1995, p. 371, example 4)

However, Miller and Weinert (1995) argue that these functions of *like* (i.e. introducing examples and giving salience) are compatible. Despite the similarities, which are probably the result of shared historical origins, the authors’ examples appear to show evidence of multi-functionality, i.e. that there are quite distinguishable functions of *like*. Aijmer (2002), who treats multi-functionality of discourse particles as a given fact, argues that it does not ‘seem to cause problems in communication because of the presence of linguistic and contextual cues functioning as interpreting strategies’ (p. 28). Indeed, in this study, when a discourse marker appears to fulfil 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 (cf. 3.3).

Other functions of *like* identified in the literature include *like* for hedging (Jucker & Smith, 1998), for indicating a search for words and for restarting (Müller, 2005); *like* preceding an explanation has rarely been discussed except in Müller’s (2005, p. 215) study. Her results report that this function occurs 137 times in the speech of both American NSs and in that of

German NNSs. Under this main function of *like*, she also identifies two sub-types of *like* operating as a discourse marker, ‘*like* introduces a repetition of something the speaker has said before’ (2005, p. 217) as in example (14) and ‘*like* introducing an alternative term’ (2005, p. 218) as in examples (15) and (16). However, this new function of *like* and its sub-types reported by Müller (2005) do not seem convincing to me.

(14) 216 and all of the sudden it went to &

216a & a shot at the dining room,

217 .. you know *like* this random shot &

217a & [in the dining] room,

(6B, NS)

(2005, p. 217, example 200)

(15) B: what was your favorite part of it?

192 *like* the funniest part?

(27B, NNS)

(2005, p. 219, example 204)

(16) 104 so like a <L2 Fuenf-Mark-Stueck L2>,

105 *like* a .. five dollar piece or something,

(44B, NS)

(2005, p. 218, example 203)

I argue against Müller’s (2005) treatment of this particular function of *like*. By looking at *like* in examples (14) and (15), it could be said that they are not significantly different from the discourse marker *like* suggested by Schourup (1985) to be equivalent to ‘for example’. In her

example (16), Müller (2005) states that ‘[t]he alternative term he [the speaker] then chooses as an explanation (preceded by *like*) is an almost literal translation [from German] into English (line 105)’ (p. 219). However, again when looking at the two lines 104 and 105 in example (16), I argue that *like* (line 105) operates in the same way as *like* (line 104), which Müller (2005) does not classify as a discourse marker. With the limited context provided, it is difficult to judge; however, arguably this occurrence of *like* could be analysed as an approximator preceding numeral expressions and not *like* preceding an explanation for two reasons. The first reason is that *like* is immediately followed by ‘*five dollar piece*’. The second reason is the word ‘*something*’ which occurs after the numeral expression and could also mean an unspecified amount. Therefore, following, for example, Schourup (1985), Underhill (1988), Meehan (1991), Miller and Weinert (1995), Andersen (1998), Jucker and Smith (1998) and Andersen (2000; 2001), this function of *like* preceding an explanation is not included in my functional analysis.

Following Schourup (1985), Müller (2005) adds *like* preceding a restart to her functional categorization but she mentions it at the end because it occurs six times in her 350,000 word corpus and is only used by American speakers. Schourup (1985) points out that there are several instances of *like* preceding a restart occurred in his data. He defines restart as ‘a point at which the present speaker stops an item under construction and recommences’ (1985, p. 54). The following example is from Müller (2005):

- (17) 282 (H) it was just funny how he got his--  
 283 ... *like* how he sometimes leaned in there,  
 (141C, NS)

(2005, p. 218, example 216)

Underhill (1988) states that *like* ‘functions with great reliability as a marker of new information and focus’ (p. 234). In the same vein, *like* is considered a discourse organizer which has a non-contrastive highlighting/focusing function (Miller & Weinert, 1995). Underhill (1988) names it ‘marking focused information’. He argues that the information marked by *like* is ‘not only new information, but usually focused’ (1988, p. 238). Other researchers have acknowledged this function of *like*. For example, Romaine and Lange (1991) distinguish two functions of *like* in English: *like* as a quotative and *like* as a non-contrastive focuser. Meehan (1991) includes focusing *like* as one of several functions of *like*. She discusses the semantic value of the focuser *like* in the utterance and provides two examples, one with *like* and the same example without *like*:

(18) She’s *like* really pregnant.

She’s really pregnant.

(1991, p. 46, example 15)

Meehan (1991) argues that there is a difference between the two examples. I agree with her that ‘[w]here *like* acts like a focus marker and is removed, the focus of the sentence is more likely to be placed on the subject “She” rather than the adjectival phrase describing the state being’ (1991, p. 46). Speaking about the scope of *like* in this function, Müller (2005) states that it covers noun phrases, verb phrases, adverbs and adjectives. She emphasizes that ‘it never has a full clause in its scope. Therefore, I have called this function of *like* “marking lexical focus”’ (2005, p. 221). From her data, instances of *like* to focus what follows are as follows:

(19) 268 (H) and he *like* .. picks her up,

269 .. a=nd takes her in there.

270 (H) and u=m .. that was *like* the end.

271 then they play .. *like* music.

(8B, NS)

(2005, p. 221, example 207)

Previous studies have indicated that age generally has an influence on the frequency of *like* as quotative and as a discourse marker (Andersen, 2001; Blyth et al, 1990). Blyth et al (1990) demonstrate that quotative *like* was employed with the highest frequency by far by the group of speakers aged 20–24 as opposed to those older than 38 years, who did not use it at all. However, my participants were all classified in one group aged 20–25, i.e. in the target population for LINDSEI, namely third and fourth year undergraduate students. Therefore, this variable is not discussed further because there is no other age group to which it can be compared. Other variables, such as gender and a period of stay in an English speaking country, are considered.

In the following sections, I examine the functions of *like* and their frequencies as they occur in the data from the SLC and LOCNEC. I begin with the qualitative analysis.

### **6.3 Functions of *like* in the SLC and LOCNEC**

*Like* in my data and in LOCNEC appears in three broad functional groups: as a quotative, as a discourse marker and as a non-discourse marker which subsumes verbs, nouns, adjectives, adverbs, prepositions, conjunctions and fixed expressions, e.g. *stuff like that*. In the next subsections, I discuss the functional categories of *like* which operate as a non-discourse marker and as discourse markers.



### 6.3.1 Non-discourse marker functions of *like*

Müller (2005) and Fox Tree (2006) argue that there are three major non-discourse marker uses of *like*: *like* as a verb, preposition and conjunction, as well as in a fixed expression, as illustrated in the examples (20), (21), (22) and (23) respectively from the SLC and LOCNEC data:

(20) <A> oh that's nice </A>

<B> sort of attacked on two fronts if you *like* and I've already got past the personnel manager so if . if anybody goes to the personnel manager and says what do you think of this person she can al already say yeah I *like* her cos she's already seen my C V so </B>

<A> (mm) </A>

<B> it's </B>

<A> that's nice </A>

(LOCNEC-E15)

(21) <A> what happened in picture number two </A>

<B> picture number two she saw he= her drawings his drawing of her </B>

<A> (uhu) </A>

<B> she didn't like it </B>

<A> because </A>

<B> because she is exactly *like* her and since I thought she she she is not pleased for herself so she asked him to make it more beautiful and apparently she is very ugly from the drawing </B>

<A> yeah </A>

(SLC-AR4)

- (22) <B> certain parts of .. er that he's drawn and er he looks very . shocked at this  
(erm) so in the third picture he's . he's re= repainting the picture .. but he's looks he  
looks *like* he's drawing it so it doesn't really look *like* her <\B>  
<A> no <laughs> <\A>  
<B> he's glorifying the picture . and then <laughs> er in the last frame er the[i:] the  
woman is showing . the finished . the finished product to her friends and there is  
<X> it looks *like* the friends are . dubious <XX> <\B>  
<A> <laughs> <\A>  
<B> that that for some reason there's this glamorous woman in the picture and er ..  
she just happens to look vaguely similar to: er <\B>  
<A> (mhm) <\A>  
<B> to the woman <\B>  
<A> (mhm) okay <\A>

(LOCNEC-E04)

- (23) <A> Does he feel that you still little </A>  
<B> No no he maybe his friends I don't want to embarrass him me myself </B>  
<A> But you are not an embarrassing person </A>  
<B> Ah unless if you weren't shaving you mustache unless if you weren't wear a  
jeans unless if you wear a skinny you know like stuff *like* that way </B>  
<A> uhu</A>

(SLC-AR09)

There are also instances of *like* preceded by forms of ‘be’, for quoting:

- (24) <B> yeah almost but it’s not that the thing the thing is become like really I dunno embarrassing accidentally I dunno I always walk the guy he so formal and so polite  
</B>  
<A> (uhu) </A>  
<B> and when there like I I don’t c= I’m caring enough but we are like in the mall it’s so quite suddenly I dunno some way he is just *like* <foreign> ittihad </foreign> I dunno some I dunno </B>  
<A> so he start clapping </A>  
<B> yeah yeah I dunno it’s something like </B>  
<A> but do you think that was because of (ur) he was(ur) like (ur) </A>  
<B> because he believes like (ur) th= I dunno how to say it like went I dunno don’t remember the sing (ur) saying like <foreign> inta 3reeb ya3ni</foreign> </B>  
<A> if you if yeah </A>

(SLC-AR3)

In all the examples highlighted above, *like* occurs as a non-discourse marker. Contrary to the instances in which *like* functions as a discourse marker, the removal of *like* affects the grammaticality and the semantic meaning of the sentences in which it appears. Therefore, following Müller (2005), Fuller (2003a; 2003b) and Dailey-O’Cain (2000), I exclude the attribution of discourse marker status to quotative *like* because it cannot be left out without rendering the utterance ungrammatical. In the following subsections, I focus on the discourse marker functions of *like*.

### 6.3.2 Searching for the appropriate expression

One of the widely recognized functions of *like* is to search for an appropriate expression. Schourup's (1985) description of this use of *like* as a 'pausal interjection' (p. 53) seems appropriate. He provides a plausible explanation of why *like* could be used in such places in which 'the material about to follow is difficult to formulate appropriately or precisely' (1985, p. 56). Andersen (2001) explains the circumstances in which the speaker uses *like* to search for an appropriate expression as if he/she is saying 'I have something on my mind, but I don't know how to put it' (p. 249). In less academic terminology, the Cobuild (1995) and Macmillan (2002) dictionaries similarly argue that this form of *like* can be used when the speaker is 'thinking about what to say next', i.e. searching for the appropriate expression.

Extract (25) from my SLC data shows that the speaker AR3 uses the discourse marker *like* to search for the word 'roommates' as shown by the fact that *like* is followed by a filled pause (er):

- (25) <A> and what was your decision when you came back because this seems to me a good experience about knowing people <overlap /> </A>  
<B> <overlap /> yeah yeah </B>  
<A> in depth </A>  
<B> th= and I know we went there you know because we are *like* (er) roommates so we buy the things for refrigerator together and </B>  
<A>(er) </A>  
<B> and he doesn't want to buy anything like he doesn't wa- ok I go and buy then we will split you know fifty fifty </B>  
<A> right </A>

(SLC-AR3)

A similar example can be found in the native speaker LOCNEC data. In example (26) we can see that the discourse marker *like* is followed by an empty pause for two seconds, which works as a contextual clue that the function of *like* in this utterance is to search for an appropriate expression/word:

- (26) <A> (mhm) .. (er) I mean . since you like Disney films and that kind of stuff have  
you already been to Disneyland or <\A>  
<B> no no <\B>  
<A> Eurodisney or <\A>  
<B> no <\B>  
<A> no would you like to go there <\A>  
<B> (erm) .. I don't know <\B>  
<A> you don't know why <\A>  
<B> well .. this is sort of *like* .. childish sort of thing to it isn't it no . well I don't  
know I I would like to go but <\B>  
<A> (uhu) <\A>  
<B> I don't think I'll ever admit to liking to go <laughs> <\B>

(LOCNEC-E02)

As another example, extract (27) shows that the speaker AR13 has clearly encountered difficulties finding a suitable term to describe precisely what she wants. She wants to express how passionate she was about doing voluntary work even though it was summer time, when the weather in Saudi is normally extremely hot. When she tries to use the appropriate word *field*, it seems that she struggles to find it. So, she repeats the word *other* twice to then say *section*. However, she realizes that *section* is not the word she is searching for. She eventually

uses *other* followed by the discourse marker *like* to search for the appropriate word, which in this case is *field*:

(27) <B> yeah and you know it was summer </B>

<A> (uhu) </A>

<B> so I basically didn't have life that time I used to come really tired and just you know sleep but I would definitely volunteer in other other section other *like* field yeah </B>

(SLC-AR13)

Müller (2005) states that for her to classify *like* as an indicator of search, there had to be at least one other hesitational marker. Such hesitational markers are, for example, 'filled and unfilled pauses, lengthening, truncated words, or hedges such as *kind of*' (2005, p. 209). However, while this seems a reasonable clue to use, there is another marker which has not been mentioned here: repetition. Johnstone (2008) points out that repetition can be a useful tool which provides 'speakers who are planning as they go with a reusable template and thus minimizing hesitations and fillers while allowing people to keep the conversational floor as they think of what to say next' (p. 173). Therefore, in my study, I add repetition as a hesitational marker which indicates that the speaker is in the state of searching for something, whether lexical or content-related. Example (28) illustrates that the repetition of the word *other* three times is a contextual clue suggesting that the speaker is using *like* to search for an appropriate expression, *field*. This is an example of *like* as a marker of a lexical search.

Extract (28) is an interesting case. Speaker AR11 was telling the interviewer about what he has achieved so far and how satisfied he was with those achievements. Then he said he was looking forward to teaching at an academic level. He uses the discourse marker *like* to search for the word *levels*. It could be concluded that *like* is used to search for a word because it is

preceded by another discourse marker, *you know*, which also serves the function of thinking about what to say next (Cobuild, 1995; Macmillan, 2002).

- (28) <B> I was satisfied with my with what I have done </B>  
<A> Yeah I can imagine you can do things </A>  
<B> so this pushed me <foreign> inshallah </foreign> to complete my studies  
which I have to the masters </B>  
<A> you must do that </A>  
<B> and I I'm looking forward to teach academic you know *like* level levels not  
you know the usual school levels </B>  
<A> that was the thing I was about to say </A>

(SLC-AR11)

In my data, just as Müller's (2005), there are some instances where speakers who use *like* manage to find the words they are looking for and some other instances where speakers fail to do so. Thus, the following example shows that the use of the discourse marker *like* to search for a word is not always successful. The speaker in example (29) is trying to remember a saying. The discourse marker *like* is followed by a filled pause (*er*) and a lengthened syllable, but eventually the speaker fails to remember the saying and says instead '*I dunno how to say it*':

- (29) <B> and when there like I I don't ca: I'm caring enough but we are like in the mall  
it's so quite suddenly I dunno some way he is just like <foreign> itthad </foreign>  
I dunno some I dunno </B>  
<A> so he start clapping </A>  
<B> yeah yeah I dunno it's something like </B>  
<A> but do you think that was because of (er) he was er like (ur) </A>

<B> because he believes *like* (ur) th: I dunno how to say it like went I dunno don't remember the saying (er) saying like <foreign> inta ghareeb yaghni </foreign>  
</B>

<A> if you if yeah </A>

(SLC-AR3)

### 6.3.3 Marking an approximate number

As has been mentioned in the literature, one of the main functions of *like* is hedging with numeral or quantitative expressions (see Anderson, 2001; Jucker & Smith, 1998; Meehan, 1991; Müller, 2005; Schourup, 1985). When *like* precedes a numeral expression, it indicates that the number is not being treated as an exact one. So, in a case like this, *like* helps the speaker to reduce their commitment to the truth of their utterance (Müller, 2005). Extract (30) from my data provides an example of this:

- (30) <A> there were a lot of buildings and towers I didn't expect that I'm in Dubai like I expected that I'm in New York for example although I didn't visit New York but I saw it in the movie okay there are a lot of towers a lot of buildings a lot of high buildings very high buildings just like Khalifah towers and Borj Arab a lot of a lot of towers there <\A>  
<B> it was beautiful <\B>  
<A> yeah of course (er) and about the malls <\A>  
<B> yeah <\B>  
<A> the malls were so so many there and I visited I visit *like* four malls Dubai mall <\A>

(SLC-AR19)



Speaker AR19 in extract (30) describes her visit to Dubai. She was so impressed by what she had seen and the massive progress Dubai had achieved in a short period of time. She was speaking about the malls and how happy she was to be able to have the opportunity to see all these places. However, she was not quite sure exactly how many malls she had visited and so she used the discourse marker *like* to be approximate and to avoid giving an inaccurate number. Even though *like* is preceded by a repetition which I have argued suggests ‘searching for the appropriate expression’, it seems less significant in the case of *like* preceding a numeral expression.

Example (31) from LOCNEC includes two instances of the discourse marker *like* marking an approximate number. Speaker (46) is describing her visit to Brazil and the weather in one of the cities which she visited. It can be observed that the speaker does not use any discourse marker before ‘hot humid’; however, the discourse marker *like* is used before numbers.

- (31) <B> so they'd be literally living over the river <\B>  
 <A> (mhm) <\A>  
 <B> sort of like reflecting the tribes of south east Asia and things . you know it was that wasn't a very pleasant place <\B>  
 <A> (mhm) <\A>  
 <B> I didn't like <name of city> at all . but saying that it was also exceptionally hot humid . and (erm) <\B>  
 <A> I guess you're not used to it when you come from Britain <laughs> <\A>  
 <B> (er) I was quite used to it by then because I'd been there for a couple of months <\B>  
 <A> (mhm) <\A>  
 <B> but you know this was stupidly hot *like* fifty <\B>

<A> . fifty degrees </A>

<B> yeah </B>

<A> oh my god </A>

<B> and (erm) it's *like* a hundred percent humidity .. <X> </B>

(LOCNEC-E46)

In the following example (32), speaker AR9 talks about what he usually does with his salary. He says that he withdraws ‘*like three hundred or two hundred and fifty*’ to give to his mother to save for him. He uses *like* here to indicate approximation, i.e. that it is not a fixed amount of money paid every month. Similarly, in example (33) from the native speaker data, *like* indicates approximation. The speaker is not sure exactly how many lines he/she could write in the dark. *Like* was used here to help the speaker to reduce her commitment to the precise number of lines which was not important information in this context.

(32) <A> so your mom is doing this role </A>

<B> yeah even my: money you know (er) to (er) like when I got my s= my salary  
<foreign> mukafaty </foreign> I= have to you know like withdraw *like* three  
hundred or two hundred fifty and give to my mom so that she give it to me <x>  
fifteen ten fifteen </B>

<A> Yeah </A>

(SLC-AR9)

(33) <B> cos for two years I had to make notes when I went to see a play and had to  
think about it and </B>

<A> oh every time </A>

<B> well we: we . kind of . didn't towards the[i:] end </B>

<A> (mhm) </A>

<B> but I mean we did generally make notes and it's awful cos you get in<?> there and you can't see so you're just writing without looking and you've got *like* four lines written on one line and you can't really proper <?> read it </B>

<A> (mhm) </A>

(LOCNEC-E09)

Another example using this function of *like* is given in extract (34). Speaker AR6 chooses to talk about an important issue she was experiencing at the time of the interview. This issue was a family problem which resulted in her mother leaving the house. The speaker says '*and it's like three months already*'. The number of months is preceded by the approximator *like* because the speaker is not sure exactly how long her mother has been away from home but still she wants to give an idea of how tough it is when her mother is not around anymore.

(34) <A> ok great so you gonna talk about an experience you have had which has taught an important lesson </A>

<B> (uhu) </B>

<A> ok </A>

<B> it's actually (er) due to time it's current (erm) my family had has a problem (er) and they are almost separated </B>

<A> oh oh I'm sorry </A>

<B> it's ok my mom (er) left the house and it's *like* three months already </B>

<A> (uhu) </A>

(SLC-AR6)

The previous examples present the discourse marker *like* used before numerical phrases. In the following examples, we will see examples of *like* used before quantity phrases. In

example (35), speaker AR4 uses the discourse marker *like* to avoid committing herself to giving an exact amount in terms of the money he lady paid to the artist to have him paint her in the way she wanted to appear. In fact, the exact value does not matter in this instance because the speaker believes that giving the hearer a rough idea of it will suffice.

(35) <A> maybe yeah </A>

<B> okay she wanted an artist to to ask him to draw her and so I guess he took a long time to draw her exactly as she is (er) but she seems like she didn't like it very much she was lying arguing with him may be about the the price she paid *like* a lot and she didn't she did in being say </B>

(SLC-AR4)

Similar to the previous example, speaker E13 wants to demonstrate to the interviewer how boring his second work experience was. He uses *like* to emphasize that he will not repeat this experience for a long time. Giving a precise length of time is not the key point in this discussion.

(36) <B> yeah .. and then I went to Germany that was when I went to Stuttgart . and . worked in a nursing home </B>

<A> oh yes </A>

<B> for six months .. which was a really good experience I really enjoyed that but then again it's another thing I don't want to do </B>

<A> (mhm) </A>

<B> for *like* a long time (erm) </B>

<A> for your entire life </A>

(LOCNEC-E13)

### 6.3.4 *Like* introducing an example

The usage of *like* as an exemplifier has been acknowledged in the literature (Andersen, 2001; Meehan, 1991; Miller & Weinert, 1995; Müller, 2005; Schourup, 1985). There are two types of *like* as an exemplifier. The first type is when *like* means ‘for example’ but is realized as a conjunction. An example of this is given by the *Oxford Advanced Learner’s Dictionary* (1989): ‘We could look at some modern poets, *like* Eliot and Hughes’. It can be said undoubtedly that *like* in this example means ‘for example’. However, this *like* should not be counted as a discourse marker because it is syntactically immobile and cannot be omitted without rendering the sentence ungrammatical. The second type of *like* which introduces examples and functions as a discourse marker is given in example (37). Extract (37) is from the native speaker’s data LOCNEC. The speaker E03 in this example discusses with the interviewer what he thinks he will do in the future. He says that he is thinking about advertising or ‘say *like* article <X> writing’. The discourse marker *like* which he uses functions as an exemplifier.

- (37) <A> oh how interesting <starts laughing> you must have a lot of imagination  
<stops laughing> <\A>  
<B> <laughs>.. yeah but (erm) .. I don't know I enjoy it but I'm not I don't think  
I'm particularly good . so I might do that as a hobby <\B>  
<A> uhu <\A>  
<B> but probably think about going into advertising <\B>  
<A> oh yeah <\A>  
<B> or say *like*<X> article writing <\B>  
<A> (mhm) <\A>  
<B> something not <\B>  
<A> for a newspaper <\A>
- (LOCNEC-E03)

Example (38) illustrates another case in which *like* functions as a discourse marker to mean ‘for example’. In this example, the girl speaks about her experience of living abroad in Paris and how this experience helps her to build up her self-confidence. She explains to the interviewer how she managed the household responsibilities. She says that she used not only to cook for her parents but also clean their home; this is another example of household responsibilities she provides. So, *like* is used to mean ‘for example’ and is considered grammatically optional.

- (38) <B> thank you so these this experience helped me alot that (er) to build my own confidence with to deal with my husband <\B>  
 <A> uhu <\A>  
 <B> the way I ma= manage er the home the way (er) you know when I was there the= we didn’t have a servant so I use to cook for my parents *like* clean the home buy the gro= groceries so (er) I can <\B>  
 <A> so it wasn’t difficult for you when you got married <\A>  
 <B> yeah this is it <\B>  
 <A> yeah <\A>

(SLC-AR04)

The speaker AR1 in example (39) starts explaining how difficult the life of the old generation was. She speaks about her parents’ generation in particular. Then she wants to give an example for illustration. She says that back in their time “*like* TV was something sound great”. In her sentence, she uses the discourse marker *like* to introduce an example. This *like* is grammatically optional, i.e. the sentence would still be well-formed without it.

- (39) <A> so good so talk about differences in generation what else can you say about that </A>

<B> well their life was like really difficult and now we have more entertainment  
you know what you call it ways or options </B>

<A> (uhu) </A>

<B> but back in their time *like* TV was <starts laughing> something sounds great  
</B>

<A> (uhu) </A>

(SLC-AR1)

In extract (40) below, speaker E18, from the native speaker LOCNEC data, speaks about where she is planning to teach in the future. She would like to teach abroad for some time and then come back to England. She gives some examples to describe the weather of other countries. She uses the discourse marker *like* to give examples, such as earthquakes and insects. Again, the *like* in this example is optional.

(40) <A> <XX> what are you planning on doing after you've finished <\A>

<B> (erm) <\B>

<A> if you know <\A>

<B> well I'd I mean I'd like to teach eventually . teach English <\B>

<A> English <laughs> <\A>

<B> yeah (erm) either as a foreign language or within England I'm I'm not I don't  
really mind so I'm probably gonna do a teacher training course <\B>

<A> (mhm) <\A>

<B> and probably do some courses in teaching English as a foreign language <\B>

<A> (mhm) <\A>

<B> and then .. hopefully I'll start teaching in England and if you know as I get  
more confident and sort of like if you know I've got some experience I might go

abroad and try teaching there <\B>

<A> (mhm) .. that would be nice .. just for a few years or for . your entire life  
would you like to: <\A>

<B> (erm) no just just for a few years I think I'd I mean well I could I don't think I  
could like live abroad for all my life I think I'd want to come back to England <\B>

<A> you would miss it <laughs> <\A>

<B> yeah <laughs> it may have bad weather but it doesn't have *like* earthquakes  
and little insects that bite you and <begin\_laughter> all that kind of thing  
<end\_laughter> <\B>

<A> that's true that's true yes <\A>

<B> yes yeah <\B>

(LOCNEC-E18)

In the discussion in extract (41), the speaker expresses his opinion concerning when children can be independent. He says that they can be independent after they get a qualification. This qualification could be from a college, institution, etc. Speaker AR11 uses the discourse marker *like* to provide an example for illustration, 'graduate *like* from college'.

(41) <A> would you like your sons to be like that in the future when you get married  
and you become old you have sons you want them to be strong <\A>

<B> of course I would help them until they became independent <\B>

<A> you don't want them to be independent and work in their own <\A>

<B> after they graduate *like* from college and start to get a job then they are free to  
get <X> <\B>

(SLC-AR11)



### 6.3.5 *Like* marking focused information

Another function of the discourse marker *like* is to focus on the following information. Regarding the focuser *like*, I argue, along with Underhill (1988) and Fuller (2003b), that there are many instances of *like* which indicate approximation but can also be explained as having a focusing function at the same time. Fuller (2003b) states that ‘the overlap of these two functions makes sense; there is no reason to indicate approximation of something that is not important in the sentence, and conversely, if something is the focus of the sentence, you want to make sure that, if approximate, this looseness of meaning is marked’ (p. 367). Therefore, I argue that with numbers, *like* is not always an approximation marker, especially when the speaker is giving specific information (cf. Underhill, 1988).

Let us consider the following example from the SLC. In this extract (42), speaker AR13 answers the interviewer’s questions about whether she would repeat the experience of working during the summer vacation. The speaker, in turn, shows her willingness to repeat it but not in the medical field because of one main problem she emphasizes – the working hours. When she mentions the time ‘from like eight to five everyday’, it is unlikely she is unsure about when her work started and finished. She said that the working hours were ‘*too long*’. This is a clue to illustrate that the speaker uses *like* in this example to highlight the starting and finishing time. Therefore, I conclude that this discourse marker *like* functions as a focus not as an approximator although it is followed by numbers.

(42) <A> so <x> final question would you repeat this experience again <\A>

<B> (er) I would but not in the medical field not because like I didn’t like it but because (er) the working hours were too long for me because I used to work there from *like* eight to five everyday <B>

<A> oh that’s true <\A>

(SLC-AR13)

In another interesting example from LOCNEC, speaker E10 chooses to speak about one of her experiences of films. She describes this experience as a particularly bad one. She says that she did English literature A-level in which King Lear was a set text. The play lasted for four and a half hours. In this excerpt, she mentions the interval, which came after the first hour. Her focus was on the remaining time left. She uses the discourse marker *like* to describe the important part in the sentence, the three and a half hours solid.

- (43) <B> yeah .. and (erm) .. there was that one interval but the[i:] interval came too early on <\B>  
<A> an interval <\A>  
<B> yeah .. that was like after the first hour <\B>  
<A> (uhu) <\A>  
<B> and then it was *like* three and a half hours solid after that and that's a long time to sit <\B>  
<A> oh yeah <\A>  
<B> and keep your concentration and like . they had like . bits weren't weren't necessary like between the play between the[i:] er <XX> they were changing stuff over . they'd have (erm) . a little musical bit which wasn't part of the play at all <XXX> <\B>  
<A> oh yes just because there was a short= a shortage of actors and they had to: to: to get changed I guess <\A>  
<B> yeah .. yeah that's right <\B>

(LOCNEC-E10)

The speaker AR12 in the extract below (44) talks about his brothers, particularly those who can speak English and those who cannot. The interviewer asks him whether his brothers are younger or older than him. The speaker wants to give a detailed answer, saying ‘and I have *like* (er) five older than me...’. Again *like* is clearly used to focus on the following information as it is highly unlikely for someone to give an approximate number of how many brothers he/she has.

- (44) <A> brilliant brilliant that’s really good do your brother speak with you some English at home <\A>
- <B> yes we do that a lot <\B>
- <A> do they speak English better than you or <\A>
- <B> some and some some of them <\B>
- <A> are they younger than you or older than you <\A>
- <B> no I’m before the youngest I’m I have the only younger brother <\B>
- <A> (uhu) <\A>
- <B> and I have *like* (er) five older than me two of them don’t speak the la= the English language don’t like it <\B>
- <A> oh really <\A>
- <B> two of them are really good in English one of them is better than me <\B>

(SLC-AR12)

In the following extract from LOCNEC, speaker (E10), an English literature student, speaks about a film called Jane Eyre. She does not like this film because, as she clarifies, it was preceded by something she loved a lot ‘a production of Wuthering Heights’. The speaker uses the discourse marker *like* to focus on how she views the production of Wuthering Heights. She says ‘it was *like* really really well done’.

- (45) <B> not you know I am Jane Eyre but I am acting sort of thing but I think the thing is <XX> spoil because the first thing we ever went to see was like a production of Wuthering Heights <X> <\B> <A> oh yes (mhm) <\A>
- <B> in Chester and it was like a professional cast it had like that bloke out of Casualty <XX> . actors people have heard of <\B> <A> oh yes <\A>
- <B> and it was *like* really really well done so like I sort of judged everything by that <\B>
- <A> (mhm) <\A>
- (LOCNEC-E10)

A similar example of *like* marking a focus on information is given in excerpt (46):

- (46) <A> (uhu) do you like the buildings here <\A>
- <B> I don't mind them to be honest <\B>
- <A> you don't like them <\A>
- <B> no <\B>
- <A> no <\A>
- <B> no . yeah I quite . I quite like them cos I'm I'm living in Pendle so <\B>
- <A> (mhm) . oh yes <\A>
- <B> that's *like* really nice there and <XX> seem to see you know <X> walk up the spine and .. I don't know it's nice it's modern really <\B>
- <A> (mhm) <\A>
- <B> well fairly modern <\B>
- <A> yeah fairly <\A>
- (LOCNEC-E30)

Another instance is from the Saudi data in the SLC. Speaker AR11 in extract (47) comments on the picture description, which is the last task in the interview. He points out how different people are in accepting the truth in general and the truth of how they look in particular. He says that ‘some people won’t like accept the truth’. Here, the discourse marker *like* focuses on what comes next – ‘accept the truth’ – which is the main point of his utterance.

(47) <A> you do you do what you think is suitable </A>

<B> okay what I think here is some people when they go to others and ask them for opinions on on on about them they might tell them the truth but some people won’t *like* accept the truth </B>

<A> (uhu) </A>

(SLC-AR11)

### 6.3.6 *Like* preceding a restart

Following Schourup (1985), Müller (2005) adds this function of *like* to her functional categorization but she mentions it at the end because it occurs six times in her entire data and is only produced by American speakers. Schourup (1985) points out that there were several instances of *like* preceding a restart in his data. He defines a restart as ‘a point at which the present speaker stops an item under construction and recommences’ (p. 54).

In both my SLC and LOCNEC data, this function appears five times, four in the SLC and one in LOCNEC. The following is an example of this function:

(48) <B> I don’t know maybe be= because actually this this the point we we do not know when they are so strict like that the student w= would not feel that he is calm and he can’t do very well he can’t do his best so actually I recommend that the

professors be (er) be much more f: friendly *like* more friendly than this so the students feel free and can do his best (er) </B>

(SLC-AR5)

This is another case in which *like* occurs as a restart. However, in this example, the discourse marker *like* restarts the utterance with more information ‘a postgraduate certificate’.

- (49) <B> and then back up there in our third year and fourth year and we do lots of  
 </B>  
 <A> four years </A>  
 <B> (mm) </B>  
 <A> oh that's interesting </A>  
 <B> it's yeah there's a lot more <X> a lot more work to do because you have to  
 you cover so much you know all the different areas of the curriculum </B>  
 <A> (mhm) (mhm) </A>  
 <B> but you can there's two ways of going into teaching you can do this degree  
 which is four years or you can do a normal degree in one subject and then do a . a P  
 G <X> *like* a postgraduate certificate </B>  
 <A> (uhu) </A>  
 <B> in education at the[i:] end </B>

(LOCNEC-E12)

### 6.3.7 Unclear instances

Instances of *like* were coded as ‘unclear’ in contexts which were acoustically unintelligible, incomplete, or with all the words available but still difficult to classify functionally. In extract (50) it is hard to judge the function of the first *like* due to the unintelligible passage after it. It

was followed by an empty pause and a preposition ‘and’, then a filled pause ‘erm’, then another empty pause. Similarly, in the second extract (51), the function of *like* cannot possibly be gauged because of the passage which followed it.

(50) <A> yeah so is it your first year at university <\A>

<B> yeah <\B>

<A> (mhm) . and why come did you here to study linguistics <\A>

<B> <X> .. right well .. (erm) I went to Blackburn College to do my A-levels and I did English language <\B>

<A> (mhm) <\A>

<B> and eh . one of my teach= one of my tutors who was there .. (erm) called (erm) . <name of lecturer> she: . was familiar . with Lancaster University and the Linguistics Department .. and I said to her <X> I was quite interested in doing linguistics at . degree level and she said oh .. apply to Lancaster it's got a very good linguistics department <\B>

<A> (mhm) <\A>

<B> <X> one of the best in<?> the country and I thought well fair enough<?> and I applied to a few others *like* .. and (erm) .. I applied to: .. Bangor as well . in Wales <\B>

(LOCNEC-E02)

(51) <B> yeah and he still you know the thing is what makes me furious about it *like* the the and when I start two thousand eight two thousand I think ten was my friend wedding and we went there another friend came to me and said you know <first name of interviewee> I'm sorry and really embarrassed to talk to you about your friend <foreign> flaan </foreign></B>

(SLC-AR3)

## 6.4 Quantitative results for functions of *like*

**Table 6.1 Functions of *like* and their abbreviations**

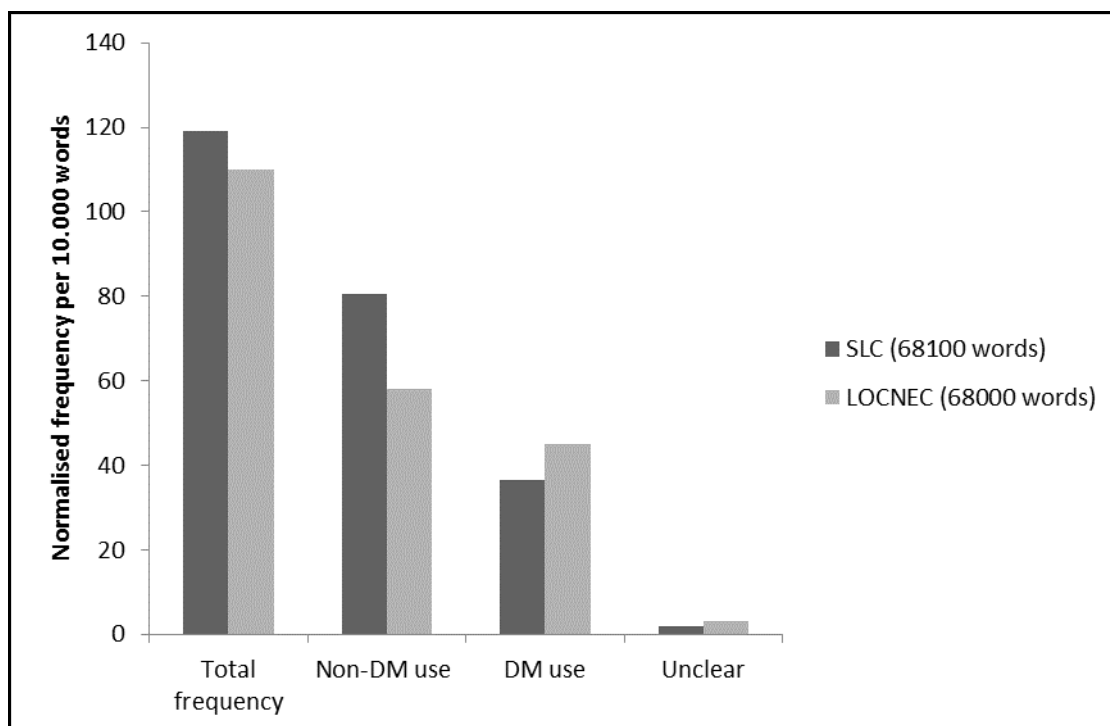
FUNCTION	ABBREVIATION
Searching for the appropriate expression	Search
Marking an approximate number	Approximation
Introducing an example	Example
<i>Like</i> marking focused information	Focus
<i>Like</i> preceding a restart	Restart

Before discussing the functions of the discourse marker *like*, I shed light on the occurrence of *like* in general in both corpora. Table 6.2 followed by Figure 6.1 demonstrate the distribution of the raw frequency of *like*, the non-discourse marker *like*, the discourse marker *like*, and finally unclear instances of *like* in the two corpora. Regarding the raw frequency and the non-discourse marker use, it can be observed that the Saudi speakers used *like* slightly more than the native speakers. However, when it comes to the discourse marker *like*, we find that the native speakers used *like* more than Saudis. The unclear instances of *like* were found more in the native speaker data than in the Saudi corpus. This may be due to the fact that because I transcribed the entire Saudi learner corpus, I became more familiar with it.

**Table 6.2 Numbers of tokens of *like* in the SLC and LOCNEC (frequencies per 10,000 words in parentheses)**

	Total Frequency	Non-DM use	DM use	Unclear
SLC (68,100 words)	808 (119)	547 (80.4)	249 (36.5)	14 (2)
LOCNEC (68,000 words)	751 (110)	396 (58)	307 (45)	22 (3)



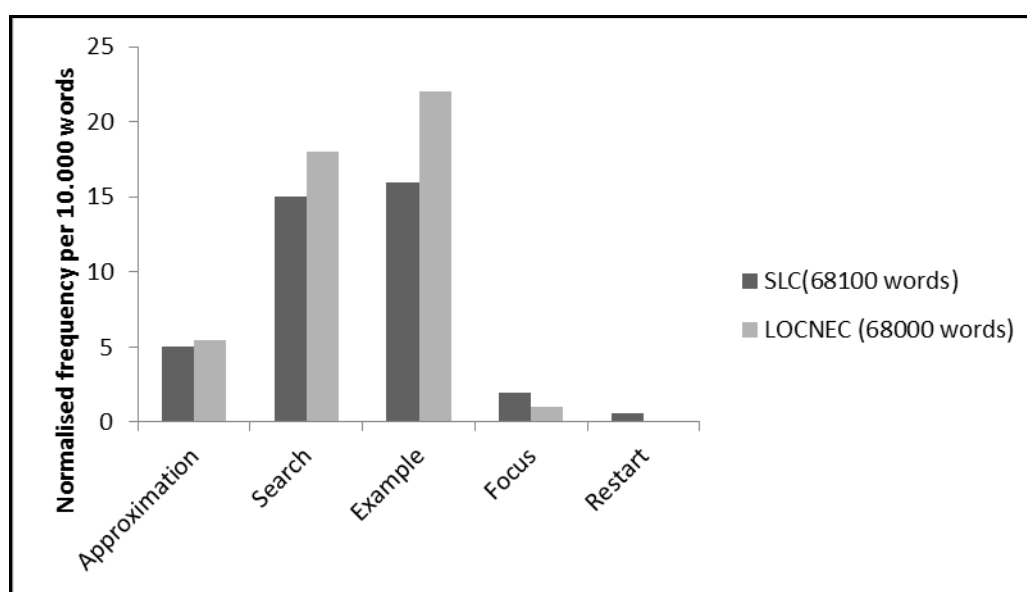


**Figure 6.1 Distribution of the linguistic item *like* – SLC versus LOCNEC**

The quantitative results of the comparison between NSs (LOCNEC) and Saudi English language speakers (SLC) at the level of the individual functions of *like* are summarized in Table 6.3 and Figure 6.2. In Figure 6.2, we can see that ‘introducing an example’ was the most frequent discourse marker function among the Saudis (111 tokens, or 14.5 per 10,000 words) and the NSs (149 tokens, or 22 per 10,000 words). The second most frequent discourse marker function, almost similar in frequency to the previous one, was the use of *like* in searching for an appropriate expression. Saudi speakers used it 103 times or 15 per 10,000 words, whereas NSs used it 114 times or 18 per 10,000 words. In both corpora, *like* for approximation came the third with almost 60% less use than the previous discourse marker function. The use of *like* for focusing on information was employed more by Saudi speakers (11 tokens, or 2 per 10,000 words) than by NSs who used this function 8 times. There were 14 unclear occurrences of *like* in the SLC and 19 occurrences in LOCNEC. However, the use of no function was found to achieve statistical significance.

**Table 6.3 Distribution of the discourse marker *like* in the SLC and LOCNEC**  
(frequencies per 10,000 words in parentheses)

	Approximation	Search	Example	Focus	Restart
SLC (68,100 words)	34 (5)	103 (15)	111 (16)	11 (2)	4 (0.60)
LOCNEC (68,000 words)	38 (5.50)	114 (18)	149 (22)	8 (1)	1 (0.10)

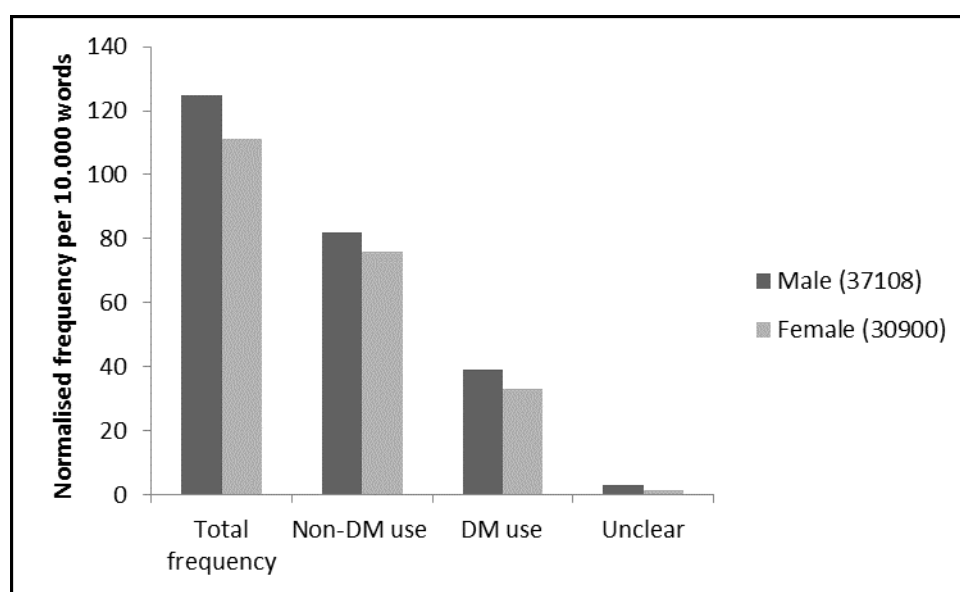


**Figure 6.2 Distribution of *like* as discourse marker – SLC versus LOCNEC**

With regard to the discourse function of *like*, the statistical analysis shows that gender has little impact on the functional distribution: there is a high level of similarity between the males and females. Figure 6.4 illustrates that males and females both used *like*, with males using it slightly more than women (145 tokens, 39 per 10,000 words vs. 101 tokens, 33 per 10,000 words). Although, unexpectedly, these results differ from some previous studies (e.g. Dailey-O’Cain, 2000; Ferrara & Bell, 1995; Fuller, 2003b; Romaine & Lange, 1991, Siegel, 2002), they are consistent with Müller (2005). Her data show that American females used *like* less frequently than their male counterparts. In contrast, German females used *like* more often than German males (Müller, 2005).

**Table 6.4 Number of tokens of *like* in the SLC according to gender (frequencies per 10,000 words in parentheses)**

	Total Frequency	Non-DM use	DM	Unclear
Male (37,108 words)	463 (125)	304 (82)	145 (39)	10 (3)
Female (30,900 words)	344 (111)	236 (76)	104 (33)	4 (1.30)

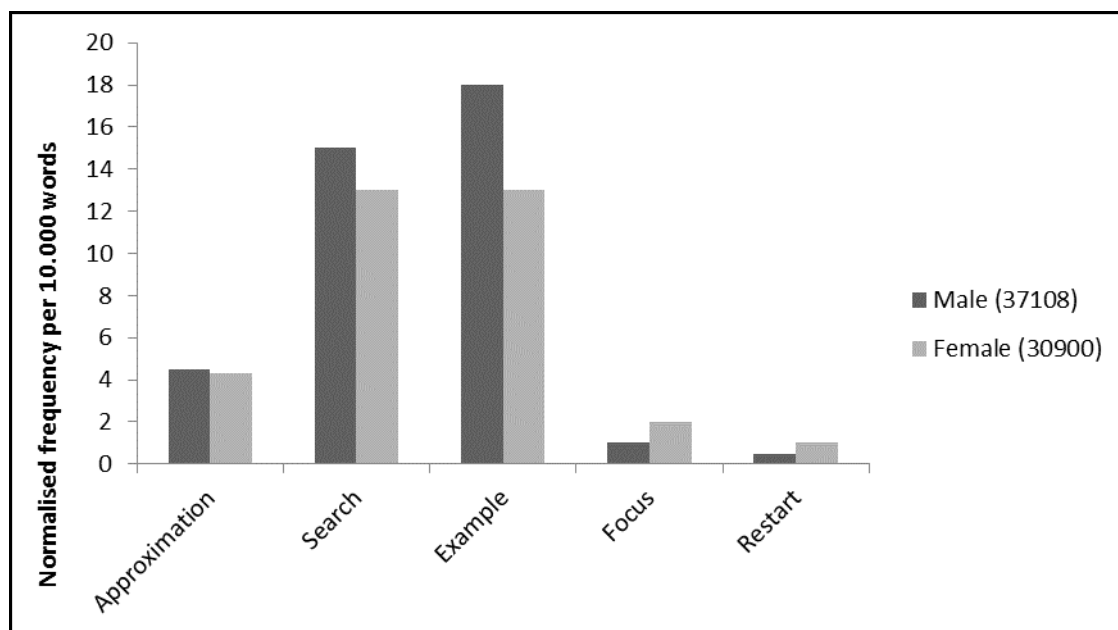


**Figure 6.3 Distribution of the linguistic *like* in the SLC according to gender**

Table 6.5 and Figure 6.4 provide more detail concerning how males and females employed the discourse marker *like*. It can be observed that males used three functions out of five more than females: approximation, search and example.

**Table 6.5 Functions of *so* in the SLC according to gender (frequencies per 10,000 words in parentheses)**

	Approximation	Search	Example	Focus	Restart
Male (37,108 words)	17 (4.50)	55 (15)	68 (18)	5 (1)	2 (0.50)
Female (30,900 words)	16 (4.30)	40 (13)	41 (13)	6 (2)	2 (1)



**Figure 6.4 Distribution of discourse marker functions of *like* in the SLC according to gender**

My data reveal that 25 out of 27 NSs used the discourse marker *like* at least once, while this was the case for only 41 out of 50 Saudi English language speakers (see Appendix 13). In my data from the SLC, the use of the discourse marker *like* ranged from 1 to 22 tokens. Speaker AR3 used *like* the most with 37 tokens. This speaker did an English language course in Malaysia for one month. Some might argue that this could be the reason why he employed *like* more than all the other speakers. However, among the speakers in the SLC, there were some participants who lived abroad for years but used fewer instances of *like* than those who had never been abroad. A possible explanation for this might be the topic he chose to speak about. Although he was not the only participant who went for topic one,<sup>27</sup> there were many occasions when he needed to give numbers and examples. So, he used *like* as an approximator and for exemplification in 18 out of 37 tokens. In the SLC, there was another participant who had spent some time abroad: speaker AR49. Speaker AR49, who used *like* 17

<sup>27</sup> Topic one is about an experience he/she has had which has taught him/her an important lesson.

times, lived in France for four years. Having been abroad and having spoken French encouraged her to learn English and speak it fluently.

Similar to the SLC, in LOCNEC, the number of times the discourse marker *like* was used ranged from 1 token used by three speakers to 35 as the highest frequency used by one speaker. These quantitative results given by the LOCNEC data make it easier to hypothesize why different interviewees in the SLC used *like* differently. The data clearly indicate that even NSs of English varied in terms of how many times they used *like*, from as many as 35 to one speaker who did not use *like* at all.

### **6.5 Comparison of findings in the corpus and textbooks**

To validate my second hypothesis, I examined the seven textbooks studied by the participants in my research at school. Despite the fact that *like* was the most frequently occurring discourse marker in my data, unexpectedly the result of a search in the textbooks was rather disappointing. Similar to what Müller (2005) found in the textbooks she examined in her study, *like* as a discourse marker does not occur at all. So, it could easily be concluded that *like* is not a discourse marker that Saudi English language learners learn in classrooms.

This begs the question, then, of how it was possible that they used *like* effectively to a level that came close to native speaker use. Like Hellermann and Vergun (2007, p. 167), I would assume that, even though *like* is neither introduced in Saudi English language textbooks nor taught, Saudi English language speakers are probably able to acquire it through their exposure to the media and through their interaction in English with their peers.

## 6.6 Summary and conclusion

In Chapter 6, I have explained and illustrated the discourse marker functions for which *like* was used by the participants in my SLC data and in the native speaker LOCNEC data. In my data, *like* frequently indicates that the speaker is searching for the appropriate expression to represent what s/he has in mind. When *like* precedes a number or a quantitative expression, it marks this expression as an approximator. Another acknowledged function of *like* can be paraphrased as ‘for example’. The concept following *like* represents only an exemplification of what the speaker is thinking about. As a focuser, *like* highlights the following word or expression for a number of possible reasons. The fifth and the last function of *like* in this study is *like* preceding a restart. According to Brinton’s (1996) classification of functions (cf. 2.3), the functions of the discourse marker *like* in this study are all at the textual level. No function was found on the interactional level. This result is similar to that of Müller (2005).

## **CHAPTER 7: CONCLUSIONS**

### **7.1 Introduction**

This concluding chapter aims to give a summary of the major findings of this study. It consists of four major parts. Section 7.2 gives a summary of all the previous chapters. In section 7.3, the research questions of the study are revisited. Section 7.4 discusses the limitations of the study. Section 7.5 is where recommendations for further research are introduced.

### **7.2 Summary of the study**

At the beginning of Chapter 1, the gist of the thesis was provided. The rationale for choosing to investigate discourse makers in the spoken language produced by advanced Saudi English language learners rather than in written language was then given. To make the study possible, it was necessary to create my own corpus, which then becomes the Arabic component of the Louvain International Database of Spoken English Interlanguage (LINDSEI) corpus of spoken learner texts. This led to the purpose of this thesis which was to investigate the use of discourse markers in the spoken English of Saudi learners as well as the representation of these discourse markers in the textbooks to compare the findings derived from authentic use of English provided by the corpora and what is in those textbooks.

Chapter 2 discussed the previous studies on discourse markers. It began by addressing some controversial issues concerning the terminology of discourse markers and the labels given to them by different researchers. This was followed by the characteristics of discourse markers and a review of the functions they fulfil in spoken discourse as discussed in the literature. The characteristics which underpin this thesis were established. The major research approaches

adopted to study discourse markers in English were discussed. Because the aim of my thesis is to investigate how NNSs use discourse markers, it was significant to review some of the studies which looked at how NNSs used English discourse markers. There were some studies which looked at particular Arabic discourse markers used by Arabs while others explored Arabic discourse markers and compared them to English ones used by Arab participants. Although the literature review shows diverse and fruitful results in the study of the discourse markers, there are still gaps which remain to be filled. So, the present study, which examines English discourse markers used by Saudi participants and compared their use to the NSs', is an attempt to fill in the gap in the literature and contribute to the better understanding of discourse markers.

Chapter 3 provides a through description of the database and outlines the research process and analytical tools used in the present study. The LINDSEI database was chosen for its intercultural nature which allows comparison between different corpora. It also gives access to LOCNEC which is currently restricted to the project collaborators. There was also an introduction of why and how the Saudi textbooks are going to be analysed.

The three subsequent chapters investigate the use of *so*, *you know* and *like* which occurred most frequently in my data. The three discourse markers were examined in the three databases: the SLC, LOCNEC, and the textbook database. Chapter 4 presented a detailed examination of the use of the most frequent discourse markers in this study: *so*. The examination of *so* reveals that it is mainly used as a result marker, although it also achieves four other functions in the textual and interactional domains. Similar to *like*, NSs used almost all functions of *so* more than speakers from the SLC. Regarding gender, the analysis indicated that both genders used the functions almost equally with no clear difference



between them. In contrast to *like*, there were a few examples in which *so* occurred as a discourse marker in the textbooks.

In Chapter 5, the discourse marker, *you know*, was examined. The analysis of the data shows that six discourse marker functions were found in both corpora. However, as a discourse marker, the SLC data demonstrates that the frequency of use was higher than in LOCNEC. In addition, the most frequent function was *you know* for speaker/hearer shared knowledge, which was also used by NNSs. In terms of gender, similar to Macaulay (2002), females in my data used *you know* more frequently than males. *You know* as a discourse marker was not introduced at all in the textbooks.

In Chapter 6, the last discourse marker *like* was analysed. The analysis shows that *like* has five major functions in the corpus data. All these functions were found at the textual level. As a discourse marker use of *like*, the results showed that NSs used it more frequently than Saudi speakers. As far as the gender of speakers is concerned, *like* is found to be used more by males than females. This applies to almost all functions of *like*. No single discourse marker use of *like* was detected in any of the textbooks which Saudi speakers studied at school.

### **7.3 Research questions revisited**

This section revisits the research questions of this thesis and suggests how this research contributes to the study of various aspects of discourse markers.

1. To what extent, if at all, do advanced Saudi learners of English (as represented by third and fourth year undergraduate students) use discourse markers in their spoken English language?

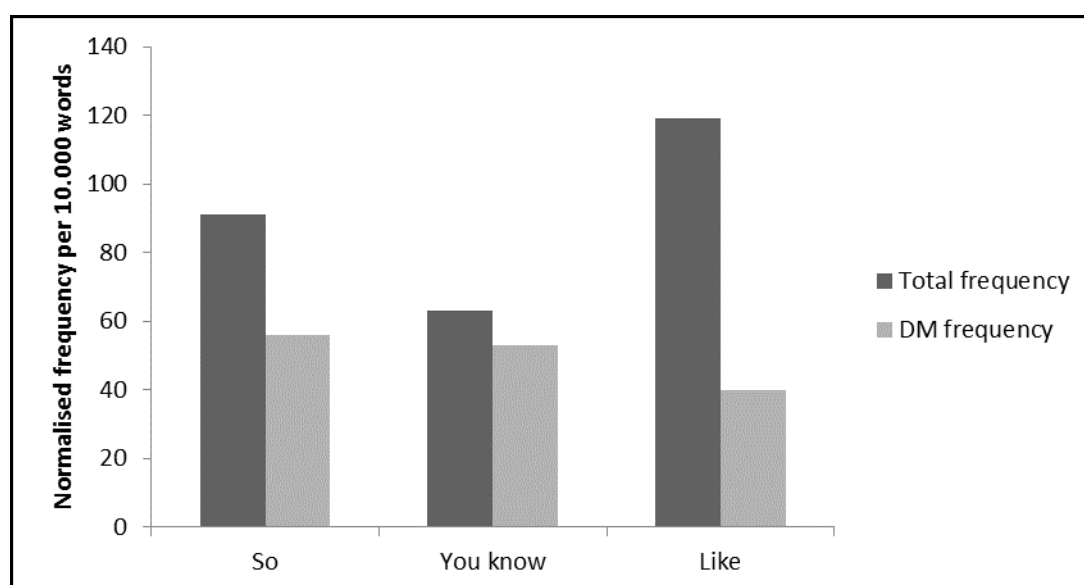
Because this is the first study to investigate the use of English discourse markers by Saudi learners of English, this question was particularly important. Before the data for this study were collected, it was difficult to say whether advanced Saudi learners use discourse markers at all in their spoken English. Thus, it was essential to find an answer to this question. The data were analysed twice. The first analysis was diagnostic and involved 20% of the data (see Table 3.5). This analysis was carried out to verify that Saudi learners do use English discourse markers. The second analysis was carried out after all 50 recordings were collected to determine the richness of the data in terms of the inclusion of discourse markers. The results of both analyses demonstrated that the advanced Saudi learners of English use a different range of discourse markers in their spoken English with different frequencies.

2. Which discourse markers are used most frequently in the discourse of advanced Saudi learners of English?

After I had collected the 50 recordings for the SLC, I ran an analysis to examine which discourse markers were most frequently used. The classification of what constituted a discourse marker and what did not was discussed in Chapter 3. This analysis was similar to that undertaken with the first 10 recordings obtained during data collection (see Table 3.6). The analysis of the 50 recordings showed that *like*, *so* and *you know* were the most frequent discourse markers in the SLC (see Table 7.1 and Figure 7.1).

**Table 7.1 Most frequent discourse markers in the SLC (frequencies per 10,000 words in parentheses)**

	Total frequency	DM frequency
<i>So</i>	622 (91)	380 (56)
<i>You know</i>	428 (63)	358 (53)
<i>Like</i>	808 (119)	273 (40)
Total	1858 (273)	1011 (148)



**Figure 7.1 Frequency of *like*, *so* and *you know* in the SLC**

Because they appeared most frequently and were previously covered in the literature on discourse markers, I decided to investigate them to see what functions they fulfilled in the speech of Saudi learners of English.

As my data show, *so* and *you know* occurred more frequently than *like* in the speech of Saudi learners. Why should this be the case? First of all, the distribution of *so* in the textbooks suggests that students might be more used to employing *so* than *you know* and *like*. The second factor is the interference of L1 with L2. As I discussed in Chapter 4, the discourse marker *so* has an equivalent, *fa*, in Arabic (Kurdi, 2008; Saeed & Fareh, 2006). In addition, in Chapter 5, *you know* was found to have a similar marker in Arabic, *garaft* (Gaddafi, 1999). Bhela (1999) points out that '[W]hen [...] speaking the target language (L2), second language learners tend to rely on their native language (L1) structures to produce a response' (p. 22). I argue that this interference is one of the factors influencing Saudi learners to use these two discourse markers with high frequency even though they have not been exposed to them in the textbooks studied at schools. Trillo (2001, as cited in Müller, 2005) examined the use of discourse markers by Spanish EFL speakers and compared these to their equivalents in a native Spanish corpus. He concluded that the frequency of use of the English discourse markers was related to their equivalents in Spanish. Kurdi (2008) stated that the explanation for the use of the discourse marker *so* as a marker of transition by the participants in her study was due to the influence of L1. Thus, I also argue that *so* and *you know* occurred more frequently than *like* because the former have more or less direct equivalents in Arabic whereas the latter does not.

3. What discourse functions do the selected discourse markers fulfil in the discourse of advanced Saudi learners of English?

The quantitative analysis of the SLC data showed that Saudi learners of English used different functions for each discourse marker. These functions were, in the cases of *so* and *you know*, on the textual and on the interpersonal level. Brinton (1996) defines the textual

level of discourse as that level where ‘the speaker structures meaning as text, creating cohesive passages of discourse’ (p. 38), while the interpersonal level is where the speaker expresses his/her attitudes, evaluations, judgments, expectations and demands (Brinton, 1996). Table 7.2 lists all the discourse marker functions according to their levels: textual and interpersonal (cf. Brinton, 1996). Thus, it can be deduced that Saudi learners of English not only use English discourse markers, but they also demonstrate various functional uses of them. For the three discourse markers examined in this study, there were 17 different functions, 13 of them functioning at the textual level and four functioning at the interpersonal level.

**Table 7.2 All discourse marker functions in the SLC according to their levels**

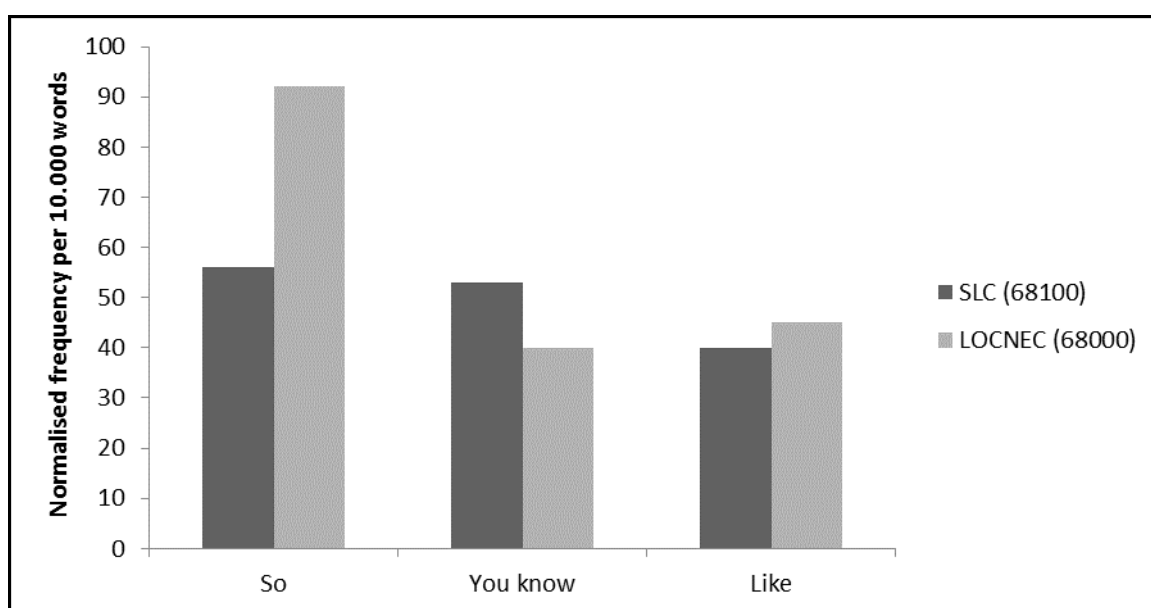
	Textual	Interpersonal
<i>So</i>	<ul style="list-style-type: none"> <li>– Indicating a result</li> <li>– Marking main idea unit</li> <li>– Introducing a new sequence</li> <li>– Introducing a turn</li> </ul>	<ul style="list-style-type: none"> <li>– Marking transition</li> </ul>
<i>You know</i>	<ul style="list-style-type: none"> <li>– Marking lexical or content search</li> <li>– Marking restart and repair</li> <li>– Introducing explanation</li> </ul>	<ul style="list-style-type: none"> <li>– Marking speaker/hearer shared knowledge</li> <li>– Marking generally shared knowledge</li> <li>– Narrative</li> </ul>
<i>Like</i>	<ul style="list-style-type: none"> <li>– Searching for the appropriate expression</li> <li>– Marking an approximate number</li> <li>– Introducing an example</li> <li>– Marking a focus on information</li> <li>– Preceding a restart</li> </ul>	

4. To what extent are Saudi students similar to or different from – both quantitatively and qualitatively – native speakers of English in their use of the selected discourse markers?

To be able to answer the quantitative side of this question, let us take a look at Table 7.3 and Figure 7.2.

**Table 7.3 Number of tokens of the three discourse markers in the SCL and LOCNEC (frequencies per 10,000 words in parentheses)**

	<i>So</i>	<i>You know</i>	<i>Like</i>
SLC (68100 words)	380 (56)	358 (53)	273(40)
LOCNEC (68000 words)	642 (92)	273 (40)	307 (45)



**Figure 7.2 Distribution of the three discourse markers between the SLC and LOCNEC**

As is apparent from Table 7.3 and Figure 7.2, the discourse marker *so* is the most frequently occurring discourse marker in both corpora. *You know* was used more by Saudi learners than the NSs. The difference between the two corpora was that *so* and *you know* achieved high

statistical significance at  $p < 0.0001$ . For *you know*, I can argue that there are two reasons why this was the case. The first reason has been acknowledged in the literature, namely some studies have found that NNSs use *you know* more frequently than NSs. This was introduced in Chapter 5 where Fuller (2003a) argued that NNSs find this discourse marker easy to acquire and in some cases adopt it in a formulaic manner. The second reason, in addition to the first, relates to interference from L1 as *you know* has an equivalent marker in Arabic, *garaft* (Gaddafi, 1999). The last discourse marker, *like*, was used by NSs but the difference was not statistically significant.

The qualitative analysis of the three discourse markers showed that all the functions (see Table 7.2) were used in both corpora. However, when we say that Saudi learners of English used all the functions as did the English NSs, this does not imply equality in the frequency. This means that there is at least one speaker who did not use all or even most of them. For instance, there were nine speakers in the SLC and two speakers in LOCNEC who never used a single discourse marker such as *like*.

5. What are the variables, also called ‘linguistic/non-linguistic’ factors by Müller (2005), which could influence the Saudi use of the selected discourse markers (e.g. gender, speaking other languages, a period of stay in an English-speaking country)?

In the literature on discourse markers, several variables have been identified as having an impact on the frequency of use of markers. These variables include social class, age, gender, etc. (cf. Erman, 2001; Fuller, 2003a; Holmes, 1986; Jucker & Smith, 1998; Müller, 2005; Östman, 1981; Stubbe & Holmes, 1995). Some of these variables were not considered in my thesis. There is no information concerning social class in the LINDSEI database and therefore

neither is there in the SLC, which is a sub-corpus within LINDSEI. As for age, because my data comprises one age group (20–25 years), the age variable was not considered. Thus, three variables were included in this thesis: gender, speaking other languages and a period spent in an English-speaking country. However, the results of the analysis of the three discourse markers revealed that neither speaking other languages nor staying in an English-speaking country had an impact on the use of the discourse markers or the functions for which they were employed.

In relation to gender, at the usage level it was found that gender had no significant influence on the use of the discourse marker *so*. Unlike *so*, *you know* was used more by females than males (cf. Macaulay, 2002) even though this high rate did not achieve statistical significance. However, on the individual function level, there were two functions which were used by males more than females: *you know* for lexical or content search and *you know* in narratives, but again no statistical significance was achieved for either of them.

6. To what extent are the selected discourse markers represented in Saudi textbooks? How might this be connected to the actual output of the speakers in the study?

As discussed in the literature review (Chapter 2), discourse markers are not taught at school (Hellermann & Vergun, 2007; Sankoff et al, 1997; Trillo, 2002). To the best of my knowledge, there is no existing study on the usage of discourse markers in the Saudi classrooms. This led me to examine the representation of discourse markers in Saudi textbooks. I examined seven textbooks: sixth year elementary, first year intermediate, second year intermediate, third year intermediate, first year secondary, second year secondary and third year secondary. These textbooks were taught at the time at which the Saudi participants



in the SLC corpus were learning English at school. The examination of the textbooks showed that the discourse marker *so* was introduced in the textbooks. There were 75 instances of *so* representing different functions. However, *you know* and *like* did not occur at all.

In my opinion, given the representation of the three discourse markers in the Saudi textbooks of English, it is not surprising that the Saudi speakers in the SLC did not have much difficulty in using *so*. In fact, *so* was the most frequent marker in the SLC data. Saudi speakers were still able to employ *you know* and *like* even though they were not taught them at school. Nevertheless, *you know* occurred more frequently than *like* because it has an equivalent in Arabic but *like* does not.

#### **7.4 Limitations of the study**

This is one of the first studies, if not the only study, to compare systematically the use of discourse markers in two different corpora and in a textbook database. Given the scope of this thesis, it was only possible to analyse three discourse markers in detail. Of course, the inclusion of more discourse markers would provide a more comprehensive account of similarities and differences.

There are a number of limitations related to the nature of LINDSEI. The first limitation has to do with the register as the data relate to interviews and not natural conversations. I agree with Müller (2005) that even though various conditions and topics can be controlled in interviews, yet interviews as an approach will still produce recorded material that does not necessarily represent natural speech data.

The second limitation is regarding the size of the corpus, which comprises no more than 50 recordings. So, it could be argued that LINDSEI is relatively small in size when compared to other corpora such as the BNC, COCA and COHA, which run into hundreds of millions of words. However, representativeness is not all about corpus size. Biber (1993) states that:

‘[...] sample size is not the most important consideration in selecting a representative sample; rather, a thorough definition of the target population and decisions concerning the method of sampling are prior considerations. Representativeness refers to the extent to which a sample includes the full range of variability in a population.’ (p. 243)

The third limitation is the limited age group of the participants, who were third and fourth year university students, aged 20–25 years. This does not allow comparisons with other age groups which could illustrate other varieties of discourse marker use.

Because the interviews steps are predetermined by LINDSEI, the researcher has no control in terms of adding or skipping a step should the circumstances of the interview require (i.e. very long or very short interviews). Indeed, if an interview does not follow the structure of the interview schedule precisely, no matter how good it is, it should be discarded. The last point I would like to make here concerns prosody. LINDSEI does not provide prosodic information in the orthographic transcription. The presence of prosody could help in identifying some particular functions. For example, because LOCNEC does not come with its recordings, it is harder to establish the function of a discourse marker such as *like* marking focused information.

Another limitation of this thesis is the functional analysis of the discourse markers. Even though discourse markers are multifunctional, this study has opted to take, in any given

context, only the most likely function of a discourse marker and has not attempted to capture all its possible functions.

The last limitation in this section is related to the analytical framework followed in this study; coherence-based framework. I argue with Schourup (1999, 240) that in the coherence-based account, discourse markers are believed to play a role in the identification of coherence relation between two textual units whereas relevance account gives no attention to coherence relations. Hearers in the relevance approach, however, are responsible to determine how an utterance achieves relevance rather than attempting to identify these relations (Blakemore, 1996: 328). Therefore, adopting relevance approach can primarily yield cognitive-related functions of discourse markers which could probably contribute to the analysis of the unclear instances where there were no enough contextual cues.

### **7.5 Recommendations for further research**

Although this thesis did achieve its main aim, investigating discourse markers used by Saudi learners of English, a number of recommendations for the future can be made which could contribute to the same aim and provide a broader picture of how discourse markers are used by Saudi learners. It is hoped that the findings of this thesis will trigger more research on the study of discourse markers, i.e. including more discourse markers, studying them within different frameworks such as relevance theory (cf. Anderson 1998, for instance) and studying the position of discourse markers within utterances.

Another recommendation is to study the classroom discourse to examine teachers' language, focusing on the use of discourse markers. In addition, an examination of Saudi university textbooks to look at the representation of discourse markers is recommended. This would

provide more answers to the question of how Saudi learners of English come to use discourse markers in their speech.

The last recommendation is to assess in detail the current textbooks at schools and to provide recommendations for activities which could be incorporated in the teaching materials to introduce discourse markers and their usage to students. Following Lam (2008), in Table 7.4 below is an example of what such activities might look like. Similar consciousness-raising activities are believed to encourage language learners ‘to notice particular features of the language, to draw conclusions from what they notice and to organize the view of language in the light of the conclusions they have drawn’ (Willis & Willis, 1996, p. 64). Willis and Willis (1996) point out that the benefit of these types of activities is to assist learners to notice language features which may be overlooked in real-time communication.

**Table 7.4 Example of a suggested activity to introduce discourse markers in textbooks**

- Activity 1: Distinguishing discourse functions from grammatical functions
- Study the following concordance lines which are examples of *like* serving discourse functions and grammatical functions:
1. so the change *like* really smart from that point of view but when I get to the university
  2. dish washing and we are like professional it takes 1= just *like* ten minutes
  3. because she is exactly *like* her and since I thought she she she is not pleased
  4. tell her that this paint doesn't look *like* you and you are more beautiful than her
  5. the thing is I know I know the guy for *like* three years
  6. I dunno don't remember the sing (ur) saying *like* <foreign> inta 3reeb ya3ni </foreign>
  7. I *like* cooking and actually we don't have a servant
  8. especially with baking I *like* my cookies and muffin
  9. I was *like* wow they had so much fun
  10. yeah he was he was he wasn't *like* (erm) outgoing enough

Now discuss the differences between the discourse and grammatical functions of *like*.

## 7.6 Concluding remarks

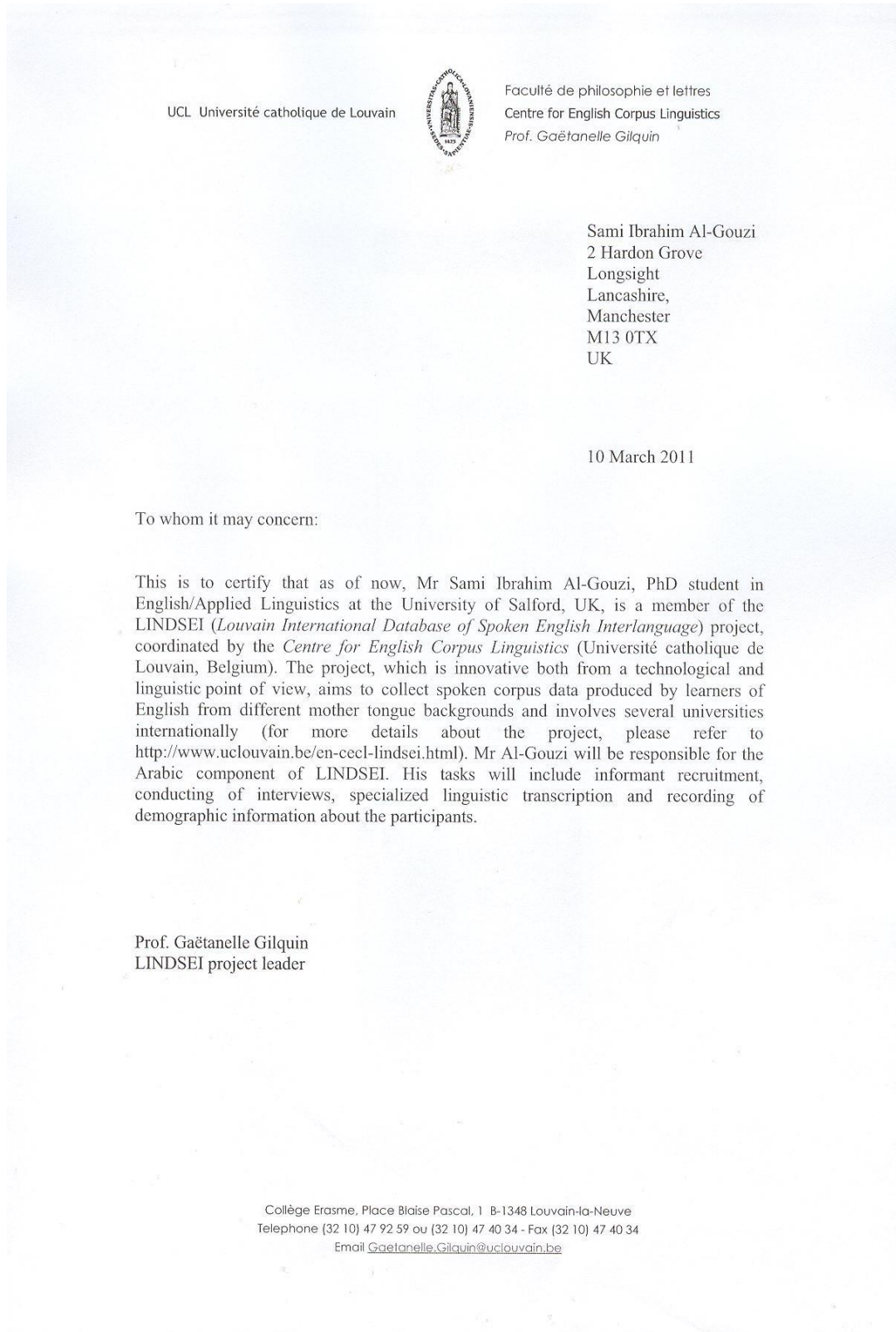
In this, the last part of this thesis, it is worth making some final concluding remarks about the potential contributions it can make to different disciplines. Although the corpus-based investigation of spoken English is not recent, this research is a very promising contribution to the field as it provides, for the first time, a **spoken** learner corpus of speakers of English whose first language is Arabic.<sup>28</sup> This should open doors for linguists around the world to investigate this corpus on its own or in conjunction with other corpora from LINDSEI. This corpus also allows linguists to investigate various aspects of learner English, such as lexis, syntax, phraseology, discourse or pragmatics. Another contribution to the field of discourse analysis is that the study sets out both qualitative and quantitative findings related to the use of English discourse markers by NNSs/language learners. This forms a starting point from which more research on discourse markers can be undertaken, relating or comparing the findings to those from this study.

---

<sup>28</sup> There is a written learner corpus of speakers of English whose first language is Arabic available at: <http://www.uclouvain.be/en-cecl-icle.html>

## APPENDIX 1

### Membership of LINDSEI letter



## APPENDIX 2

### Approval letter from the Department of English from University of Umm Alqura

Kingdom of Saudi Arabia  
Ministry Of Higher Education  
Umm Al-Qura University



المملكة العربية السعودية  
وزارة التعليم العالي  
جامعة أم القرى

15/3/2011

Dear Mr. Sami Al-Gouzi,

I have received your request to conduct your linguistics experiment on some students in our department this semester. I would like to inform you that we approve your experiment and appreciate such work. However, we need to have a sample of your test and arrange for the time and procedures with the chair deputy at the girls section.

Please feel free to contact me for any further information.

Truly yours,  
  
Dr. Abdulhafeez Turkistani

Chair,

Department of English

Umm Al-Qura University

Mahhah, Saudi Arabia

Email: english@uqu.edu.sa .



الرقم: التاريخ: المشفوعات: مطابع جامعة أم القرى

### APPENDIX 3

<Pilot recording= "P01">

<A> hi <first name of interviewee> </A>

<B> ok </B>

<A> thank you very much for accepting to taking part with me . in my pilot study . I gave you the three topics to choose one of them . and I think you've already chosen one . have n't you </A>

<S>

<B> yeah (er) actually I . chose the third topic (er) .. I cannot say th= wheth= whether it was good or bad but i . when I've been a child . (er) . I saw (er) a film . it a horror movie like (er) a father who is scientist or something . he has a daughter . only daughter . and he gave her (er) like a small crocodile to bring it up </B>

<A> (uhu)</A>

<B> and after a while (er) the girl felt bored from that . crocodile . and she left it in the toilet . and she flashed it . (er) mean= meanwhile (er) th= her father has (er) a kind of medicine . which moves everything to become a giant </B>

<A> (uhu)</A>

<B> and he has extra dose of it or he has so many . so much amount of it . so he throw it in the toilet . did not know that the </B>

<A> the croc= the crocodile </A>

<B> yeah the daughter has already throw it (er) and after that . crocodile becomes so huge and started attacking people and <XX> and people swimming at the sea and he(er) it attacked (er) them (er) .. the thing I liked the the scenes maybe the (er) the story itself the the[i:] . the actions scenes I donno maybe the way of acting y: you feel as if it were . true </B>

<A> (uhu) </A>

<B> (er) but I the thing I wouldn't like like (er) th: the scene of blood </B>

<A> (uhu) </A>

<B> and it's really was frightening for a child of about five six year old </B>

<A> (uhu) I agree</A>

<B> but when I grew up I looked at it from another by coincidence I've it seen again </B>

<A> (uhu) </A>

<B> and I looked at it I look at the (er) (er) creativity of you know creating such movies I know it's false now . and I feel it's boring </B>

<A> yeah</A>

<B> but it (urm) it was coincidence that I've seen twice once I've been so young and once when I've been so old </B>

<A> were they different from each other the first and the second </A>

<B> no the same version </B>

<A> yeah I mean I mean you were looking at it from different prospective maybe or </A>

<B> yeah</B>

<B> yep when I see the blood I thought t it's true. and I remembered the scene there was a baby on the. beach you know . and everybody was running away . and this (er) this baby was emotionless like he is so and the mom . was running to catch . to save . the child and people hold her arms because they don't want her to lose her life and she was thinking to rescue and in the last minute . the baby was you know [?] it it full [X] it's of suspense you know suspense and it's very exciting but .. when I grew up I . I I look at it as I enjoyed it I don't feel pity for people so much because < overlap/>I know it's false </B>



</S>

<F>

<A> yeah ok so if I ask you to tell me about a lesson you= you've taught from that movie . especially when you're a grown up and (er) I mean now can you can you just give a lesson about it .. can you read it again</A>

<B> ah it's good question (ur) (ur) the moral lesson I thought I I've learnt from it when I . was child that (ur) the father is stupid he should raise th= that child awareness that ok girl don't throw anything . you know when because he spoiled so much you know and the girl just because felt bored of this she throw it away . the next thing is that laboratory if you have a child at home and you are doing you have lab= laboratory should be more careful about not throwing (ur) these you know dangerous thing which can go (ur) throw the sewage system</B>

<A> yeah hazardous materials </A>

<B> yeah yeah it it was catastrophe cresols because th= there was wedding . and people suddenly the the crocodile (er) bite (ur) the (ur) or bit the . leg of the bride and the white dress becomes red it was horrible and the cake I was worried about the cake <starts laughing> that <stops laughing> pull over anything is upside down . but now if I want to look at it I . I'm against this films . because really it affects me badly its (ur) it sti= it stuck in my memory . because like (er) .. it plays on my nerves I donno I . I become so excited so (er) watching out what's going to: happen next . but at the same time it's . I know it it this film costs maybe (ur) two kilos of tomatoes but like when I was child I thought oh blood blood you know I . I become scared if I see a cat or any small: insect on the floor oh its beco= it will turn to become a giant and attack me . there was no= not moral lesson from it . it was made just me for suspense and maybe for business </B>

<A> well at least you don't recommend kids to watch . some kind of horrible </A>

<B> yeah but still it affects teenagers </B>

<A> yeah </A>

<B> yeah teenagers are very much . enthusiastic to: .. to follow such</B>

<A> have ever seen any similar movie . to this one</A>

<B> (ur) I've seen more horror movies but this one why it was horrible . because it was not exaggeration like anaconda (?) you know or like . I don't remember the name of the film the fly which becomes giant I donno know . but this one was . you can believe it because it's scientific scientifically speaking it can occur . I mean like if a material can enlarge a creature . maybe but </B>

<A> but do do you believe that there is a material or liq= a liquid or an= anything </A>

<B> no because if there was a ghost I don't believe in ghost if there was a ghost or: there was something supernatural . or dinosaur or like something like that or like . supernatural creature you don't believe it . but this it has a touch of: truth or touch of reality . maybe . of course there is not to this degree I mean but: like . you know like everything now is artificial the vegetables animals the hen the: the hen grow quickly <starts laughing> </B>

<A> yeah </A>

<B> and it takes less you know <stops laughing> </B>

<A> yeah </A>

<B> so maybe </B>

<A> yeah they can play with genes </A>

<B> yeah they can play with it so that's why I it stuck in my memory because you can believe it . but if it were like plastic or better (?) creature or something it would be silly for me I don't want even watch it </B>

<A> so do you normally fancy the horror movies </A>

<B> not really </B>

<A> so what's you what's your best type of movies you always prefer to watch </A>

<B> (er)( urm) maybe soap opera soap opera because (er) romantic movies are . I feel th= they are not down to earth you know they don't discuss daily life issues you know I like drama social drama and: like soap opera </B>

<A> are you are you watching </A>

<B> like friends . friends this series and American series I think . yeah </B>

<A> how about any English series you watch . do you watch </A>

<B> Big brother is nice . my favourite is Britain's got talent .. and fortunately now there is Arab's got talent </B>

<A> yeah </A>

<B> yeah it's amusing you see people not only singers dancers (er) (er) ma= magicians maybe </B>

<A> Everything yeah .they can just reflects talent </A>

<B> yea . yeah performance </B>

</F>

<P>

<A> Yeah that's great (urm) ok can we just move into the[i:] last . task as I told you there will be four pictures to that can tell a story . I ask you to look at them and try to make a story around them can you just tell me anything you think about it </A>

<B> (eh) I have two answers for it if you allow me </B>

<A> ok yeah of course yeah </A>

<B> (er) When I first look at it I remember a story maybe it's right or wrong . they say we s: see it in cartoon . films . like (er) when .. nobody knows the mirror be= before as any mentioned . so when (er) a man from china I donno maybe the first in life . in the history maybe china was the first I'm not so sure . but in the east I remember a man bo= bought mirror and he went home . and then his wife was unpacking and suddenly she found . she looked at the mirror and she become so jealous . she said who is that woman <starts laughing> because they don't know there invention called mirror <stops laughing>. and then later on he was shocked . and she looked at him as (er) you know you betrayed or something she was so mad with him . but (er) later on it's it . everybody looks at it he will see himself or herself . but now when I . l= looked again . (er) I have a second thought about it and it's like. the woman or a lady she is so much beautiful . and the (er) (er) artist draw drew her and she said no it's she was not satisfied or unhappy with it </B>

<A> she was socked maybe because she's seen her </A>

<B> And she said no (er). no it's not me I don't like it and then he re (er) drew it again but . it's not like her just he wants to draw a beautiful woman just maybe sat= satisfy her as a customer maybe it's a start of the customer service just to do what people like </B>

<A> yeah </A>

<A> yeah anything just to please them </A>

<B> and she was happy and proud of it and: people were looking at it . and I think people are . they don't believe that she is he herself or </B>

<A> so what do you think . the people in the fourth picture or her friends when they were watching her her p= portrait </A>

<B> making a social complement they are flattering . I think they are from th= the deep their minds they deepest of their mind they they they she know is (er) not her the picture is not hers so h= they just (ur) I think they are doing like what he did </B>

<A> yeah </A>

<B> just watching because maybe she is: rich </B>

<A> (uhu) and they are </A>

<B> or they want to flatter her </B>

<A> complement her because </A>  
 <B> social complement or: </B>  
 <A> so i= if I </A>  
 <B> because it= it's totally not her . but because she however he because she is: superficial maybe her superficial thinking because she doesn't look her (ur) picture . should be realistic and she she wants any picture which is beautiful even if it was not her </B>  
 <A> (uhu) </A>  
 <B> and (ur) this is the only solution mayb he copes with he= her he is peaceful man . we donno know if if he is her husband it would be nice of him it would be like he is tactful I can say but if he is (er) poor man he wants money after all it's customer service (ur) but I don't blame him I blo= blame hi= them </B>  
 <A> with her requests </A>  
 <A> them </A>  
 <B> yeah her: friends </B>  
 <A> but if I can just draw and start from the (u)r the point you've started to and carry on your (um) (u)r reading to the pictures . I can see that the lady here . are not (urm) realistic because she doesn't want to appear . and to see her see her in reality . and people are really making laugh a= </A>  
 <B> she doesn't have self-content </B>  
 </A> yeah </A>  
 <B> she doesn't have self-confidence </B>  
 </A> her self-confidence is very low if the [i:] if the reading of the picture is like what you have mentioned . but if you were in her shoes for example what are you going to do </A>  
 <B> yeah (ur) at the beginning she asked somebody to draw hers . so she: . she should accept the photo because it's her after all because it's not (ur) dress she can . (ur) you know or anything she can manipulate or dress she can cut and do and . it's (ur) it's just her photo </B>  
 <A> yeah </A>  
 <B> so the artist . (ur) we cannot blame the artist . so he is just drawing what he </B>  
 <A> so the blameworthy here is </A>  
 <B> from his artistic point of you so </B>  
 <A> so the lady here is th= is the to blame I think </A>  
 <B> yeah </B>  
 <A> and the painter has nothing to do </A>  
 <B> actually I don't like such quality of people . but . they are there. in our life . daily life so we have to to accept them as they are </B>  
 <A> so what the category of those people .. you said that they are there in our life </A>  
 <B> they are living in their own l= maybe they .. (ur) they do want to . accept life as it is because you have to adapt yourself to any problem </B>  
 <A> to to the reality </A>  
 <B> for those people yeah if they made accident ad they lose part of their body they become mel= melancholic and maybe they commit suicide . because they cannot accept any change even maybe . maybe she was: very beautiful when she she's been young younger but now she becomes old and she doesn't accept reality she doesn't she doesn't know that every season has it beauty even autumn sometimes nice maybe she . she is not realistic she is </B>  
 <A> (uhu) </A>  
 <A> brilliant </A>  
 <B> fantastic or like donkey shot called yeah living in the out of date <starts laughing> fantasia <stops laughing> </B>  
 <A> thank you very much to be honest this is the best reading I've . I've ever received  
 <B> please don't be like the artist </B>

<A> oh <starts laughing>no no of course <starts laughing> </A>

<B> draw me as I'm </B>

<A> yeah thank you very much for the reading and for the interpretation of these four pictures . and thank you again for taking part with me and we just come to the end of the interview thank you <first name of interviewee> </A>

<B> but may I add something </B>

<A> yeah of course </A>

<B> you are (urm) it's very nice idea like if you look at it a story consist of like consist of like one two three four person people</B>

<A> (uhu) (uhu) </A>

<B> and that's excellent you know </B>

<A> yeah </A>

<B> it can . it's like a piece of cake it's like (er) one episode of reality everyday life </B>

<A> (uhu) (uhu) </A>

<A> thank you very much and I wish you a very best of luck cheers </A>

</P>

</h>

## APPENDIX 4

Letter sent to University of Umm Al-Qura



Sami Al-Gouzi (PhD Student)  
School of English, Sociology, Politics &  
Contemporary History  
University of Salford  
Crescent House, The Crescent  
Salford M5 4WT  
United Kingdom  
Tel: +44 (0) 161 2485084  
Mobile: +4477 8815 6133  
Email. s.i.al-gouzi@edu.salford.ac.uk

**(University agreement letter)**

**Date 25/02/2011**

**Dear Dr. Turkistani,**

I am a postgraduate research student in Applied Linguistics at the University of Salford. As part of my course I am undertaking a research study which examines various aspects of Saudi learners' English language usage.

Prior to undertaking the study I would like to request your approval to approach male and female students within your department to take part in the study.

Participants will be contacted and interviewed by members of staff with whom I have already informally discussed the project and what is required from them to do. I hope to recruit 50 participants (25 males and 25 females). Their involvement entails one recording. In line with

the LINDSEI (Louvain International Database of Spoken English Interlanguage) design, participants will be asked to speak about a familiar topic that will encourage participants to speak in a relatively unmonitored style. For more details, a participation information sheet and a sample consent form are attached.

I can assure you that the study will not disrupt the working environment or teaching schedule and provision in any way, and that all necessary steps will be taken to ensure students' anonymity and confidentiality. For this purpose, the study has been approved by the Ethics Panel at the University of Salford.

If you have any questions about this project, would like more information, or would like to raise any concerns you might have, please do not hesitate to contact me or my supervisors at the details below.

Dr. Nicholas Smith Lecturer in English Language & Linguistics School of English, Sociology, Politics & Contemporary History University of Salford Crescent House Salford M5 4WT UK. Tel: +44 (0)161 295 6791 Fax: +44 (0)161 295 5077 e-mail: n.smith@salford.ac.uk	Dr Heike Pichler Lecturer in Sociolinguistics School of Languages University of Salford Maxwell Building Salford M5 4WT UK. Tel: +44 (0)161 295 4575 Fax: +44 (0)161 295 5335 e-mail: h.pichler@salford.ac.uk
--	--

I would like to thank you for your co-operation in this matter.

Yours sincerely,

Sami Al-Gouzi

## APPENDIX 5

### Approval letter from the Department of English from University of Umm Alqura

Kingdom of Saudi Arabia  
Ministry Of Higher Education  
Umm Al-Qura University



المملكة العربية السعودية  
وزارة التعليم العالي  
جامعة أم القرى

15/3/2011

Dear Mr. Sami Al-Gouzi,

I have received your request to conduct your linguistics experiment on some students in our department this semester. I would like to inform you that we approve your experiment and appreciate such work. However, we need to have a sample of your test and arrange for the time and procedures with the chair deputy at the girls section.

Please feel free to contact me for any further information.

Truly yours,  
  
Dr. Abdulhafeez Turkistani

Chair,

Department of English

Umm Al-Qura University

Mahhah, Saudi Arabia

Email: english@uqu.edu.sa .



## APPENDIX 6



**School of English, Sociology, Politics  
& Contemporary History**  
**University of Salford**  
**Crescent House, The Crescent**  
**Salford M5 4WT**  
**Tel: +44 (0) 161 2485084**  
**Mobile (UK): +4477 8815 6133**  
**Mobile (KSA): +966 505715718**  
**Email. s.i.al-gouzi@edu.salfors.ac.uk**

### Participant Consent Form

**Title of project:** A corpus-based study of language use by Saudi learners of English

**Name of project supervisors:** Dr Nicholas Smith and Dr Heike Pichler

**Name of researcher:** Sami Al-Gouzi, PhD student – Applied Linguistics

#### Tick the appropriate box please:

1. I confirm that I have read and understood the participant information sheet for the above study.

YES	NO
-----	----

2. I have been given the opportunity to have my questions answered by the researcher (via telephone or e-mail) or the interviewer (face-to-face).

YES	NO
-----	----

3. I agree to take part in the project and be recorded while being interviewed by my university lecturer and speaking to a fellow student.

YES	NO
-----	----

4. I accept that I will receive no payment for my participation in this project.

YES	NO
-----	----

5. I understand that my participation is completely voluntary and that I can withdraw from participation at any time without any consequences and without giving any explanation.

YES	NO
-----	----

6. I give my consent to the researcher and his supervisors to access my recordings.

YES	NO
-----	----



7. I understand that my anonymity will be guaranteed should extracts from my recording be published in any form or media.

YES	NO
-----	----

8. I agree that the written transcription of my interview recording and accompanying material will be:

(i) quoted in the researcher's PhD thesis and published work or used in public presentations;

(ii) part of the large data archive of non-native Englishes which is continually up-dated project and made available to legitimate researchers.

YES	NO
-----	----

9. I agree that the recording of my interview will be part of the large data archive of non-native Englishes which is continually up-dated project and released to legitimate researchers.

YES	NO
-----	----

Name of participant giving consent .....

Signature of participant giving consent .....

Date consent was given .....

Name of researcher taking consent .....

Signature of researcher taking consent .....

Date consent was taken .....

**A copy of this consent form will be left with you and a copy will be taken by the researcher.**

## APPENDIX 7

Learner profile questionnaire

### LEARNER PROFILE

=====

Text code: (to be filled in by the researcher)

=====

**Surname: First name(s):**

Age:

Male ☐ Female ☐

Nationality:

Country:

Native language:

Father's mother tongue:

Mother's mother tongue:

Language(s) spoken at home: (if more than one, please give the average % use of each)

**Education:**

Primary school - medium of instruction:

Secondary school - medium of instruction:

Current studies:

Current year of study:

Institution:

Medium of instruction:

English only ☐

Other language(s) (specify) ☐

Both ☐

=====

Years of English at school:

Years of English at university:

**Time spent in an English-speaking country:**

Where?

When?

How long?

=====

**Other foreign languages in decreasing order of proficiency:**

=====

**I hereby give permission for my interview to be used for research purposes.**

Date: ..... Signature: .....

.....

**Section to be filled in by the interviewer**

**Interviewer:** Male ☐ Female ☐

Native language:

Foreign languages (in decreasing order of proficiency):

Relation with learner: Familiar ☐ Vaguely familiar ☐ Unfamiliar ☐

(If possible, please be more specific, e.g. learner's professor, TA, etc:.....)

## APPENDIX 8

### Interview tasks

#### LOUVAIN INTERNATIONAL DATABASE OF SPOKEN ENGLISH INTERLANGUAGE

I'd like to interview you informally on things of interest in your life for fifteen minutes. To get the conversation started could you please choose one of the following topics and think about what you are going to say. You should aim to be able to talk for 3–5 minutes. The conversation will then continue informally.

**Topic 1:** An experience you have had which has taught you an important lesson.  
You should describe the experience and say what you have learnt from it.

**Topic 2:** A country you have visited which has impressed you. Describe your visit and say why you found the country particularly impressive.

**Topic 3:** A film/play you've seen which you thought was particularly good/bad.  
Describe the film/play and say why you thought it was good/bad.

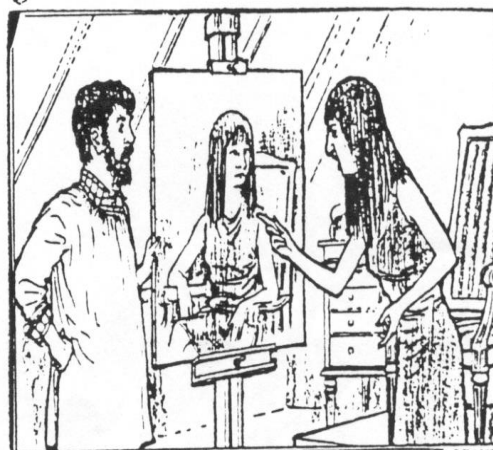
Please don't take any notes as I would like it to be a spontaneous talk.

## APPENDIX 9

### Picture description task

#### Annex 2: Story for retelling

The four pictures below tell a story. Study the pictures and then make up a story around them.



## APPENDIX 10

### **LINDSEI transcription guidelines**

#### **1. Interview identification**

Each interview is preceded by a code of this type: <h nt="FR" nr="FR+*three-figure number*">

e.g. <h nt="FR" nr="FR004"> (4th interview with French mother tongue student)

Examples of country codes:

DUTCH	=	DU001
GERMAN	=	GE001
NORWEGIAN	=	NO001
SPANISH	=	SP001
SWEDISH	=	SW001

All interviews should end with the following tag (on a separate line): </h>

#### **2. Speaker turns**

Speaker turns are displayed in vertical format, i.e. one below the other. Whilst the letter "A" enclosed between angle brackets always signifies the interviewer's turn, the letter "B" between angle brackets indicates the interviewee's (learner's) turn. The end of each turn is indicated by either </A> or </B>.

e.g. <A> okay so which topic have you chosen </A>

<B> the film or play that I thought was particularly good or bad really </B>

#### **3. Overlapping speech**

The tag <overlap /> (with a space between "overlap" and the slash) is used to indicate the beginning of overlapping speech. It should be indicated in both turns. The end of overlapping speech is not indicated.

e.g. <B> yeah I went on a bus to London once and I'll never <overlap /> do it again </B>

<A> <overlap /> that's even worse </A>

#### **4. Punctuation**

No punctuation marks are used to indicate sentence or clause boundaries.

#### **5. Empty pauses**

Empty pauses are defined as a blank on the tape, i.e. no sound, or when someone is just breathing. The following three-tier system is used: one dot for a "short" pause (<1 second), two dots for a "medium" pause (1-3 seconds) and three dots for "long" pauses (> 3 seconds).

e.g. <B> (erm) .. it's a British film there aren't many of those these days </B>

#### **6. Filled pauses and backchannelling**

Filled pauses and backchannelling are marked as (eh) [brief], (er), (em), (erm), (mm), (uhu) and (mhm). No other fillers should be used.

e.g. <B> yeah . well Namur was warmer (er) it was (eh) a really little town </B>

#### **7. Unclear passages**

A three-tier system is used to indicate the length of unclear passages: <X> represents an unclear syllable or sound up to one word, <XX> represents two unclear words, and <XXX> represents more than two words.



e.g. <B> <X> they're just begging <XX> there's there's honestly he did a course .. for a few weeks </B>

If transcribers are not entirely sure of a word or word ending, they should indicate this by having the word directly followed by the symbol <?>.

e.g. <B> I went to see a<?> friend at university there and stayed </B>

Unclear names of towns or titles of films for example may be indicated as <name of city> or <title of film>.

e.g. <B> where else did we go (er) <name of city> it's in Bolivia </B>

## **8. Anonymization**

Data should be anonymised (names of famous people like singers or actors can be kept).

Transcribers can use tags like <first name of interviewee>, <first name and full name of interviewer> or <name of professor> to replace names.

e.g. <A> I'm <first name of interviewer> . what's your name? </A>

## **9. Truncated words**

Truncated words are immediately followed by an equals sign.

e.g. <B> it still resem= resembled the theatre </B>

## **10. Spelling and capitalization**

British spelling conventions should be followed. Capital letters are only kept when required by spelling conventions on certain specific words (proper names, I, Mrs, etc.) – not at the beginning of turns.

## 11. Contracted forms

All standard contracted forms are retained as they are typical features of speech.

## 12. Non-standard forms

Non-standard forms that appear in the dictionary are transcribed orthographically in their dictionary accepted way: *cos*, *dunno*, *gonna*, *gotta*, *kinda*, *wanna* and *yeah*.

## 13. Acronyms

If acronyms are pronounced as sequences of letters, they are transcribed as a series of upper-case letters separated by spaces.

e.g. <B> yes not really I did sort of basic G C S E French and German </B>

If, on the other hand, acronyms are pronounced as words, they are transcribed as a series of upper-case letters not separated by spaces.

e.g. <A> (mhm) (er) you're doing a MAELT </A>

## 14. Dates and numbers

Figures have to be written out in words. This avoids the ambiguity of, for example, "1901", which could be spoken in a number of different ways.

e.g. <B> an awful lot of people complain and say well the grants were two thousand two hundred </B>

## 15. Foreign words and pronunciation

Foreign words are indicated by <foreign> (before the word) and </foreign> (after the word).

e.g. <B> we couldn't go with (er) knives and so on <foreign> enfin </foreign> we were (er)  
</B>

As a rule, foreign pronunciation is not noted, except in the case where the foreign word and the English word are identical. If in this case the word is pronounced as a foreign word, this is also marked using the <foreign> tag.

e.g. <B> I didn't have the (erm) . <foreign> distinction </foreign> </B>

## 16. Phonetic features

### (a) Syllable lengthening

A colon is added at the end of a word to indicate that the last syllable is lengthened. It is typically used with small words like *to*, *so* or *or*. Colons should not be inserted within words.

e.g. <B> that's something I'll I'll plan to: to learn </B>

### (b) Articles

-when pronounced as [ei], the article *a* is transcribed as a[ei];

e.g. <B> and it's about (erm) . life in a[ei] (eh) public school in America I think </B>

-when pronounced as [i:], the article *the* is transcribed as the[i:].

e.g. <B> and the[i:] villa we were staying in was in one of the valleys </B>

## 17. Prosodic information: voice quality

If a particular stretch of text is said laughing or whispering for instance, this is marked by inserting <starts laughing> or <starts whispering> immediately before the specific stretch of speech and <stops laughing> or <stops whispering> at the end of it.

e.g. <B> <starts laughing> I don't have to assess it I only have to write it <stops laughing>  
</B>

## **18. Nonverbal vocal sounds**

Nonverbal vocal sounds are enclosed between angle brackets.

e.g. <B> I hope so I've I've got some <coughs> friends out there </B>

e.g. <B> so I went back into Breda .. and sat down again <imitates the sound of a guitar>  
</B>

## **19. Contextual comments**

Non-linguistic events are indicated between angle brackets only if they are deemed relevant to the interaction (if one of the participants reacts to it, for example).

e.g. <A> no it's true it's nice to have your own bathroom </A>

<somebody enters the room>

<B> hi </B>

## **20. Tasks**

The three tasks making up the interview (set topic, free discussion and picture description) should be separated from each other. This is done using the following tags: <S> (before the set topic), </S> (after the set topic), <F> (before the free discussion), </F> (after the free discussion), <P> (before the picture description), </P> (after the picture description). These tags should occupy a separate line and should not interrupt a turn.

e.g. <S>

<A> did you . manage to choose a topic </A>

### **? Questions?**

If you have any questions regarding these transcription guidelines, don't hesitate to get in touch with us!

## APPENDIX 11

Use of discourse marker *so* per each participant in SLC

Participant's code	Frequency of DM use	Participant's code	Frequency of DM use
AR1	24	AR26	16
AR2	2	AR27	3
AR3	6	AR28	0
AR4	10	AR29	2
AR5	18	AR30	0
AR6	3	AR31	8
AR7	5	AR32	3
AR8	9	AR33	10
AR9	9	AR34	6
AR10	0	AR35	11
AR11	43	AR36	2
AR12	16	AR37	0
AR13	34	AR38	3
AR14	8	AR39	6
AR15	1	AR40	16
AR16	7	AR41	4
AR17	6	AR42	10
AR18	2	AR43	0
AR19	15	AR44	6
AR20	5	AR45	10
AR21	2	AR46	0
AR22	5	AR47	1
AR23	6	AR48	7
AR24	12	AR49	0
AR25	0	AR50	8

## APPENDIX 12

Use of discourse marker *you know* per each participant in SLC

Participant's code	Frequency of DM use	Participant's code	Frequency of DM use
AR1	50	AR26	1
AR2	4	AR27	2
AR3	18	AR28	61
AR4	3	AR29	0
AR5	1	AR30	0
AR6	0	AR31	5
AR7	1	AR32	11
AR8	0	AR33	2
AR9	34	AR34	6
AR10	4	AR35	64
AR11	3	AR36	0
AR12	0	AR37	0
AR13	8	AR38	6
AR14	0	AR39	0
AR15	0	AR40	1
AR16	3	AR41	3
AR17	0	AR42	0
AR18	0	AR43	0
AR19	0	AR44	11
AR20	3	AR45	0
AR21	20	AR46	0
AR22	0	AR47	0
AR23	4	AR48	18
AR24	0	AR49	8
AR25	0	AR50	3

### APPENDIX 13

Use of discourse marker *like* per each participant in SLC

Participant's code	Frequency of DM use	Participant's code	Frequency of DM use
AR1	29	AR26	1
AR2	7	AR27	2
AR3	37	AR28	2
AR4	1	AR29	5
AR5	7	AR30	4
AR6	2	AR31	10
AR7	0	AR32	1
AR8	1	AR33	0
AR9	17	AR34	4
AR10	0	AR35	6
AR11	22	AR36	1
AR12	3	AR37	5
AR13	17	AR38	3
AR14	4	AR39	5
AR15	1	AR40	0
AR16	2	AR41	0
AR17	0	AR42	3
AR18	1	AR43	4
AR19	3	AR44	3
AR20	6	AR45	1
AR21	8	AR46	1
AR22	0	AR47	0
AR23	1	AR48	1
AR24	1	AR49	17
AR25	0	AR50	1



## REFERENCES

- Aijmer, K. (1996) *Conversational Routines in English. Convention and Creativity*. London, Longman.
- Aijmer, K. (2002) *English Discourse Particles. Evidence from a Corpus*. Amsterdam/Philadelphia, John Benjamins.
- Aijmer, K. (2004) Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies*, 3(1), 173-190.
- Aijmer, K. (2009) "So er I just sort I dunno I think it's just because..." A corpus study of I don't know and dunno in learners' spoken English. *Language and Computers*, 68(1), 151-168.
- Aijmer, K. (2011) "Well I'm not sure I think... The use of well by non-native speakers." *International Journal of Corpus Linguistics* 16(2), 231-254.
- Aijmer, K. & A.-M. Simon-Vandenberghe (2011) "Pragmatic markers." *Discursive Pragmatics*, 8, 223-247.
- Al Kohlani, F. A. (2010) *The Function of Discourse Markers in Arabic Newspaper Opinion Articles*. Unpublished PhD dissertation. Georgetown University.
- Alghamdi, E. A. (2014) Discourse Markers in ESL Personal Narrative and Argumentative Papers: A Qualitative and Quantitative Analysis, *International Journal of Humanities and Social Science*, 4(4), 294-305.
- Algouzi, S. (2008) *Barriers to developing speaking competency in English: A case study of Saudi students in state schools*. Unpublished MA dissertation. Nottingham University.
- Alhuqbani, M. N. (2013) The English But and Its Equivalent in Standard Arabic: Universality vs. Locality. *Theory & Practice in Language Studies*, 3(12).
- Al-Khalil, T. (2005) *Discourse markers in Syrian Arabic: a study of halla? yacne, tayyeb, and lakan*. Unpublished PhD thesis. Essex University.

- Alptekin, C. (2002) Towards intercultural communicative competence in ELT. *ELT Journal*, 56(1), 57–64.
- Al-Saif, A., & Markert, K. (2010, May) The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In *LREC*.
- Andersen, G. (1998) “The pragmatic marker like from a relevance-theoretic perspective”. *Pragmatics and beyond new series*, 147–170.
- Andersen, G. (2000) “The role of the pragmaticmarker like in utterance interpretation”. In Andersen, G. & Fretheim, T. (eds), *PragmaticMarkers and Propositional Attitude*. Amsterdam, John Benjamins, pp. 17–38.
- Andersen, G. (2001) *Pragmatic Markers of Sociolinguistic Variation: A Relevance-Theoretic Approach to the Language of Adolescents*. Amsterdam, John Benjamins.
- Ariel, M. (1998) Discourse markers and form-function correlations. In Jucker, A. H. & Ziv, Y. (eds) *Discourse Markers: Description and Theory*. Amsterdam, John Benjamins. pp. 223–259
- Atkins, S., Clear, J., & Ostler, N. (1992) Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1–16.
- Baker, M. (1995) Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223–243.
- Baker, P. (2010) Corpus methods in linguistics. *Research Methods in Linguistics*. London, UK: Continuum International Publishing Group, pp. 93–113.
- Bardovi-Harlig, K. (2001), Evaluating the empirical evidence: Grounds for instruction in pragmatics. In Rose, K. R. & Kasper, G. (eds), *Pragmatics in language teaching*, New York, Cambridge University Press. pp. 13–32.
- Barlow, M. (1996) Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1(1), 1–37.

- Baron, Naomi S. (2000) *Alphabet to email: How written English evolved and where it's heading*. London: Routledge.
- Beeching, K. & Detges, U. (2014) *Discourse Functions at the Left and Right Periphery Crosslinguistic Investigations of Language Use and Language Change*. Leiden, Brill.
- Bhela, B. (1999) Native language interference in learning a second language: Exploratory case studies of native language interference with target language usage. *International Education Journal*, 1(1), 22–31.
- Biber, D, Conrad, S. & Reppen, R. (1998) *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge, Cambridge University Press.
- Biber, D. (1993) Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243–257.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London Longman.
- Biber, D. (2006) *University language: A corpus-based study of spoken and written registers. Studies in corpus linguistics*. Amsterdam, John Benjamins.
- Biber, D., & Conrad, S. (2009) *Register, genre, and style*. Cambridge, Cambridge University Press.
- Blakemore, D. (1987) *Semantic Constraints on Relevance*. Oxford, Blackwell.
- Blakemore, D. (1992) *Understanding Utterances*. Oxford, Blackwell.
- Blakemore, D. (1998) On the context for so-called discourse markers. In Williams, J. & Malmkjaer, K. (eds) *Context in language understanding and language learning*. Cambridge, Cambridge University Press, pp. 44–60.
- Blakemore, D. (2002) *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge, Cambridge University Press.

- Blakemore, D. (2004) Discourse markers. In Horn, L. R. & Ward, G. (eds) *The Handbook of Pragmatics*. Oxford, Blackwell, pp. 221–240.
- Blass, R. (1990) Constraints on relevance: a key to particle typology. *Notes on Linguistics*, 48, 8–20.
- Blyth, C., Recktenwald, S., & Wang, J. (1990) I'm like, "say what?!" A new quotative in American oral narrative. *American Speech*, 215–227.
- Bolden, G. B. (2006) Little words that matter: discourse markers “so” and “oh” and the doing of other-attentiveness in social interaction. *Journal of Communication*, 56 (4), 661–688.
- Bolden, G. B. (2008) “So what's up?” Using the discourse marker so to launch conversational business. *Research on Language and Social Interaction*, 41(3), 302–337.
- Bolden, G. B. (2009) Implementing incipient actions: The discourse marker ‘so’ in English conversation. *Journal of Pragmatics*, 41(5), 974–998.
- Brand, C., & Kämmerer, S. (2006) The Louvain International Database of Spoken English Interlanguage(LINDSEI): Compiling the German Component. *Corpus Technology and Language Pedagogy*. Frankfurt aM, Peter Lang, pp. 127–140.
- Brinton, L. J. (1990) The development of discourse markers in English. In Fisiak, J. (ed) *Historical Linguistics and Philology*. Berlin, Mouton de Gruyter, pp. 45–71.
- Brinton, L. J. (1996) *Pragmatic Markers in English. Grammaticalization and Discourse Functions*. Berlin, Mouton de Gruyter.
- Brinton, L. J. (1998) The flowers are lovely; only, they have no scent: The evolution of a pragmatic marker. *Anglistentag 1997 Giessen. Proceedings. Trier: WVT*, 9–33.
- Brown, G., & Yule, G. (1983) *Discourse Analysis*. Cambridge: Cambridge University Press
- Burns, A. (1998) Teaching Speaking. *Annual Review of Applied Linguistics*, 18, 102–123.

- Buyse, L. (2007) Discourse marker so in the English of Flemish university students. *Belgian Journal of English Language and Literatures (BELL), New Series*, 5, 79–95.
- Buyse, L. (2012). So as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44(13), 1764-1782.
- Callies, M. & Szczesniak, K. (2008) ‘Argument realization, information status and syntactic weight: a learner-corpus study of the dative alternation’, in Walter, M. and Grommes, P. (eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, pp. 165-78
- Canale, M., & Swain, M. (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Carter, R. & McCarthy, M. (2006) *Cambridge Grammar of English: A comprehensive guide to spoken and written grammar and usage*. Cambridge, Cambridge University Press.
- Carter, R. (1998) Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52(1), 43–56.
- Carter, R. A. & McCarthy, M. J. (1997) *Exploring Spoken English*. Cambridge, Cambridge University Press.
- Carter, R., Hughes, R., & McCarthy, M. (2000). *Exploring Grammar in Context*. Cambridge: Cambridge University Press.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995) Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5–35.
- Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz, & S. J. Samuels (Eds.), *Comprehending oral and written language*. San Diego: Academic Press, pp.83 – 113.

- Collins Cobuild. (2006) *Collins COBUILD Advanced Learner's English Dictionary*. London, Harper Collins.
- Corbin, J., & Strauss, A. (eds) (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. (2nd ed). Thousand Oaks, CA, Sage.
- Crystal, D. (1988) Another look at, well, you know.... *English Today*, 4(01), 47–49.
- Dailey O'Cain, J. (2000) The sociolinguistic distribution of and attitudes toward focuser like and quotative like. *Journal of Sociolinguistics*, (4)1, 60–80.
- De Cock, S. (1999) Repetitive phrasal chunkiness, and advanced EFL speech and writing. In Mair, C. & Hundt, M. (eds) *Corpus Linguistics and Linguistics Theory*. Amsterdam, Rodopi 51-68.
- De Cock, S. (2004) Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), New Series*, 2(2), 225–246.
- De Klerk, V. (2005) Procedural meanings of well in a corpus of Xhosa English. *Journal of Pragmatics*, 37. 1183–1205.
- Demerci, M. & Kleiner, B. (1997) Discourse markers in second language research. *Journal of Intensive English Studies*, 11. 131–142.
- Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford, Oxford University Press.
- Dunning, T. (1993) Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1). 61-74.
- Edmondson, W. (1981) *Spoken discourse: A model for analysis*. London and New York: Longman.
- Erman, B. (1987) *Pragmatic Expressions in English*. 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.

- Eslami-Rasekh, Z. (2005) Raising the pragmatic awareness of language learners. *ELT Journal*, 59(3), 199–208.
- Ferrara, K. & Bell, B. (1995) “Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of be+like”. *American Speech*, 70, 265–290.
- Fischer, K. (1998) Validating semantic analyses of discourse particles. *Journal of Pragmatics*, 29(2), 111–127.
- Fischer, K. (2000) Discourse particles, turn-taking, and the semantics-pragmatics interface. *Revue de Sémantique et Pragmatique*, 8, 111–137.
- Fox Tree, J. E. (2006) Placing like in telling stories. *Discourse Studies*, 8(6), 723–743.
- Fox Tree, J. E., & Schrock, J. C. (2002) Basic meanings of “you know” and “I mean”. *Journal of Pragmatics*, (34)6, 727–747.
- Fraser, B. (1988) Types of English discourse markers. *Acta Linguistica Hungarica*, 38, 19–33.
- Fraser, B. (1990) An approach to discourse markers. *Journal of Pragmatics*, 14, 383–395.
- Fraser, B. (1996) Pragmatic markers. *Pragmatics*, 6, 167–190.
- Fraser, B. (1999) What are discourse markers in English? *Journal of Pragmatics*, 31, 931–952.
- Fraser, B. (2006) Towards a Theory of Discourse Markers. In: Kerstin Fischer (ed) *Approaches to Discourse Particles*. Amsterdam, Elsevier. pp. 189–204.
- Fukuya, Y., Reeve, M., Gisi, J. & Christianson, M. (1998) *Does focus on form work for teaching sociopragmatics?* Paper presented at the annual meeting of the International Conference on Pragmatics and Language Learning, Urbana, IL.
- Fuller, J. M. (2003a) Discourse marker use across speech contexts: a comparison of native and non-native speaker performance. *Multilingual*, 22, 185–208.

- Fuller, J. M. (2003b) Use of the discourse marker like in interviews. *Journal of Sociolinguistics*, (7), 365–377.
- Fuller, J. M. (2003c) The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, 35(1), 23–45.
- Fung, L. & Carter R. (2007) Discourse markers and spoken English: native and learner use in pedagogic settings. *Applied Linguistics*, (28)3, 410–429.
- Gaddafi, A.M. (1990) *A study of discourse markers in Libyan spoken Arabic*. Unpublished PhD thesis. London Birkbeck University.
- Gahl, S., Yao, Y., & Johnson, K. (2012) Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789-806.
- Gast, V. (2006), Introduction. In: *Zeitschrift für Anglistik und Amerikanistik*. Special Issue on the Scope and Limits of Corpus Linguistics, 54(2), 113-120.
- Ghobrial, A. (1993) *Discourse markers in colloquial cairene Egyptian Arabic. A pragmatic perspective*. Unpublished PhD thesis. Boston University.
- Gilquin, G. (2008) Hesitation markers among EFL learners: Pragmatic deficiency or difference? *Pragmatics and Corpus Linguistics: A mutualistic entente*, 2, 119.
- Gilquin, G., De Cock, S. & Granger, S. (eds) (2010) *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve, Presses Universitaires de Louvain.
- Goldbergju, L. (1980) *Discourse particles: An analysis of the role of 'y'know', 'I mean', 'well', and 'actually' in conversation*. Unpublished PhD thesis. Cambridge University.
- Goss, E. & Salmons, J. (2000) The evolution of a Bilingual Discourse Marking System: modal particles and English markers in German-American dialects. *International Journal of Bilingualism*, 4(4), 469–484.



- Götz, S. & Schilk, M. (2011) Formulaic sequences in spoken ENL, ESL, and EFL. In Marianne Hundt & Joybrato Mukherjee (eds) *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam, John Benjamins, pp. 79–100.
- Granger, S. (1998) The computer learner corpus: a versatile new source of data for SLA research. *Learner English on Computer*, 3–18.
- Granger, S. (2009) The contribution of learner corpora to second language acquisition and foreign language teaching. In Aijmer, K. (ed) *Corpora and Language Teaching*. Amsterdam, John Benjamins, pp. 13–32.
- Halliday, M. A. K (1980). Modes of meaning and modes of expression: Types of grammatical structure and their determination by different semantic functions. In D. J. Allerton (ed.), *Function and context in linguistic analysis*. Cambridge University Press, pp. 57-79
- Halliday, M. A. K. (1973) *Explorations in the functions of language*. London: Edward Arnold.
- Halliday, M.A.K & Hasan, R. (1976) *Cohesion in English*. London, Longman.
- Halliday, M.A.K. (2004). *An Introduction to Functional Grammar*. 3rd edition, revised by Christian M.I.M. Matthiessen. London: Arnold.
- Hansen, Maj-Britt M. (1997) *Alors* and *donc* in spoken French: A reanalysis. *Journal of Pragmatics*, 28, 153–287.
- Hansen, Maj-Britt M. (1998) *The Function of Discourse Particles. A Study with Special Reference to Spoken Standard French*. Amsterdam, John Benjamins.
- Hasko, V. (2012) Qualitative corpus analysis. In Chapelle, C.A. (ed) *The Encyclopedia of Applied Linguistics*. Malden, MA, Wiley-Blackwell.

- Hasselgren, A. (2002) Learner corpora and language testing. Small words as markers of learner fluency. In Granger, S., Hung, J.& Petch-Tyson, S. (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia, John Benjamins, pp. 143–173.
- Hays, P. R. (1992) Discourse markers and L2 acquisition. In Staub, D.& Delk, C. (eds) *Proceedings of the Twelfth Second Language Research Forum*. Michigan: Papers in Applied Linguistics –Michigan. pp. 24–34.
- 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.
- Hellermann, J & Vergun, A. (2007) Language which is not taught: the discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39, 157–179.
- Hlavac, J. (2006) Bilingual discourse markers: evidence from Croatian English code-switching. *Journal of Pragmatics* 38(11), 1870–1900.
- Holmes, J. (1986) “Functions of you know in women’s and men’s speech”. *Language in Society*, 15, 1–22.
- Horn, M. et al (2001) Cue words and the topic structure of spoken discourse: the case of Swedish men ‘but’. *Journal of Pragmatics*, 33, 1061–1081.
- House, J. (1996) Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18(02), 225-252.
- House, J. (2010) The pragmatics of English as a lingua franca. In Trosborg, A. (ed) *Pragmatics across Languages and Cultures*. Handbook of Pragmatics, vol. 7. Berlin/New York, Mouton de Gruyter, pp. 363–390.

- House, J. (2013) Developing pragmatic competence in English as a lingua franca: using discourse markers to express (inter) subjectivity and connectivity. *Journal of Pragmatics*, 59, 57–67.
- Hunston, S. & Francis, G. (1999) *Pattern Grammar: A Corpus-driven Approach to the Grammar of English*. Amsterdam, Benjamins.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge, Cambridge University Press.
- Hunston, S. (2006) Corpus linguistics. *Linguistics*, 7(2), 215–244.
- Hussein, M. (2009) *Relevance theory and procedural meaning: the semantics and pragmatics of discourse markers in English and Arabic*. Unpublished PhD thesis. Newcastle University.
- Johnson, A. (2002) “So ...?”: Pragmatic implications of so-prefaced questions in formal police interviews. *Language in the Legal Process*, 91–110.
- Johnstone, B. (2008) *Discourse analysis*. Malden, MA: Blackwell.
- Jucker, A. & Smith, S. (1998) And people just you know like “wow”: discourse markers as negotiating strategies. In Jucker, A. & Ziv Y. (eds) *Discourse Markers: Description and Theory*. Amsterdam/Philadelphia, Benjamins. pp. 171–202.
- Jucker, A. (1993) The discourse marker well: A relevance theoretical account. *Journal of Pragmatics*, 19, 435–452.
- Jucker, A. H. & Ziv, Y. (1998a) *Discourse Markers: Descriptions and Theory*. Benjamins, Amsterdam, Netherlands.
- Jucker, A. H. & Ziv, Y. (1998b) Discourse markers: introduction. In Jucker, A. & Yael, Z. (eds). *Discourse Markers. Theory and Descriptions*. (Pragmatics & Beyond New Series 57). Amsterdam, Benjamins. pp. 1–12.
- Kasper, G. & Blum-Kulka, S. (1993) *Interlanguage Pragmatics*. Oxford, Oxford University Press.

- Kasper, G. & Roever, C. (2004) Pragmatics in second language learning. In Hinkel, E. (ed) *Handbook of Research in Second Language Teaching and Learning*. Mulway, N.J, Lawrence Erlbaum Associates. pp. 317–334.
- Kasper, G. & Schmidt, R. (1996) Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition*, 18, 149–169.
- Kasper, G. (1982) Teaching-induced aspects of interlanguage discourse. *Studies in Second Language Acquisition*, 4, 99–113.
- Kasper, G. (1989) Interactive procedures in interlanguage discourse. In Olesky, W. (ed) *Contrastive Pragmatics*. Amsterdam, John Benjamins. pp. 189–229.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Addison Wesley Longman.
- King, B. (2009) Building and analysing corpora of computer-mediated communication. In Paul Baker (Ed.), *Contemporary corpus linguistics*. London: Continuum, pp. 301–320.
- Knott, A. & Dale, R. (1994) Using a set of linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1), 35–62.
- Kurdi, H. (2008) *The use of discourse markers by Syrian Arabic learners of English*. Unpublished PhD thesis. University of Manchester.
- Labov, W. (1981) Field methods used by the project on linguistics change and variation. *Sociolinguistic Working Paper* 81. Austin, Texas, South Western Educational Development Laboratory.
- Lam, Phoenix W.Y., (2010) Toward a functional framework for discourse particles: a comparison of well and so. *Text & Talk* 30 (6), 657–677.
- Lam, W. Y. P. (2008) *Discourse particles in an intercultural corpus of spoken English*. Unpublished PhD thesis. The Hong Kong Polytechnic University.

- Lee, H. (2004), Discourse markers use in native and non-native English speakers. In Moder, C. & Martinovic-Zic, A. (eds) *Discourse across Languages and Cultures*. Amsterdam, John Benjamins, pp. 117–127.
- Leech, G. (1983) *Principles of pragmatics*. London: Longman.
- Leech, G. (1997) Teaching and language corpora: a convergence. In Wichmann, A., Fligelstone, S., McEnery, T. & Knowles, G. (eds) *Teaching and Language Corpora*. London and New York, Longman.
- Leech, G. (2000) Grammars of Spoken English: New Outcomes of Corpus-Oriented Research. *Language Learning*, 50(4), 675-724.
- Lee-Wong, S. M. (2001). Coherence, focus and structure: The role of discourse particle NE. *Pragmatics*, 11(2).
- Lenk, U. (1995) Discourse markers and conversational coherence. In Wårvik, B., Tanskanen, S.-K. & Hiltunen, R. (eds) *Organization in Discourse*. Proceedings from the Turku Conference. Turku, University of Turku. pp. 341–352.
- Lenk, U. (1998) *Marking Discourse Coherence: Functions of Discourse Markers in Spoken English*. Tübingen, Gunter Narr Verlag.
- Levinson, S. C. (1983) *Pragmatics*. Cambridge, Cambridge University Press.
- Liao, S. (2009) Variation in the use of discourse markers by Chinese teaching assistants in the US. *Journal of Pragmatics*, 41, 1313–1328.
- Lörscher, W. (1986) Conversational structures in the foreign language classroom. In Kasper, G. (ed) *Learning, Teaching and Communication in the Foreign Language Classroom*. Aarhus, Aarhus University Press. pp. 11–22.
- Lüdeling, A. & Kytö, M. (2008) *Corpus Linguistics: An International Handbook*. Vol. 1. Berlin, New York, Mouton de Gruyter.
- Macaulay, R. (2002) You know, it depends. *Journal of Pragmatics*, 34(6), 749–767.

- Mahboob, A. & Elyas, T. (2014) English in the Kingdom of Saudi Arabia. *World Englishes*, 33(1), 128–142.
- Matsuda, M. (1999) Interlanguage pragmatics: what can it offer to language teachers. *The CATESOL Journal*, 11(1), 39–59.
- Mayor, M. (ed) (2009) *Longman Dictionary of Contemporary English*. Essex, Pearson Education.
- McCarthy, M. J. & Carter, R. A. (1994) *Language as Discourse: Perspectives for Language Teaching*. London, Longman.
- McEnery, T. & Wilson, A. (1996) *Corpus Linguistics*. Edinburgh, Edinburgh University Press.
- Meehan, T. (1991) It's like, 'What's happening in the evolution of like?' A theory of grammaticalization. *Kansas Working Papers in Linguistics*, 16, 37–51.
- Meyer, C. (2002) *English Corpus Linguistics: An Introduction*. Cambridge, Cambridge University Press.
- Miller, J. & Weinert, R. (1995). The function of LIKE in dialogue. *Journal of Pragmatics*, 23(4), 365–393.
- Ministry of Education (2002) *General Directorate of Curricula (2002): English for Saudi Arabia*. Jeddah, Almadina Almunawara Press.
- Moreno, A. E. (2001) Native speaker – non-native speaker interaction: the use of discourse markers. *ELIA*, 2, 129-142
- 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, John Benjamins.

- Nikula, T. (1996) *Pragmatic Force Modifiers. A study in Interlanguage Pragmatics*. Jyväskylä, University of Jyväskylä.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007) *From Corpus to Classroom*. Cambridge: Cambridge University Press.
- Oakes, M.P. (1998) *Statistics for Corpus Linguistics*. Edinburgh, Edinburgh University Press.
- Östman, J. O. (1981) *'You Know': A discourse-functional study*. Amsterdam, John Benjamins
- Östman, J. O. (1982) The symbiotic relationship between pragmatic particles and impromptu speech. In Enkvist, N. (ed), *Impromptu Speech: A symposium*, pp.147–177. Abe, Abe Akademi.
- Östman, J. O.(1995) Pragmatic particles twenty years after. In Wårvik, B., Tanskanen, S.-K. & Hiltunen, R. (eds) *Organization in Discourse*. Proceedings from the Turku Conference. Turku, University of Turku. pp. 95–108.
- Oxford Advanced Learner's Dictionary (1989) *Oxford Advanced Learner's Dictionary of Current English*, Hornby, A. S. (4th ed.). Oxford, Oxford University Press.
- Oxford Advanced Learner's Dictionary (2000) *Oxford Advanced Learner's Dictionary of Current English*, Hornby, A. S. (6th ed., edited by Wehmeier, S.). Oxford, Oxford University Press.
- Özbek, N. (1995) *Discourse markers in Turkish and English: A comparative study*. Unpublished PhD Thesis. Nottingham University.
- Pichler, H. (2010) Methods in discourse variation analysis: reflections on the way forward. *Journal of Sociolinguistics*, 14(5), 581–608.
- Raymond, G. (2004) Prompting action: the stand-alone "so" in ordinary conversation. *Research on Language and Social Interaction*, 37(2), 185–218.
- Rayson, P. & Garside, R. (2000) Comparing corpora using frequency profiling. In *proceedings of the workshop on Comparing Corpora, held in conjunction with the*

- 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1-8 October 2000, Hong Kong, pp. 1 – 6.
- 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*. Amsterdam, Elsevier. pp. 339–358.
- Redeker, G. (1990) Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14, 367–381.
- Risselada, R. & Spooren, W. (1998) Introduction: discourse markers and coherence relations. *Journal of Pragmatics*, 30(2), 131–133.
- Romaine, S. & Lange, D. (1991) The use of like as a marker of reported speech and thought: a case of grammaticalization in progress. *American Speech*, 66(3), 227–279.
- Rose, K. R. & Kasper, G. (eds) (2001) *Pragmatics in Language Teaching*. New York, Cambridge University Press.
- Sacks, Harvey, Schegloff, Emanuel A., Jefferson, Gail, (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 (4-1), 696-735.
- Saeed, A. T. & Fareh, S. (2006) Difficulties encountered by bilingual Arab learners in translating Arabic ‘fa’ into English. *International Journal of Bilingual Education and Bilingualism*, 9(1), 19–32.
- Sampson, G., & McCarthy, D. (eds) (2005) *Corpus Linguistics: Readings in a Widening Discipline*. London, Continuum.
- Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M. & Gagnon, L. (1997) Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, 9, 191–217.



- Schachter, J. (1990) Communicative competence revisited. In Harley, B., Allen, P., Cummins, J. & Swain, M. (eds) *The Development of Second Language Proficiency*. New York, Cambridge University Press, pp. 39–49.
- Schiffrin, D. (1987) *Discourse Markers*. Cambridge, Cambridge University Press.
- Schiffrin, D. (2001) Discourse markers: language, meaning, and context. In Schiffrin, D., Tannen, D. & Hamilton, H. E. (eds) *The Handbook of Discourse Analysis*. Malden, MA, Blackwell Publishers, pp. 54–75.
- Schleef, E. (2004) Gender, power, discipline, and context: on the sociolinguistic variation of okay, right, like, and you know in English academic discourse. *Texas Linguistic Forum*, 48, 177–186.
- Schmied, J. (1993) Qualitative and quantitative research approaches to English relative constructions. In: Souter, C., Atwell, E. (Eds.), *Corpus Based Computational Linguistics*. Rodopi, Amsterdam pp. 85–96.
- Schourup, L. (1985) *Common Discourse Particles in English Conversation*. New York, Garland.
- Schourup, L. (1999) Discourse markers. *Lingua*, 107, 227–265.
- Sedgwick, R. (2001) *Education in Saudi Arabia*, WENR, 14(6). Available at <http://www.wes.org/ewenr/01nov/practical.htm> [Accessed 13th January 2011].
- Seidman, I. (2013) *Interviewing as Qualitative research: A Guide for Researchers in Education and the Social Sciences* (4th ed), New York, Teachers College Press.
- Siegel, M. E. (2002) Like: the discourse particle and semantics. *Journal of Semantics*, 19(1), 35–71.
- Siepmann, D. (2005) *Discourse Markers across Languages: A Contrastive Study of Second-level Discourse Markers in Native and Non-native Text with Implications for General and Pedagogic Lexicography*. New York, Routledge.

- Sinclair, J. (2004a) Introduction. In Sinclair, J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam and Philadelphia, Benjamin, pp. 1–10.
- Sinclair, J. (2004b) *Trust the Text: Language, Corpus and Discourse*. London, Routledge.
- Sinclair, J. M. & Mauranen, A. (2006) *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam, John Benjamins.
- Smith, N., Hoffmann, S. & Rayson, P. (2008) Corpus tools and methods, today and tomorrow: incorporating linguists' manual annotations. *Literary and Linguistic Computing*, 23(2), 163–180.
- Sperber, D. & D. Wilson. (1995) *Relevance: Communication and Cognition*. (2nd ed.) Oxford, Blackwell.
- Stevenson, A. (ed). (2010) *Oxford Dictionary of English*. Oxford, Oxford University Press.
- 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 and Communication*, 15(1) 63–88.
- Tagliamonte, S. (2005) So who? Like how? Just what? Discourse markers in the conversations of young Canadians. *Journal of Pragmatics*, 37, 1896–1915.
- Takahara, P. O. (1998) Pragmatic functions of the English discourse marker anyway and its corresponding contrastive Japanese discourse markers. In Jucker, A.H. & Ziv, Y. (eds) *Discourse Markers: Descriptions and Theory*. Amsterdam, Benjamins, pp. 327–351.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam, John Benjamins.
- Traugott, E. C. (1995) The role of the development of discourse markers in a theory of grammaticalization. *ichl XII*, Manchester, 1–23.
- Trillo, J. R. (2002) The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, 34(6), 769–784.

- Underhill, N. (1987) *Testing Spoken Language (A Handbook of Oral Testing Techniques)*. Cambridge, Cambridge University Press.
- Underhill, R. (1988) “Like is, like, focus”. *American Speech*, 63, 234–246.
- Van Bogaert, J. (2009) A reassessment of the syntactic classification of pragmatic expressions: the positions of you know and I think with special attention to you know as a marker of metalinguistic awareness. *Language and Computers*, 69(1), 131-154.
- Van Dijk, T. A. (1979) Pragmatic connectives. *Journal of Pragmatics*, 3, 447–456.
- Watts, Richard 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.
- Waugh, L. R., Fonseca-Greber, B., Vickers, C., & Eröz, B. (2007) Multiple empirical approaches to a complex analysis of discourse. In: Coulson, S., Gonzalez-Marquez, M., Mittelberg, I. & Spivey, M. (eds) *Methods in Cognitive Linguistics*. Philadelphia, John Benjamins, pp. 120–148.
- Weydt, H. (2006). What are particles good for? In Fischer, K. (ed) *Approaches to Discourse Particles*. Amsterdam, Elsevier, pp. 13–39.
- Willis, D. & Willis, J. (1996) Consciousness-raising activities in the language classroom. In Willis, J. & Willis D. (eds) *Challenge and Change in Language Teaching*. Oxford, Heinemann, pp. 63–76.
- Xiao, R. (2008) Theory-driven corpus research: using corpora to inform aspect theory. In: Lüdeling, Anke & Kytö, Merja (eds) *Corpus Linguistics. An International Handbook*. Berlin, Mouton de Gruyter.
- Young, R. (1991) *Variation in Interlanguage Morphology*. New York, Peter Lang.