

“And, like, they said ... well, you know”

*A corpus-based study of the discourse markers
‘like’, ‘well’ and ‘you know’ in spoken Norwegian
learner language and British English*

Karoline Lilleås Sandal



A thesis presented to the Department of Literature, Area
Studies and European Languages

University of Oslo

May 2016

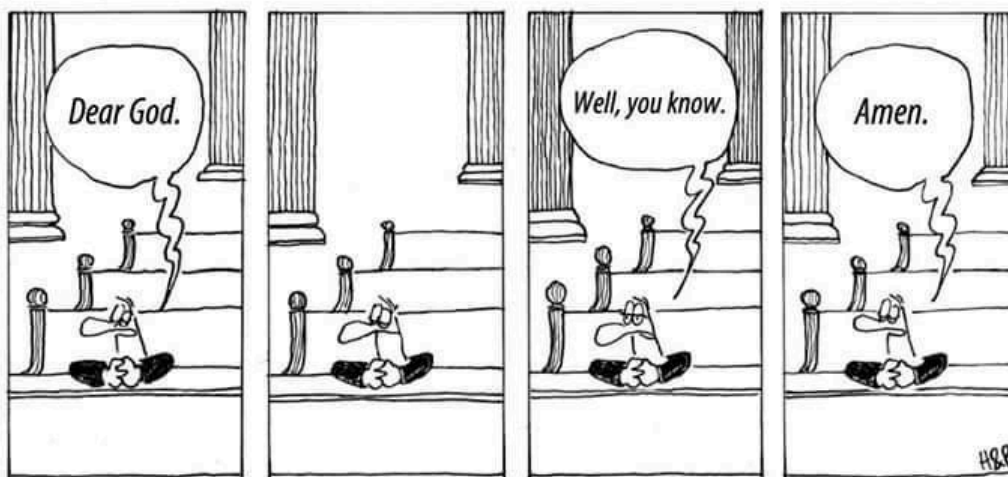
”And, like, they said ... well, you know”

A corpus-based study of the discourse markers ‘like’, ‘well’ and ‘you know’ in spoken Norwegian learner language and British English

Karoline Lilleås Sandal

University of Oslo

May 2016



© Karoline Lilleås Sandal

2016

“And, like, they said ... well, you know”: A corpus-based study of the discourse markers ‘like’, ‘well’, and ‘you know’ in spoken Norwegian learner language and British English

Karoline Lilleås Sandal

<http://www.duo.uio.no>

Print: Representeren, University of Oslo

Abstract

Drawing on data from the two corpora LINDSEI-NO and LOCNEC, this project examines and compares the three discourse markers *like*, *well* and *you know* as spoken by Norwegian advanced learners of English (NLEs) and British English speakers (BESs). Making use of Granger's (2015) Contrastive Interlanguage Analysis² model, the study investigates whether and how the two speaker groups differ in their production of these DMs, and discusses possible explanations for the observed differences. This study is founded on the assumption that Norwegian learners differ from the British English speakers in their use of DMs, which is also confirmed by the results revealing a general underrepresentation of the three DMs in the NLE data. The findings indicate that although both speaker groups display a similar functional scope of the DMs, i.e. both speaker groups use the markers for the same pragmatic functions, the frequency of these pragmatic functions is mostly lower in the NLE data than in the BES data. A qualitative analysis of a selection of these functions uncovers that the learners often demonstrate a less systematic use of the DMs than the British speakers, especially in terms collocation patterns. It is argued that lack of input in English textbooks and classroom instruction may explain some of this observed discrepancy between the two speaker groups. However, the findings suggest that this factor alone cannot explain the differences, and that other factors, such as transfer from Norwegian, learners' lack of self-confidence and cultural differences, also require consideration.

In addition to mapping the two speaker groups' use of the three DMs, this thesis contributes to the discussion of the importance of teaching DMs in school, as they are crucial elements in the learners' effort to achieve communicative competence. It is argued that by directing the learners' attention toward discourse markers in the teaching of English in Norway, they may reach a higher level of communicative competence and thereby display a more systematic and reference language-like usage of DMs.

Keywords: Advanced learner English, corpus-based analysis, corpus linguistics, contrastive interlanguage analysis, discourse markers, spoken language corpora, learner corpus research, pedagogical implications, underrepresentation, qualitative analysis, quantitative analysis

Acknowledgements

To my supervisor Professor Signe Oksefjell Ebeling for your great support and much appreciated comments and ideas throughout the process of writing this thesis –

To Professor Susan Nacey and Research Fellow Hege Larsson Aas at Hedmark University of Applied Sciences for giving me access to LINDSEI-NO –

To Professor Sylvie de Cock at Université catholique de Louvain for providing me with the metadata for LOCNEC –

To Ole, for your valuable proofreading, patience and kindness in the stressful last weeks before the deadline –

To my beloved family and friends, for showing interest in my work and supporting me when most needed –

To my wonderful fellow students and great friends at the 8th floor for the endless hours of drinking coffee, sharing ideas and discussing complete nonsense. You have made this year unforgettable –

Thank you.

Table of Contents

<i>Abstract</i>	V
<i>Acknowledgements</i>	VII
<i>List of figures and tables</i>	XI
<i>List of abbreviation</i>	XIII
1 Introduction	1
1.1 Background, aims and scope	1
1.1.1 Features studied.....	2
1.1.2 Hypothesis and research questions.....	3
1.2 Thesis outline	4
2 Theoretical background and previous studies	5
2.1 Discourse markers	5
2.1.1 Defining discourse markers	7
2.1.2 General features of discourse markers	8
2.2 Previously used frameworks for analyzing discourse markers	11
2.2.1 Schiffrin (1987)	12
2.2.2 Redeker (1990; 1991)	14
2.2.3 Aijmer (2002).....	15
2.2.4 Müller (2005)	18
2.3 Characteristics of the selected discourse markers	20
2.3.1 <i>Well</i>	21
2.3.2 <i>You know</i>	24
2.3.3 <i>Like</i>	27
2.4 Factors potentially affecting learners' acquisition of discourse markers	31
2.4.1 L1 Transfer	31
2.4.2 Lack of Input	32
2.4.3 Speaker Confidence	33
3 Method	34
3.1 Learner Corpus Research and Learner Corpora	35
3.1.1 Learner Corpora and Learner Corpus Design	35
3.1.2 Learners: who are they?	37
3.2 Contrastive Interlanguage Analysis²	38
4 Material	41
4.1 The corpora in this study	41
4.1.1 LINDSEI and LINDSEI-NO	42
4.1.2 LOCNEC	43
4.1.3 Evaluation of comparability	43
4.1.4 Evaluation of authenticity and representativeness	44
4.2 Extracting the data	44
4.2.1 Statistical calculations	45
4.3 Framework of classification	46
5 Analysis and Results	51
5.1 Preliminary frequency analysis	51
5.1.1 Corpus-Driven Frequency Analysis and Quantitative CIA ²	52
5.1.2 Individual differences within LOCNEC and LINDSEI	53

5.2	<i>Like</i>	55
5.2.1	Frequency analysis.....	55
5.2.2	Functional analysis	59
5.3	<i>Well</i>	73
5.3.1	Frequency analysis.....	73
5.3.2	Functional Analysis.....	76
5.4	<i>You know</i>	90
5.4.1	Frequency analysis.....	90
5.4.2	Functional analysis	95
5.5	Summary and discussion of results	110
5.5.1	Reference language and interlanguage variety practice	110
5.5.2	Factors contributing to the learners' underrepresentation of DMs	113
6	Concluding Remarks	116
6.1	Strengths and limitations of this study with suggestions for future studies	116
6.2	Pedagogical implications	118
	<i>References</i>	120
	<i>Appendix 1: LINDSEI-NO metadata</i>	1
	<i>Appendix 2: LOCNEC metadata</i>	3
	<i>Appendix 3: Screenshot of FilemakerPro file</i>	5
	<i>Appendix 4: Transcription guidelines</i>	6
	<i>Appendix 5: LINDSEI Picture description task</i>	10

List of tables and figures

Tables

Table 1: Aijmer's (2002) framework for categorizing DMs	18
Table 2: Müller's (2005) framework for categorizing the DMs <i>so, well, you know</i> and <i>like</i>	20
Table 3: Textual, interpersonal and qualifying functions of <i>well</i> in previous research ...	23
Table 4: Textual, interpersonal and qualifying functions for <i>you know</i> in previous research.....	27
Table 5: Textual, interpersonal and qualifying functions of <i>like</i> in previous research	30
Table 6: Framework of classification	47
Table 7: LOCNEC and LINDSEI-NO total number of occurrences (including non- discourse markers)	52
Table 8: LOCNEC and LINDSEI-NO discourse marker occurrences only.....	52
Table 9: Utterance position of <i>like</i> in LOCNEC and LINDSEI-NO	56
Table 10: Orientation of <i>like</i> in LOCNEC and LINDSEI-NO	58
Table 11: Textual, interpersonal and qualifying functions of <i>like</i> in LOCNEC and LINDSEI-NO	59
Table 12: Definitions and examples of the textual functions of <i>like</i>	60
Table 13: Definitions and examples of the interpersonal functions of <i>like</i>	65
Table 14: Utterance position of <i>well</i> in LOCNEC and LINDSEI-NO.....	74
Table 15: Orientation of <i>well</i> in LOCNEC and LINDSEI-NO	76
Table 16: Textual, interpersonal and qualifying functions of <i>well</i> in LOCNEC and LINDSEI-NO	77
Table 17: Definitions and examples of the textual functions of <i>well</i>	78
Table 18: Definitions and examples of the qualifying function of <i>well</i>	87
Table 19: Utterance position of <i>you know</i> in LOCNEC and LINDSEI-NO.....	91
Table 20: Orientation of <i>you know</i> in LOCNEC and LINDSEI-NO	94
Table 21: Textual, interpersonal and qualifying functions of <i>you know</i> in LOCNEC and LINDSEI-NO	95
Table 22: Definitions and examples of the textual functions of <i>you know</i>	96
Table 23: Definitions and examples of interpersonal functions of <i>you know</i>	101

Figures

Figure 1: Learner Corpus Design (Granger, 2008: 264)	36
Figure 2: Varieties of English by Granger (Granger, 2002: 8)	37
Figure 3: The Integrated Contrastive Model (Gilquin, 2000/2001: 100)	39
Figure 4: Contrastive Interlanguage Analysis 2 (Granger, 2015: 17)	40
Figure 5: LOCNEC and LINDSEI-NO discourse marker occurrences only.....	53
Figure 6: The frequency of <i>like, well</i> and <i>you know</i> for each speaker in LOCNEC.....	54
Figure 7: The frequency of <i>like, well</i> and <i>you know</i> for each speaker in LINDSEI-NO	54
Figure 8: Utterance position of <i>like</i> in LOCNEC and LINDSEI-NO.....	56
Figure 9: Orientation of <i>like</i> in LOCNEC and LINDSEI-NO	58

Figure 10: Textual functions of <i>like</i> in LOCNEC and LINDSEI-NO	61
Figure 11: Interpersonal functions of <i>like</i> in LOCNEC and LINDSEI-NO	66
Figure 12: Qualifying functions of <i>like</i> in LOCNEC and LINDSEI-NO.....	71
Figure 13: Utterance position of <i>well</i> in LOCNEC and LINDSEI-NO.....	75
Figure 14: Orientation of <i>well</i> in LOCNEC and LINDSEI-NO	76
Figure 15: Textual functions of <i>well</i> in LOCNEC and LINDSEI-NO	79
Figure 16: Interpersonal functions of <i>well</i> in LOCNEC and LINDSEI-NO	82
Figure 17: Qualifying functions of <i>well</i> in LOCNEC and LINDSEI-NO.....	86
Figure 18: Utterance position of <i>you know</i> in LOCNEC and LINDSEI-NO.....	92
Figure 19: Orientation of <i>you know</i> in LOCNEC and LINDSEI-NO	94
Figure 20: Textual functions of <i>you know</i> in LOCNEC and LINDSEI-NO	97
Figure 21: Interpersonal functions of <i>you know</i> in LOCNEC and LINDSEI-NO	102
Figure 22: Qualifying functions of <i>you know</i> in LOCNEC and LINDSEI-NO	108

List of Abbreviations

BES: British English speaker

CA: Contrastive Analysis

CIA: Contrastive Interlanguage Analysis

DM: Discourse marker

EFL: English as a foreign language

EOL: English as an official language

ESL: English as a second language

ILV: Interlanguage variety

L1: First language

L2: Second language

NLE: Norwegian learner of English

Pttw: Per 10,000 words

RLV: Reference language variety

RQ: Research question

SLA: Second language acquisition

Corpora used

LINDSEI: The Louvain International Database of Spoken English Interaction

LINDSEI-NO: A subcorpus of LINDSEI consisting of spoken data collected from Norwegian native speakers

LOCNEC: The Louvain Corpus of Native English Conversation

1 Introduction

1.1 Background, aims and scope

[...] a normal child acquires knowledge of sentences not only as grammatical, but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what manner. In short, a child becomes able to accomplish a repertoire of speech acts, to take part in speech events, and to evaluate their accomplishments by others. (Hymes, 1972: 277)

Communicative competence is defined by Hymes (1966; 1972) as a language speaker's knowledge of the grammatical aspects of language (syntax, morphology, phonology), the social aspects of language, and how to use language appropriately in conversation. This competence is considered crucial for a speaker's ability to succeed in conversation, and can therefore be seen as an important aspect on which to direct focus in language teaching. In her study of discourse markers, Schiffrin points to four different aspects of communicative competence (the expressive, social, cognitive and textual), and claims that "discourse markers – expressions like *well*, *but*, *oh*, and *y'know* – are one set of linguistic items that function in [the] cognitive, expressive, social and textual domains" (2001: 54). In other words, discourse markers (DMs) can be seen as important for learners striving to achieve such communicative competence.

The English subject curriculum presented by the Ministry of Education and Research in Norway further underlines the importance of communicative competence by presenting one competence aim in particular, saying that pupils should be able to "express [themselves] fluently and coherently in a detailed and precise manner suited to the purpose and situation" (Sommerseth, 2013). A comparison of two English textbooks used for high school level English education in Norway, *Passage* and *Targets* (Burgess and Sørhus, 2009; Haugen et al., 2009), and their corresponding websites¹, reveals that *Passage* focuses little on communicative competence, while *Targets* is much more dedicated to this aspect of the English language. However, apart from a few cases of utterance-initial *well* found in *Targets*, very little attention is given to discourse markers in particular in either book. This lack of focus on DMs in textbooks is a problem also addressed in previous research (Lam, 2009; Hellermann and Vergun, 2007; Fung and Carter, 2007). And, as input is found to affect

¹ Accessed at <http://www.lokus.no/open/Targets-2015> for *Targets* and <http://passage-new.cappelendamm.no/> for *Passage* (Access date 12. April 2016).

learners' output (Ortega, 2009: 59), it is viable to assume that this absence of focus may negatively affect learners' competence within this field.

With reference to the above-mentioned issues, I find it interesting and important to study how Norwegian learners of English use discourse markers. Through a comparison of spoken corpus data produced by Norwegian learners of English (NLEs) and British English speakers (BESs) studying at university level, I aim to discover to what extent the learners differ from their British peers, whether there are areas in which they struggle to achieve competence, and discuss possible explanations to the observed tendencies. This study is highly explorative in nature, and aims to reveal general tendencies of how the Norwegian learners use of the DMs under study. Hopefully, such tendencies will in the future be used as the starting point for further in-depth investigations, which are beyond the scope of this thesis.

1.1.1 Features studied

The large amount of items classified as discourse markers (see section 2.1) require the choice of an appropriate subset in order to ensure feasibility, especially when the limited time frame of a master thesis is taken into account. Note that this approach prohibits a generalization of DM use within the two speaker groups. For the present study, the three discourse markers *like*, *well* and *you know* have been selected on both practical and theoretical grounds.

The theoretical reason for choosing these three DMs is that they serve different functions in discourse and thus cover a large area within the field. For instance, *well* is typically found in question/answer adjacency pairs, *you know* is hearer-oriented, and *like* is often classified as being information-centered. Moreover, these three markers have Norwegian equivalents such as *vel*, *du skjønner*, *du vet*, *likksom*, and *jeg bare*. These Norwegian counterparts may be central in terms of L1 transfer, potentially causing the NLEs to misuse or under- or overrepresent the target DMs.

In terms of practical factors affecting my choice of DMs, the three markers of interest all provide sufficient data in the corpora that will be used for the analysis. This is essential, as “repeated events” is taken as crucial in the formulation of generalizations about language” (Tognini-Bonelli, 2001: 58). Also, my choice of discourse markers has been affected by previous research. As this study is based on a very recent subsection of the LINDSEI corpus

on which little research has been conducted, it has been of interest to study DMs that have been studied before, in order to be able to make comparisons to other interlanguage varieties with different first language (L1) backgrounds.

1.1.2 Hypothesis and research questions

Based on the motivations presented above and findings from previous research (discussed in section 2.1), I have formulated the following hypothesis and research questions for this study:

Hypothesis: If it is true that little emphasis is placed on discourse markers in English education in Norway, and furthermore that input (i.e. education) determines learners' output (i.e. production of discourse markers), I would expect to discover a discrepancy between the Norwegian learners and the British English speakers in terms of the overall frequency and the functional use of the discourse markers *like*, *well*, and *you know*.

In order to test and explore this hypothesis, the following research questions (RQs) are asked:

RQ1: Are there differences between Norwegian advanced learners of English and British speakers of English in their use of the discourse markers *like*, *you know* and *well* in spoken language?

RQ2: Do the Norwegian learners use these discourse markers for different pragmatic functions than the British speakers?

RQ3: If RQ2 can be answered with 'yes', to what extent can the observed discrepancy between the two speaker groups be explained by factors such as lack of input, L1 transfer and speaker confidence?

RQ4: Are there any other factors that appear to have affected the learners' use of the three discourse markers?

RQ1 aims to test the hypothesis, revealing either an under- or overrepresentation of the three DMs among the NLEs. RQ2 aims to explore, in a more qualitative manner, how the two speaker groups differ in their use of the three DMs, and with this question I hope to be able to point at possible problematic areas for the learners in their use of the three DMs. RQ3 is based on previous research claiming that learners' production of DMs is affected by factors such as input, transfer and anxiety (Ortega, 2009). With this question I wish to explore to what extent these factors are present in my data, and to what extent they affect the NLEs'

production of the target DMs. RQ4 opens up for a discussion of contributing factors, other than the ones listed above, that may affect the learners' production of the three DMs. The two latter questions are highly explorative, and I do not expect the data of this study to reveal any conclusive answers here. Rather, I hope that my results will point to some possible explanations, which then could be further investigated in future studies.

1.2 Thesis outline

The present thesis is built up of six chapters. Following this introduction, Chapter 2 presents relevant theoretical background about discourse markers in general, previous research within this field, and a close examination of previous research on the three DMs under investigation, namely *like*, *well* and *you know*. Furthermore, this chapter also presents theoretical perspectives on L1 transfer, input and self-confidence, i.e. three factors that possibly affect learners' production of DMs.

Chapter 3 and 4 provide a description of the method and material applied in this study. In Chapter 3, a short introduction to corpus linguistics is given, before entering the field of Learner Corpus Research (LCR) and Contrastive Interlanguage Analysis (CIA). Chapter 4 introduces the two corpora used for this study, LOCNEC and LINDSEI-NO, with a critical evaluation of their authenticity and representability. Thereafter follows a presentation of this thesis' analysis procedure and framework of classification.

In chapter 5, the quantitative and qualitative analysis of the three discourse markers *like*, *well*, and *you know* is presented, followed by a discussion of the results where I revisit the research questions and discuss potential factors that can explain the observed differences between the NLEs and BESs.

Finally, Chapter 6 sums up the findings, gives an evaluation of some strengths and limitations of the study, and considers pedagogical implications and proposes topics for further research.

2 Theoretical background and previous studies

This chapter begins with a brief introduction to the development of research on discourse markers since the 1970s, before defining what DMs are and providing an overview of their general features. Section 2.2 takes a closer look at four previous frameworks for categorizing and analyzing DMs, presented by Schiffrin (1987), Redeker (1990; 1991), Aijmer (2001) and Müller (2005), which will serve as the starting point for the development of this study's analytical framework. Following these general accounts of discourse markers, the next step is to narrow the focus. Dealing with DMs as a homogeneous group is problematic, as this category contains many diverse and multifunctional items originating from several different word classes. It is therefore necessary to dedicate section 2.3 to the three markers *well*, *you know* and *like*, which are the main focus of this study. This section draws on previous research, focusing on the markers' lexical, syntactic, semantic, functional, sociolinguistic and stylistic features. Finally, section 2.4 introduces three potential factors considered to affect learners' acquisition and production of DMs, namely L1 transfer, input and speaker-confidence.

2.1 Discourse markers

Research on discourse markers as part of conversational grammar fully emerged in the 1960s and 1970s, with scholars such as Weydt (1969) and Gülich (1970) as some of the forerunners. People started to analyze discourse markers, which earlier had been seen as hesitational phenomena or performance errors (Östman, 1995: 98), as systemic pragmatic elements. Since then, interest in these markers has increased significantly, broadening the field to not only focus on DMs through a pragmatic and discourse analytic point of view, but to also look at them from the aspect of sociolinguistics, second language acquisition (SLA) and language pedagogy. This has led to useful insights about both DMs as a group (Schiffrin, 1987; Aijmer, 2002; Müller, 2005) and about individual markers such as *well* (Greasley, 1994; Schourup, 2001; Müller, 2004; Cuenca, 2008; Aijmer, 2011), *now* (Schourup, 2011), *no* (Lee-Goldman, 2011), and *you know* (Erman, 2001; Macaulay, 2002; Fox Tree and Schrock, 2002; Mukherjee, 2009). Scholars have, for example, investigated differences between men and women's use of the discourse marker *you know* (Holmes, 1986; Macaulay, 2002) and differences between adults and adolescents' use of *you know* (Erman, 2001). With regard to

the difference between men and women, Holmes (1986: 17) reports that there is no significant difference in the overall usage, whereas Macaulay claims that “women are more likely to use *you know* than men” (2002: 765). As for adult use as opposed to adolescent use, Erman (2001: 1356) discovers a functional difference where adults prefer a text-oriented use of *you know*, whereas adolescents tend to use it more as social and metalinguistic monitor.

Most research on DMs has looked at how the markers are used in the English language (Fraser, 1990; Blakemore, 2006; Bell, 2010), but also other languages have been studied, such as Italian (Bazzanella, 1990), Singaporean English (Gupta, 1992), Japanese (Cook, 1990), Indonesian (Wouk, 2001), Spanish (Montolío Durán and Unamuno, 2001), French (Cadiot et al., 1985), Hebrew (Maschler, 1997; Ariel, 1998), and German (Abraham, 1991). Since discourse markers are considered informal, most research on them is conducted on spoken material (Fuller, 2003; Fung and Carter, 2007; Aijmer, 2011; Buysse, 2012). Yet, some scholars also look at written discourse (Castele and Collewaert, 2013) and different genres of written discourse such as narratives (Koike, 1996) and newspapers (Cotter, 1996).

Müller (2005: 1) argues that discourse markers are understudied phenomena in research within SLA, where the applicants are non-native speakers learning English as a second or foreign language and not bilingual speakers having acquired English as children. Since this was stated, research on language learners’ use of pragmatic markers in spoken language has increased rapidly with studies conducted by for example Fung & Carter (2007), Hirzalla (2007), Romero Trillo (2002), Mukherjee (2009), Hellermann & Vergun (2007), Aijmer (2011), Brand & Götz (2011) and Buysse (2012). Scholars within this field are particularly interested in how the learner groups use the markers compared with native speakers in terms of functions and expressed meaning (Mukherjee, 2009; Aijmer, 2011; Buysse, 2012). They also focus on errors and fluency (Brand and Götz, 2011), why differences in usage emerge (Liu, 2013), and how use of pragmatic markers can be taught in the classroom (Romero Trillo, 2002; Fung and Carter, 2007; Hellermann and Vergun, 2007; Mukherjee, 2009). Several scholars have discovered an overrepresentation of discourse markers by learners of English (Müller, 2004; Gilquin, 2008; Aijmer, 2011; Buysse, 2012; Unaldi, 2013). This finding is partly explained as being a result of learners’ need to use discourse markers more often for monitoring functions, “i.e. for monitoring the speaker’s progression through the discourse” (Aijmer, 2011: 251). But as opposed to these scholars, Fung & Carter (2007) and Müller (2005) report that DMs are underrepresented in learner language both generally and

across different pragmatic functions, and claim that lack of focus on DMs in English teaching may explain parts of the observed discrepancy.

2.1.1 Defining discourse markers

Defining discourse markers is a challenging task as they formally have little in common and can serve a number of different functions. Attempts have been made to treat discourse markers as any other word class, but since “their primary task in language is not related to the propositional aspects of sentences, but to the pragmatic functioning of language” (Östman, 1995: 98), these attempts have been misguided. Thus, we are left with a heterogeneous group of particles difficult to classify and define across genres and languages. This heterogeneous group is built up of items belonging to a number of different grammatical categories, such as connectives (*now, so, but*), pragmatic uses of modal adverbs (*typically, possibly*), interjections (*oh, well, um*), routines (*how are you*), feedback signals (*yeah, oh*), vocatives, disjuncts (*frankly, fortunately*), approximants (such as hedges), and reformulation markers (*that is, in other words*) (Gülich and Kotschi, 1983: 227; He and Young, 1998; Aijmer and Simon-Vandenberg, 2011). There is little consensus among researchers on how to define this group of items, which markers are classified as DMs, and which term to use when referring to them. Therefore, the above list is not finite, and can be both extended and shortened depending on which definition one follows.

The term ‘discourse marker’, as used by scholars such as Schiffrin (1987) and Müller (2005), is only one of many terms used to refer to the items belonging to the heterogeneous group described above. Some other terms commonly used are ‘cue phrases’ (Knott and Dale, 1994), ‘discourse connectives’ (Blakemore, 1987), ‘discourse operators’ (Redeker, 1990), ‘discourse particles’ (Schourup, 1999), ‘discourse signaling devices’ (Polanyi and Scha, 1983), ‘phatic connectives’ (Bazzanella, 1990), ‘pragmatic connectives’ (Stubbs, 1983), ‘pragmatic expressions’ (Erman, 1987), ‘pragmatic operators’ (Ariel, 1993), ‘pragmatic particles’ (Östman, 1995), ‘semantic conjuncts’ (Quirk et al., 1985), and ‘sentence connectives’ (Halliday and Hasan, 1976). All these terms attribute approximately the same characteristics to the markers, albeit with some variation. As Fung and Carter put it, “[t]he multiplicity of terminology surrounding DMs reflects diverse research interests and analytical categories, as well as difficulties in accounting for them adequately in theoretical terms” (2007: 411). The present study will not participate in the discussion of which term is most suitable and which

is not. The term ‘discourse marker’ has been chosen, as this appears to be the most neutral and including term of them all.

Below follows a short list of definitions of DMs as presented by different scholars within the field. For this study, I will mainly follow Aijmer’s definition, as I find the passage “facilitating the hearer’s interpretation of the utterance” (2002: 2) to be particularly accurate for the present study’s results where all three DMs are frequently found with textual, interpersonal and qualifying functions serving to guide the hearer to make the appropriate interpretation of the utterance.

“[Discourse markers are] sequentially dependent elements which bracket units of talk” (Schiffrin, 1987: 31).

“[Discourse markers] signal a sequential relationship between the current basic message and the previous discourse” (Fraser, 1990: 383).

“[Discourse markers are] linguistic items of variable scope, and whose primary function is connective. By ‘variable scope’ I mean that the discourse segment hosting a marker may be of almost any size or form [...]” (Hansen, 1997: 160).

“[Discourse markers are] a syntactically heterogeneous class of expressions which are distinguished by their function in the discourse and the kind of meaning they encode” (Blakemore, 2006: 221).

“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 clues” (Aijmer, 2002: 2).

2.1.2 General features of discourse markers

As mentioned, DMs can be described as comprising several features and as serving a number of different functions in an utterance. Inspired by Aijmer and Simon-Vandenberg’s (2011: 226) summary of the classifications of the marker’s characteristics, the DMs have been grouped into five different categories according to their features: 1) phonological and lexical features, 2) syntactic features, 3) semantic features, 4) functional features, and 5) sociolinguistic and stylistic features. These categories represent the main characteristics of interest in this study’s analysis. The phonological, lexical, syntactic and semantic features will mainly serve to help define whether an item is considered to be a discourse marker or not, whereas the functional features will serve to categorize the DMs in terms of the

pragmatic function they perform in the utterance. Little attention will be given to the DMs' sociolinguistic and stylistic features, as this would require a different approach to the data than the one engaged for the present study. It is important to note that not all characteristics described below are shared by all DMs. Examples are included from this study's data where appropriate.

With regard to **phonological and lexical features**, the markers are characterized as short (Östman, 1982: 149) and phonologically reduced items (Schiffrin, 1987) which are difficult to specify lexically (Schiffrin, 1986: 42, 47, 62), and which form a phonologically independent tone unit (Quirk et al., 1985: 1112; Hansen, 1997: 156). This is illustrated by examples (2.1) and (2.2) below. As these are phonological features, they do not always become clear from the written transcriptions. In ambiguous cases, listening to the recorded files has facilitated the categorization.

- 2.1 *and the first time we were **like** . what . what's this . why is the sea red (N826)*
2.2 *I guess she wanted (eh) ... **well** . gues she felt she: (eh) she looks better than (N145)*

In terms of **syntactic features**, DMs most commonly occur sentence-initially (Brinton, 1996). However, studies have shown that DMs also occur sentence-medially and sentence-finally in certain contexts (Aijmer, 2002). Example (2.1) above illustrates the use of an utterance-medial DM, whereas *well* in example (2.2) is used utterance-initially. Furthermore, DMs appear as syntactically optional (Schourup, 1999: 230; Fraser, 1988: 22), and are defined as having weak clause association (Brinton, 1996; Hansen, 1997). This means that they are only loosely attached to the clause and can be omitted without altering the grammaticality of the host sentence. In example (2.3), for instance, *you know* can be removed without making the sentence ungrammatical.

- 2.3 *all of a sudden this thunderstorm and it was a nightmare yeah it was awful but **you know** it wasn't too windy because you (E98)*

Concerning **semantic features**, DMs are characterized as items with little or no propositional meaning (Östman, 1982). The markers are also considered non-truth-conditional, i.e. they do not contribute to the truth-conditions of the utterance (Hölker, 1991; Schourup, 1999: 232). This point relates to the syntactic features discussed above, and indicates that DMs can be

omitted from the clause without altering the meaning of the utterance. Furthermore, Svartvik (1980: 169) claims that markers are difficult to translate into other languages, often because of their lack (or reduction) of semantic meaning.

In terms of **functional features**, DMs are considered multifunctional (Aijmer, 2002: 55), i.e. they operate on several linguistic levels simultaneously. They are also characterized as marking connectivity (Schourup, 1999: 230; Fraser, 1990: 383; Hansen, 1997: 160), having indexical functions (Schiffrin, 1987), and as being reflexive in the sense that “they comment on the utterance and thus assist in the interpretation of the utterance” (Aijmer and Simon-Vandenberg, 2011: 225). There will be a closer examination of the different functions the individual markers of this study may serve in section 2.3. Yet, as a preliminary example, sentence (2.4) below illustrates the use of *well* serving both to introduce an explanation and to appeal for the hearer’s attention.

2.4 *I loved it I really did but I because it's so em .. well it's classed as quite an arty film . all my friends are like <?> oh no I'm not interested I wanna watch Reservoir Dogs (E238)*

In the final category for classifying discourse markers, **sociolinguistic and stylistic features**, DMs are classified as belonging mostly to the oral and informal discourse (Brinton, 1996). This is illustrated by example (2.5), where the speaker clearly applies an informal way of expressing his opinion about Italian pizza.

2.5 *the pizza in Italy was really . weird . they had like pizza with (eh) Nutella on it . and then they had pizza with pommes frites (N809)*

As becomes clear from the many characteristics of discourse markers presented above, it is difficult to provide a short definition which includes all possible items without being too extensive. Consequently, we are often left with broad definitions as the ones presented in section 2.1.1 above, and further classifications of each discourse marker are required. Therefore, section 2.3 is dedicated the three DMs examined in this study, and aims to define and classify them in terms of the five categories of features described above. But prior to this, an introduction to some of the major previous works and frameworks for categorizing discourse markers will be offered. These frameworks will serve as a starting point and as a source of inspiration for this thesis’ analytical framework, which is presented in section 4.3.

2.2 Previously used frameworks for analyzing discourse markers

Discourse markers have been studied on the basis of several theoretical frameworks within the field of pragmatics and semantics. For instance, Brown and Levinson's (1987) study of markers as an illocutionary force, and Mittwoch's (1976) study of them as speech act adverbials both base their analysis on Grice's (1975) influential Speech Act Theory. Schiffrin's (1987) and Redeker's (1990) approaches are strongly influenced by Halliday and Hasan's (1976) model of coherence, and can thus be seen as studies within the frameworks of discourse coherence. Furthermore, Blakemore (1987), Watts (1988) and Jucker (1993) approach discourse markers from a relevance theoretic point of view, while Traugott (1995), Aijmer (2002) and Fraser (1990; 1999) approach the topic from a grammatical-pragmatic point of view. Müller (2005) is an influential work on discourse markers produced by learners of English, and applies a bottom-up approach where she attempts to analyze their usage with as little influence from already existing theories as possible. Due to the limited scope of this thesis, I will not go into an in-depth discussion about all the different theoretical approaches mentioned above. They are only included to illustrate the broad scope of research that exists on discourse markers, and to make clear that the present approach is only one out of many possible ways to conduct research within the field of semantics and (grammatical) pragmatics.

Instead, this section, which serves as a theoretical foundation for the compilation of the present study, will focus on the most relevant individual works for this thesis: Schiffrin (1987), Redeker (1990), Aijmer (2002), and Müller (2005). These four scholars differ greatly in their approach to discourse markers. Yet, they all seem to discover approximately the same functions for the target DMs and draw similar conclusions. The following section is dedicated to provide a detailed description of these four works, presenting the different scholars' theoretical frameworks and how they have chosen to classify and analyze their target discourse markers.

2.2.1 Schiffrin (1987)

Schiffrin (1987) presents one of the earliest and most prominent studies on discourse markers. Operating within the field of interactional sociolinguistics analyzing discourse markers from the theoretical framework of discourse coherence, she studies the use of the discourse markers *oh, well, and, but, or, so, because, now, then, I mean, and y'know* in data collected from sociolinguistic group interviews with seven Jewish Americans from Philadelphia.

As already mentioned, Schiffrin's model of discourse coherence is influenced by Halliday and Hasan (1976). They define cohesion as a semantic concept which "refers to relations of meaning that exist within the text, and that define it as a text", and argue that cohesion is incorporated into the language and "occurs where the *interpretation* of some element in the discourse is dependent on that of another" (Halliday and Hasan, 1976: 4-5). The one *presupposes* the other. Thus, cohesion is seen as a textual concept. Building on this view, Schiffrin (1987) emphasizes that cohesion is not only a textual concept, but can also apply to sentences and words of other aspects of discourse. She thus analyzes discourse markers as items contributing to discourse cohesion.

Based on her analysis of the eleven DMs, Schiffrin (1987) presents a model of discourse coherence built up of five different planes of talk, each contributing its own type of coherence. These five planes are a) **Exchange Structure**, b) **Action Structure**, c) **Ideational Structure**, d) **Participation Framework**, and e) **Informational Structure**. She argues that all markers have one primary function and other secondary functions within these planes. This view of DMs as having one core meaning/function is very common in discourse marker research, and is shared by scholars such as Fraser (1990), Schourup (1985), and Redeker (1991).

The first plane, **Exchange Structure**, includes turns of "conditionally relevant adjacency-pair[s]" (Schiffrin, 1987: 24), i.e. anticipated turns initiated by adjacency pairs such as greetings, questions and answers. This fixed system of turn-taking is by Goffman (1981a) referred to as the system constraint of talk, and refers to speech situations where the first turn in the adjacency pair determines, or at least affects, the respondent's following turn. **Action Structure** reflects the order of speech acts, the way in which the interlocutors manage

themselves and others in terms of appropriateness and speech conventions, and the decision procedures upon which these speech acts are based (Schiffirin, 1987: 25). This category focuses on what Goffman (1981a: 21) refers to as ritual constraints, which are concerned with how the speaker manages him/herself and the interlocutor in terms of politeness and appropriateness. Schiffirin views both the Exchange Structure and the Action Structure as pragmatic categories.

Ideational Structure, however, is viewed more as a semantic category. The units within this structure reflect ideas or propositions, and how these ideas relate to each other within the discourse. Here, Schiffirin distinguishes between three different relations: cohesive relations, topic relations, and functional relations. Cohesive relations are closely connected to Halliday and Hasan's (1976) definition of cohesion, and reflect ties that occur when the interpretation of one element presupposes information from another clause. Topic relations reflect the way in which topics and subtopics are organized, and functional relations reflect the role ideas play in relation to one another within the overall discourse (Schiffirin, 1987: 26).

Participation Framework refers to the relation between the speaker and the utterance and between the speaker and the hearer (Schiffirin, 1987: 27). An example of a speaker-utterance relationship is how a speaker chooses to tell a story; either by reporting only what actually happened, or by also including subjective evaluation. The latter option will open up for hearer-evaluation, and is thus an example showing that speaker-utterance relations also affect speaker-hearer relations. There are also numerous ways and levels in which the hearer and speaker may relate to one another, such as teacher/student relationships, or hearers being intended recipients of talk vs. hearers being unintended recipients of talk, and these different relationships affect the discourse.

The **Information State** also refers to the ways in which the speaker and hearer relate to one another. But as the Participation Framework refers to the interactional aspect, the Information State refers to the cognitive aspects in which the interlocutors may relate. This cognitive state includes what Schiffirin refers to as management of speaker/hearer knowledge and speaker/hearer meta-knowledge. Speaker/hearer knowledge refers to the interlocutors' specific knowledge about a certain topic, whereas speaker/hearer meta-knowledge refers to the interlocutors' conscious awareness of this respective knowledge and of the other person's knowledge (Schiffirin, 1987: 28).

2.2.2 Redeker (1990; 1991)

Redeker (1990; 1991) also approaches DMs from a discourse-coherence point of view. Yet, she is critical of Halliday and Hasan (1976), and claims that utterances can be cohesive without overlapping references in the text, and that overlaps do not necessarily mean that an utterance is cohesive. She uses the two examples below to illustrate her point, as both examples demonstrate a cohesive text although they lack overlapping textual references. The examples in (2.6) illustrate *semantic coherence*, where cohesion occurs between the ideas or semantic meanings the utterances convey (Redeker, 1990: 368). Example (2.7) illustrates *pragmatic coherence*, where cohesion is a result of pragmatic relations rather than textual relations. Consequently, these examples illustrate that “coherence always has both an *ideational* and a *pragmatic* component” (Redeker, 1990: 369), but that one of them usually is weighted heavier than the other.

- 2.6 (1a) *Sally is crying.*
(1b) *Nanny has thrown out the time-worn teddy bear.*
(1c) *The holes were getting too large to fix. (Redeker, 1990: 367)*

- 2.7 (2a) *Take those dirty shoes off!*
(2b) *There’s a brand-new carpet in the hallway.*
(2c) *Mom’s ALREADY mad at me. (Redeker, 1990: 368)*

Redeker further classifies the two components of discourse coherence, *semantic* and *pragmatic coherence*, by dividing *pragmatic coherence* into a) *rhetorical relations*, and b) *sequential units*. *Rhetorical relations* comprises units considered cohesive based on the observation that the strongest relation lies between the utterances themselves or between “the beliefs and intentions motivating them” (i.e. antithesis, concession, evidence, justifications, conclusions, etc.), as in example (2.7 (2a-c)) above (Redeker, 1990: 369). *Sequential units*, on the other hand, include utterances that are considered cohesive without displaying any obvious ideational or rhetorical relation (paratactic sequential relations: transitions, change of topic, and hypotactic sequential relations: leading out of a commentary, correction, paraphrase, digression etc. (Redeker, 1990: 369).

Redeker thus builds on Schiffrin’s (1987) model, but criticizes it for relying too “heavily on the markers themselves in identifying the intended relations” (Redeker, 1991: 379). This is especially seen in Schiffrin’s (1987) two planes of talk *Information Structure* and *Participation Framework*. As an improvement of these two planes, Redeker (1991: 1167)

proposes a broader model within which any utterance is considered to participate in at least two of the components, and where one constituent will be more dominant than the other. This model involves three components of coherence: 1) **Ideational structure**, 2) **Rhetorical structure**, and 3) **Sequential Structure**. According to Redeker, two discourse units are related 1) **ideationally** “if their utterances in the given context entails the speaker’s commitment to the existence of that relation in the world the discourse describes” (i.e. temporal sequence, elaboration, cause, reason, consequence etc.), 2) **rhetorically** “if the strongest relation is not between the propositions expressed in the two units but between the illocutionary intentions they convey” (i.e. antithesis, concession, evidence, justification, and conclusion), and 3) **sequentially** if there is a paratactic relation (“transition between issues or topics that either follows a preplanned list or is locally occasioned”) or a hypotactic relation (transitions “leading into or out of a commentary, correction, paraphrase, aside, digression, or interruption segment”) (1991: 1168).

2.2.3 Aijmer (2002)

Aijmer (2002) presents an empirical study of the discourse markers *now, oh, ah, just, sort of, actually* and *and that sort of thing* based on data from the 500,000 word London-Lund Corpus of Spoken English. This corpus includes spoken data collected from informal private conversations and public discussions and prepared speech produced by English speakers of both genders and from different social groups (Aijmer, 2002: 5).

Aijmer analyzes the discourse markers in the framework of grammatical pragmatics, believing that these markers have gone, or are going, through a process of grammaticalization. Grammaticalization refers to the development of lexical items or phrases from being used only “in certain highly constrained local contexts to be reanalyzed as having syntactic and morphological functions” (Traugott, 1995: 32). Aijmer (2002: 5) argues that this grammaticalization process may help explain why so many discourse markers are multifunctional and thus difficult to classify and define. She also sees core meaning as a very “abstract notion” (Aijmer, 2002: 23), and therefore finds Hansen’s approach to the analysis of discourse markers as having serving different functions according to the contexts in which they occur (1998: 87) as a better solution. This view leads Aijmer to describe discourse markers as “lexical words which have undergone a change of function from propositional meaning to textual and interpersonal function” (2002: 55). However, although Aijmer

provides a different explanation to how discourse markers have emerged than Schiffrin and Redeker, it seems as if they all see them as cohesive devices to a certain degree. This is especially clear in Aijmer's analysis of DMs, where her three main functions of DMs appear to be very similar to those presented by Schiffrin and Redeker. I therefore find it relevant and useful to draw on Aijmer's (2002) framework in addition to Schiffrin (1987) and Redeker's (1990; 1991) when creating the framework of classification for the present analysis.

In her analysis Aijmer (2002) discovers that discourse markers can a) occupy several positions in the utterance, b) be either prospective or retrospective, and c) serve three main functions: as phatic connectives in the interpersonal function, as framers in the textual function, and as qualifiers in the qualifying function. In terms of positions in the utterance, Aijmer (2002: 37) finds that discourse markers can serve as 'themes' occurring in the pre-front field, as 'insertions' occurring somewhere in the middle of the utterance, or as 'tails' or 'afterthoughts' occurring in the post-end field. Thus, discourse markers are not only phenomena occurring sentence-initially as Brinton (1996) claims.

In terms of utterance orientation, discourse markers are found to function both as prospective markers, directing the listener's attention forward to something that is about to come, and as retrospective markers, commenting on something that has already been said. Prospective markers are also called 'attention-getters' (Aijmer, 2002: 37), and typically serve to announce a new topic, a new point in discussion or refer to items on a list. The retrospective markers signal a reaction to something that has already been said, and this reaction is often followed by an explanation, elaboration, or justification (Aijmer, 2002: 37).

As for the main functions in the utterance, Aijmer reports that the markers can serve at three levels, the interpersonal, textual and qualifying. **Interpersonal functions**, or what Bazzanella (1990: 630) refers to as 'phatic connectives', includes cases where the discourse markers are used to express feelings and attitudes, to mark pauses and planning of speech, or to hedge or boost the illocutionary force of the utterance, i.e. the speaker's intention (Aijmer, 2002: 50). This category can be viewed as an extension of Schiffrin's two categories *Participation Framework* and *Information State*. Since the illocutionary force is a culturally dependent speech act, as different cultures have different strategies for expressing the intended meaning, one can say that the interpersonal function also deals with face and politeness (Aijmer, 2002: 39). Aijmer's **textual level** is inspired by Halliday's category with the same name. Halliday

defines textual meaning as an item's "relevance to the context" (1985: 53), and this is also what discourse markers functioning within the textual domain do in Aijmer's categorization. Accordingly, markers functioning within this domain may serve to mark transitions, introduce new turns, introduce explanations, justifications of backgrounds, introduce/close digressions, indicate self-corrections, or introduce direct speech. Thus, speakers use discourse markers with the textual function in order to help the listener to keep track of topic changes, repairs and other rapid changes in discourse that often occur in conversations. This category also bears resemblance to Redeker's (1990) *Sequential Relations* and Schiffrin's (1987) *Ideational Structure*.

Discourse markers used with **qualifying functions** serve to indicate that "some qualification is needed because the dialogue does not 'go well'" (Aijmer, 2002: 46). This usage can for instance occur in beginnings of disagreements, in exchanges, or before arguments where the speaker feels the need to express his/her response to what has been said. Discourse markers are also used as qualifiers when the speaker is listing several items. This category can be read as an extension of Schiffrin's (1987) *Exchange* and *Action Structure*, which is also concerned with speaker-hearer relations, appropriateness and politeness strategies.

Table 1 below presents an overview of Aijmer's framework for categorizing the discourse markers *now*, *oh*, *ah*, *just*, *sort of*, *actually* and *and that sort of thing* in terms of pragmatic functions and utterance orientation. This framework has a great influence on this study's framework (presented in section 4.3), where the same structure is applied and several of the pragmatic functions are included.

Table 1: Aijmer's (2002) framework for categorizing DMs

Variables	Values
Textual functions (Framers)	Marking transition
	Introducing a new turn
	Introducing an explanation, justification or background
	Introducing or closing a digression
	Self-corrections
	Introducing direct speech
Interpersonal functions (Phatic connectives)	Expressing attitudes, feelings and evaluations
	Hedges expressing uncertainty
	Boosters
	Hearer-oriented appeals for confirmation
	Expressing responses or reactions to the preceding utterance
	Backchannelling
	Face and politeness
Qualifying functions (Qualifiers)	Indicating agreement or disagreement
	Response to a question
	Indicating comparison or contrast
	Listing
Orientation	Prospective
	Retrospective

2.2.4 Müller (2005)

Simone Müller (2005: 23) introduces her work by claiming that, although discourse markers have been thoroughly investigated by several scholars during the past decades, few attempts have been made to systematically connect research on discourse markers with research on learner languages. Based on this argument, she aims to fill this gap. By investigating data extracted from the Giessen-Long Beach Chaplin Corpus (GLBCC), a 350,000 word corpus of native (mostly American) and non-native (mostly L1 German) English spoken data, Müller (2005: 24) analyzes how the discourse markers *so*, *well*, *you know*, and *like* are used by the two speaker groups both quantitatively and qualitatively.

Müller gives an account of the already existing frameworks presented by Schiffrin (1987) and Redeker (1991), and agrees with Redeker's criticism of Schiffrin. Yet, she does not find Redeker's (1991) revised model to be an improvement, as it does not open up for all potential functions of DMs, and by such does not appear to be any more precise than Schiffrin's model (Müller, 2005: 30). As a result, Müller decides to analyze her material with a bottom-up approach, where she, with as little influence from previous theories as possible, attempts to categorize her data based on how the target discourse markers behave in her material. She also disagrees with the assumption made by several other scholars such as Schourup (1985),

Schiffrin (1987), Fraser (1990), and Redeker (1991) that discourse markers have a core meaning, and rather sees them as multifunctional markers that possibly can serve several functions depending on the situation in which they occur. This brings her to the conclusion that the four discourse markers under investigation function on two main levels, textual and interpersonal, and that the markers within these two main levels may serve several sub-functions and thus convey several meanings (Müller, 2005: 30-31). Except for the fact that Müller only proposes two main levels, this approach is highly comparable with Aijmer (2002).

Müller's **textual** function is, like Aijmer's, similar to Schiffrin's (1987) ideational structure, but "goes slightly beyond" (Müller, 2005: 30). Instead of addressing the hearer, it serves to direct focus toward "expressions and propositional content expressed in units of various length" (Müller, 2005: 30). This can for instance include cases where the discourse markers are used to mark false start and repair, or to introduce quotations and examples. Other cases of discourse markers functioning at the textual level may for example include DMs used to structure discourse by marking transitions or indicating lexical/content search.

Discourse markers functioning at the **interpersonal** level focus on the speaker-hearer relationship rather than on the textual and structural properties of the utterance. Müller discovers that these cases most often occur when there is a shift or transition, and normally function to, for example, "mark a speech act, a response, an opinion, or an evaluation" (2005: 31). Consequently, this category can be seen as a combination of Schiffrin's (1987) *Action* and *Exchange Structure* and *Participation Framework* and *Information state*. Discourse markers within this category often serve as an appeal to the hearer in order to direct the hearer's focus toward a particular word or phrase or in order to elicit a certain response. This latter point is particularly relevant for *you know* and *like*. It also seems, according to Müller (2005: 31), that the functions within the interpersonal level are more dependent on the individual marker than functions at the textual level. She further states that these two levels are not closed groups, and that different sub-functions will occur depending on the markers and the contexts in which they occur. Table 2 below illustrates Müller's classification and results in terms of the textual and interactional functions of the discourse markers *so*, *well*, *you know*, and *like* as used by native and non-native speakers of English (Müller, 2005: 246).

Table 2: Müller's (2005) framework for categorizing the DMs *so*, *well*, *you know* and *like*

Level/marker	Textual	Interactional
<i>So</i>	Marking result or consequence	Speech act marker - question or request
	Main idea unit marker	Speech act marker - opinion
	Summarizing/rewording/giving an example	Marking implied result
	Sequential so	Marker of a transition relevance place
	Boundary marker	
<i>Well</i>	*Searching for the right phrase	*Indirect answer
	Rephrasing/correcting	Direct answer
	Quotative <i>well</i>	Response to self-raised expectations
	Introducing the next scene	Contributing an opinion
	*Conclusive well	*Continuing an opinion/answer
	Marking lexical/content search	Evaluating a previous statement
<i>You know</i>	Marking lexical/content search	"Imagine the scene"
	Marking false start and repair	"See the implication"
	Marking approximation	Reference to shared knowledge
	Introducing an explanation	Appeal for understanding
	Quotative <i>you know</i>	Acknowledge that the speaker is right
<i>Like</i>	Searching for the appropriate expression	
	Marking an appropriate number of quantity	
	Introducing an example	
	Introducing an explanation	
	Marking lexical focus	

As can be read from Table 2 above, Müller (2005) discovers certain differences between the native speakers and the German learners of English. Those functions marked with **bold** indicate that the native (American) speakers use this particular DM with this function significantly more often than the German speakers. And those marked with ***bold** indicate that the German speakers use this function for this particular marker significantly more often than the native speakers do. Thus, we can see that the native speakers for example use *you know* to 'mark lexical/content search' significantly more often than the non-native German speakers, whereas the non-natives use *well* to 'search for the right answer' significantly more often than the native speakers do. Also, *like* is found only at the textual level, whereas the three other markers can occur at both the textual and the interpersonal level.

2.3 Characteristics of the selected discourse markers

This section directs the attention to the three discourse markers under study, *well*, *you know* and *like*, with the purpose of delving deeper into their meanings and functions, as identified by previous research. The pragmatic functions identified for each DM are divided into a *textual*, *interpersonal* and *qualifying* level, following Aijmer's (2002) categorization of discourse markers. At the end of each sub-section, a short discussion of irrelevant cases, where *like*, *well*, and *you know* serve other functions than that of DM, will be addressed. This

section, together with the theory and previous research presented earlier in this chapter, will serve as a starting point for my own analysis, and it will remain to see whether this study's data reveals the same results and tendencies as the ones discussed here.

2.3.1 *Well*

Well is one of the discourse markers that has been most thoroughly investigated over the past decades, with scholars such as Lakoff (1973), Svartvik (1980) and Schourup (1985) as some of the pioneers. The discourse marker is classified as a reception marker (Jucker and Smith, 1998: 197) and as a “sharing device” (Svartvik, 1980: 168), and is described as anchoring “the speaker into a conversation precisely at those points where upcoming coherence is not guaranteed” (Schiffrin, 1987: 126). According to Schiffrin (1987: 103), *well* is used as a device with the purpose of creating coherence between phrases and words that not necessarily fit together in the speech context.

Lakoff (1973) particularly looks at how *well* is used in responses to questions, and discovers that it is used in situations where the speaker wishes to mark that the information provided to a certain degree is insufficient, and that it is up to the hearer to imagine the rest (1973: 463). In other words, an answer is only preceded by *well* if it is an indirect answer yielding insufficient information, or if the information preceded by *well* is only partly the answer and more information is to follow. This view also applies to narratives, where *well*, according to Lakoff (1973: 464), indicates that details have been omitted and that the story is not told in its entirety. This is illustrated in example (2.8) below.

2.8 *I went to (eh) I went on a road trip a road trip yeah on a bus . and (em) . yeah well the bus took about yeah it could be sort of twenty-one hours or something (N513)*

Since Lakoff's study was published, several scholars have looked at *well* from different perspectives, including Schiffrin's (1987) approach to *well* from the theoretical framework of discourse coherence, Jucker's (1993) analysis of *well* from a relevance-theoretic point of view, Norrick's (2001) analysis of *well* in oral narratives, and Müller's (2004; 2005) and Aijmer's (2011) focus on the use of *well* in learner language. This section pays special attention to the latter two approaches, as this study concerns *well* as used by learners of English.

Textual functions

Well is often used at the textual level to organize discourse (Müller, 2004; Müller, 2005; Aijmer, 2011), or to serve as a “framing device” (Svartvik, 1980: 174) with the purpose of helping the listener follow the course of the conversation. As a frame, *well* “shifts the topic focus to one of the topics which have already been under discussion”, “introduces explanations, clarifications, etc.”, “[indicates] the beginning of direct speech”, or serves as an “editing marker for self-correction” (Svartvik, 1980: 174f). One such use of *well* is illustrated by example (2.9) below, where the DM serves to shift the topic. Müller further lists ‘introducing a conclusion’, ‘introducing the next scene’, and ‘move to story’ as some of the central textual functions for *well* (Müller, 2004: 1163). Example (2.10) illustrates the use of *well* to introduce a conclusion.

2.9 *once you go to another country and you live there you soon pick it up* <A> *yeah you have to anyway* *yeah you have to manage mm well anyway I think that a lot of people there speak English so even if you have problems (E604)*

2.10 *but what can you do ... we’ve tried . and we’ve had several talks and we’ve . (er) come up with suggestions what do and . well . nothing changes ... so . yeah (N11)*

Interpersonal functions

As for functions at the interpersonal level, Svartvik claims that *well* sometimes serves as a discourse technique to hold the floor, mark a pause, a hesitator or to initiate a response/answer from the interlocutor (1980: 176). It can also serve to indicate that what follows is an incomplete or indirect answer to a question (Müller, 2004: 1163). Aijmer further adds the functions ‘indicating pause’, ‘self-correction’ and ‘planning of speech’ (2011: 236). Example (2.11) below illustrates the use of *well* to hold the floor.

2.11 *Harry Potter (uhu) as the . main character . but I saw it here in Norway so for me the actors wasn’t really famous . and . well . there was a lot of suspense around this play sort of . because at one point the actor gets naked . (N10)*

Qualifying functions

Well is often used at the qualifying level where there is disagreement, or where qualification is needed because the conversation is not going very well. Jucker (1993) states that *well* can be used as a negative politeness strategy to mitigate possible face-threats in situations where there is disagreement or misunderstandings, or where a request is rejected. Such uses of *well*

include expressing incongruity, denying a previous utterance, correcting a misunderstanding, refusing to answer a question directly, and rejecting an offer (Aijmer, 2011). Yet, *well* can also serve as a positive politeness strategy to indicate “agreement, positive reaction or attitude”, “reinforcement” (Svartvik, 1980: 173f), or to contribute an opinion and evaluate a previous answer (Müller, 2004: 1163). Example (2.12) illustrates the use of *well* to express negative politeness by correcting a misunderstanding, whereas example (2.13) shows the use of *well* to express positive politeness by confirming a previous statement.

- 2.12 *I don't I don't wanna be in that situation yeah . but then when you get a job*
 <A> *where: . where are you thinking about working* **well** *I have a job now . I work at a: (em) kiosk in the cinema okay selling popcorn . and candies (N450)*
- 2.13 <A> *I mean .. mm there's a sort of gap between them and you I guess* **well** *there is a difference yeah a big difference (E1060)*

Table 3: Textual, interpersonal and qualifying functions of *well* in previous research

Variables	Values
Textual functions	Topic shift
	Introducing explanation
	Introducing justification
	Introducing clarification
	Introducing direct speech
	Introducing conclusion
	Marking pause
	Marking planning of speech
Interpersonal functions	Hold the floor
	Initiate a response from the interlocutor
	Searching for the right word/phrase
	Indirect answer
Qualifying functions	Denying or rejecting a challenging statement
	Correcting a misunderstanding
	Refusing to answer a question directly
	Qualifying an opinion
	Confirming a previous statement
	Contributing an opinion

A note on the non-discourse marker functions of *well*

The usage of *well* as a discourse marker described above must be distinguished from other, non-discourse marker, functions of *well* as for instance an adverb, adjective, noun or verb. According to the *Oxford Advanced Learner's Dictionary* (OALD, 2005: 1733), *well* as an adverb may mean a) “in a good, right or acceptable way”, b) “thoroughly and completely”, c) “to a great extent or degree”, d) “**can/could** – easily”, e) “**can/could/may/might** – probably”, or f) “**can/could/may/might** – with a good reason”, as shown in example (2.14) below.

2.14 *The kids all behaved **well**. (OALD, 2005: 1733)*

Well as an adjective may, according to the *Oxford Advanced Learner's Dictionary* mean “in good health”, “in a good state or position”, as in (2.15), or “(as) – (to do sth) sensible” (OALD, 2005: 1734), as illustrated in example (2.15) below.

2.15 *It would be just as **well** to call and say we might be late (OALD, 2005: 1734)*

Furthermore, *well* as a noun refers to “a deep hole in the ground from which people obtain water”, “oil well”, a stairwell, or “the space in front of the judge in a court, where the lawyers sit” (OALD, 2005: 1733). And *well* functioning as a verb includes the meanings “(of a liquid) to rise to the surface of sth and start to flow”, as illustrated in example (2.16), or “(literary) (of an emotion) to become stronger”,.

2.16 *Tears were **welling** up in her eyes. (OALD, 2005: 1733)*

These uses of *well* are all considered non-discourse marker uses, and will therefore not be included in this thesis' analysis.

2.3.2 *You know*

In the literature, *you know* is referred to as being one of the most multifunctional discourse markers, and this makes it very challenging to classify and define (Müller, 2005: 147). Östman, for example, describes *you know* as a marker that implies “that the speaker wants to give the addressee a feeling of great power” (1981: 19), and claims that it also sometimes serves as a politeness marker. According to Schourup (1985), *you know* presumes that there is some kind of common knowledge between the two speakers, but that there exists some uncertainty about this common ground. The speaker therefore feels the need to use *you know* to make sure that the hearer knows and understands what is talked about. *You know* is further defined as an informal expression (OALD, 2005: 854), and characterized as an addressee-oriented marker which serves as a tool for the speaker to include the addressee in conversation (Östman, 1981: 18) and gain attention from the addressee (Schiffrin, 1987: 267).

Textual functions of *you know*

You know used at the textual level primarily works as a textual resource aimed at organizing the discourse and creating coherence. This use of *you know* helps the addressee to follow the conversation with all its rapid changes in topics, digressions and pauses. Scholars have presented several functions that may serve to help guide the addressee textually. Among these are “marking transitions and topic shifts” (Erman, 2001; Fox Tree and Schrock, 2002: 740), “closing a point/narrative” (Fox Tree and Schrock, 2002: 740), “introduce clarifications and qualifying information” (Holmes, 1986: 11; Fox Tree and Schrock, 2002), “marking lexical or content search” (COED, 1976; Erman, 2001), “marking false start and repair” (Holmes, 1986: 10; Müller, 2005), “marking approximations” (Fox Tree and Schrock, 2002: 737), “introducing parenthetical comments” (Erman, 2001), “highlighting particular points” (Fox Tree and Schrock, 2002: 740), and “introducing given information” (Östman, 1981; Schiffrin, 1987). In addition, some scholars claim that *you know* serve as a pause-filler (Östman, 1981) or a “verbal filler” (Brown, 1977: 107) and a staller for time (Erman, 2001: 1340). Example (2.17) and (2.18) below illustrate the use of *you know* to mark topic shift and to introduce a clarification, respectively.

2.17 *this feeling of of erm .. sort of isolation and and and just a totally erm .. unregimented society you know which had been **you know** they tried to put the values of the southern sort of fairy cultured society [mhm [in the south of England too (E78)*

2.18 *I've never been trained to do either and .. people say that to me **you know** . people who can't act or think they can't act .. say that they really think . they (E347)*

Interpersonal functions of *you know*

Functions of *you know* at the interpersonal level serve to communicate messages to the hearer that are related to the interactive situation. This may for example be to take the turn (Östman, 1981; Erman, 2001; Fox Tree and Schrock, 2002), yield the turn, (Östman, 1981; Erman, 2001), invite addressee inferences (Fox Tree and Schrock, 2002), or appeal for a certain feedback from the addressee. Such an appeal for feedback may for example plea for the hearer's understanding (Erman, 2001), comprehension (Müller, 2005) attention (Östman, 1981; Müller, 2005) or cooperation (Östman, 1981). Müller (2005: 157) also adds the functions “see the implication” and “imagine the scene” to the interpersonal level, and Östman (1981: 27) includes the function “I won't say more”. Example (2.19) below

illustrates a case where *you know* is used to take the turn, and example (2.20) shows a case where *you know* is used to indicate “see the implication”.

2.19 *I I don't need to go through that again and <A> [yeah I understand [**you know** I did that once so . er an I and I like being in town just because I'm not .. (E619)*

2.20 *the more she: was able to: . show us of herself the more: .. i= i= it was easy to see what kind of person she was and that's **you know** .. and then in the start of the second year she said that there's two kinds of people (N108)*

Qualifying functions of *you know*

Cases of *you know* functioning at the qualifying level are concerned with expressing qualifying messages to the addressee in situations where the communication is not going very well (Aijmer, 2002). Such situations may be discussions, disagreements or misunderstandings or politeness and face-saving situations. Here, *you know* may be used for emphatic reasons, to “emphasize, intensify or boost the strength of a speech act”, to gain confidence, or to express to the addressee that what is said is valid (Holmes, 1986: 8). *You know* may also be used at the qualifying level as a politeness strategy (Östman, 1981; Holmes, 1986). It is here distinguished between positive and negative politeness, where positive politeness encompasses cases of *you know* where the speaker distances himself from potential face-threatening situations, or where s/he opens up for the addressee's interpretations in order to seem less direct and assertive. Negative feedback comprises cases where the speaker uses *you know* to express uncertainty either in terms of the contents of the message or lexical knowledge (Holmes, 1986: 7). In such cases, *you know* is used to appeal for reassurance and to soften the force of the utterance, again in order to seem less assertive. Example (2.21) shows a case where the speaker uses *you know* to soften the force of the utterance (and at the same time referring to common knowledge).

2.21 *no I think I like Galway . even though it's in the western part and it's all **you know** windy and rainy and .. yeah (eh) June no it wasn't it was lovely but but I think (N4)*

Table 4 summarizes the textual, interpersonal and qualifying functions of *you know* identified in the literature.

Table 4: Textual, interpersonal and qualifying functions for *you know* in previous research

Variables	Values
Textual level	Marking transition and topic changes
	Closing off prior discourse
	Marking false start/repair
	Introducing background information
	Marking parenthetical comments
	Highlighting a particular point
	Introducing qualifying information or clarifications
Interpersonal level	Turn-taking
	Turn-yielding
	Appeal for understanding
	Plea for cooperation
	Plea for attention
	Marking approximation
	Emphasis
	Introduce given information/shared knowledge
	Invite addressee inferences
	"See the implication"
	"Imagine the scene"
"I won't say anything more"	
Qualifying level	Positive politeness: Expressing confidence concerning the addressee's relevant background knowledge or his/her experiences, attitudes, anticipated response
	Negative politeness: Expressing message-oriented uncertainty, lexical uncertainty or addressee-oriented uncertainty
	Hedging

A note on non-discourse uses of *you know*

As mentioned in section 2.1.2, items, i.e. *you know*, are not considered discourse markers if they do not form “part of the syntax of the clause”, and if they “could not be omitted” from the clause without altering the meaning of the clause or making it ungrammatical (Macaulay, 2002: 752). Such non-discourse marker uses for the DM *you know* for instance comprise instances where the marker functions as the subject and verb of a *do-you-know*-question, as illustrated by example (2.22) below, or where the verb refers to the act of *knowing* something or someone. In the present study, all such cases of *you know* are excluded from the analysis.

2.22 *do you know* why they do that I mean erm most small towns don't want people walking <?> the streets (E146)

2.3.3 *Like*

Like is classified as an information-centered discourse marker (in contrast to the addressee-centered DM *you know*) and a presentation marker (Jucker and Smith, 1998: 172, 197). Schourup describes one of its functions as serving to express a “possible unspecified minor

nonequivalence of what is said and what is meant” (1985: 42), and Hasund finds that the speaker to a certain degree distances himself from what is expressed by the utterance (2002: 130). In other words, *like* is used when the speaker wants or chooses to provide a loose fit between his/her chosen words and their intended meaning. In other words, one should not take what is said too literally when it is preceded or followed by *like*, since the words that are chosen not necessarily are identical to what the speaker has in mind.

Different scholars focus on different meanings of *like*, claiming they serve the main or core meaning of the DM. Underhill (1988) argues that *like* primarily serves to introduce quotations, and that it this function can be paraphrased with the word *say*. However, *like* in this context is not only used to refer to direct discourse, but also to inner monologue and the speaker’s feelings and attitudes. Andersen (1997; 1998) and Schourup (1985) claim that the main meaning of *like* is to indicate looseness of meaning, as described above. Underhill (1988) states that *like* mainly serves to introduce new information and that it mainly functions as a marker of focus directing the hearer’s attention towards what is new and important.

Textual functions of *like*

Scholars have identified several pragmatic functions that the discourse marker *like* may serve. Jucker and Smith (1998: 184) identify four main functions: 1) approximator (often paraphrased as “approximately” or “about”), 2) hedger (often paraphrased as “kind of”), 3) exemplifier (often paraphrased as “for example” or “for instance”), and 4) introducer of quotations (often realized by the construction BE+*like* or FEEL+*like*). Some scholars exclude quotative *like* from their analysis as they claim that this is a grammaticalized form of the marker (Müller, 2005; Blyth et al., 1990), but this function will be included in the present study’s analytical framework. Among the functions listed above, number 3), ‘introducing examples’, and number 4) ‘introducing direct speech’ are considered textual functions. Other textual function identified for *like* in the literature is ‘pause filler’ (Schourup, 1985), often occurring after questions and as an interjection, ‘introducing explanation’ (Müller, 2005), ‘hold the floor’, where *like* indicates “intended continuation” (Schourup, 1985: 54), *like* as a discourse link connecting “syntactically (sometimes even thematically) unrelated sentences” (Andersen, 2001: 255), and *like* serving to restart or repair a stopped construction (Schourup, 1985). Example (2.23) illustrates *like* used to introduce direct speech, whereas (2.24) shows *like* used to introduce an example.

- 2.23 *are you rea= are you sure because I googled it and I'm **like** okay . you know . doesn't really look different from anywhere else (N395)*
- 2.24 *you don't know any of the tasks and then you go in on the day you get a task **like** produce a erm booklet for a ten to twelve year olds about . erm learning to drive [things like that (E1311)*

Interpersonal functions of *like*

From a relevance-theoretic point of view, *like* is seen as having procedural meaning, i.e. *like* is used to express the speaker's attitude or stance towards what is being said (Helt and Foster-Cohen, 1996: 316). An example of such attitude is a case where the speaker sees the proposition as important, and therefore uses *like* to direct the hearer's attention toward it (Helt and Foster-Cohen, 1996: 316). This function will be referred to as 'attention-getter' in the present study. Another interpersonal function, as already mentioned above, occurs when *like* is used as a "loose fit between the utterance and the thought it represents" (Müller, 2005: 200), or when the speaker expresses distance from what is being said (Andersen, 1997). Such uses can be found with values referring to for example numbers, degrees or distances, and is characterized as 'approximators' (Schourup, 1985; Meehan, 1991; Jucker and Smith, 1998; Andersen, 2001). This usage of *like* is by Underhill (1988) not interpreted as discourse marker use, but rather as a grammaticalized use of *like*. Yet, this thesis will treat it as a DM. Example (2.25) and (2.26) illustrate cases where *like* is used to appeal for attention and mark approximation, respectively.

- 2.25 *they give eh everything in plastic cups so you can't smash anything or throw anything . and sometimes it's just nice **like** if you're late and you get it and you buy a drink .. and you don't want to sort of take two seconds to finish it (E848)*
- 2.26 *I tried to pull in . turn into the school (eh) . the road (eh) the car slipped . and I . I went for **like** . I'm sure fifty metres down the road (N136)*

Furthermore, Andersen (2001: 249) claims that *like* can function as a 'hesitator' indicating that what is said, or how something is said, is not necessarily expressed in the most accurate manner. As mentioned above, Underhill (1988: 238) emphasizes *like*'s function as a 'focuser', where it is used to highlight important information. This function is also underlined by Andersen (2001), Meehan (1991) and Fuller (2003), and will be referred to as 'emphasizer' in this study. *Like* is also categorized as meaning *similar to*, indicating that what is said is not necessarily an exact reflection of reality or the best possible way to explain

something (Meehan, 1991: 40). This interpersonal function is in the present study referred to as ‘hedger’, and is exemplified in extract (2.27) below.

2.27 *they get (eh) support from . different kinds of (eh) . things around Hamar and they also . get **like** (em) . paid for the children so they have . quite a bit (N149)*

Table 5 summarizes the textual, interpersonal and qualifying functions of *like* identified in the literature.

Table 5: Textual, interpersonal and qualifying functions of *like* in previous research

Variables	Values
Textual functions	Exemplifier
	Hold the floor
	introducing explanation
	Quotative <i>like</i>
	Restart/repair
	Pausal interjection
Interpersonal functions	Approximator
	Emphasizer
	Attention-getter
	Hesitator
	Hedger

A note on non-discourse marker functions of *like*

The word form *like* can serve a number of functions both pragmatically and syntactically, and some of them are not classified as discourse marker functions. The *Oxford Advanced Learner’s Dictionary* lists six different syntactic functions of *like*; **preposition**, **verb**, **conjunction**, **noun**, and **adjective**, as illustrated by example (2.28) to (2.32) below, respectively. Romaine and Lange (1991: 224) list **suffix** as an additional function, which is illustrated by example (2.33).

2.28 *She looks **nothing like** her mother.*

2.29 *She is nice. I **like** her.*

2.30 *No one sings the blues **like** she did.*

2.31 *We all have different **likes** and **dislikes**.*

2.32 *A chance to meet people of **like** mind. (OALD, 2005: 891).*

2.33 *The sculpture looked quite human-**like** (Romaine and Lange, 1991: 224).*

2.4 Factors potentially affecting learners' acquisition of discourse markers

This section will address a set of factors within the field of SLA which may serve to explain certain problems that learners encounter when acquiring a foreign language. Three factors in particular are focused on here: L1 transfer, lack of input of the target L2 features, and speaker-confidence. These are all identified as potential explanations in previous studies to how discourse markers are used differently in learner language than in a reference language variety. This list is far from complete, as there are several other potential elements that may also contribute to learners' difficulties with using discourse markers in English, such as motivation, age, and attitudes (Ortega, 2009). However, the aim of this study is not to provide a full picture of all potential reasons to why learners differ in their use of English as compared to native speakers, but rather to point to a few possible explanatory factors that can be further studied in the future.

2.4.1 L1 Transfer

L1 transfer is seen as a possible factor explaining learners' difficulties in acquiring certain aspects of a foreign language (Ortega, 2009). Transfer is in this context understood as “a function of learners' (conscious or subconscious) intuitions about how transferable certain phenomena are” (Ortega, 2009: 38). Typically, phenomena that are considered to be specific for a certain language are less likely to be transferred than phenomena considered common or ‘universal’ for two or more languages. Moreover, transfer can occur in at least two ways. Firstly, learners may transfer a certain feature from their L1 into the target language, which consequently may end up being overrepresented or misused relatively to the reference language norm. In such cases it is easy to point at the error or feature that is overrepresented, and identify the corresponding feature in the learners' L1. Another way of transferring is often referred to as ‘negative transfer’ or ‘avoidance’. In this case, learners may, due to their L1 knowledge, end up avoiding or misusing certain features of the target language (Ortega, 2009: 40), and this often leads to underrepresentation of these phenomena in learner language compared to a reference language variety. This type of transfer is more difficult to identify, as there often is no physical evidence of the error due to the omission.

As previous interlanguage studies on DMs report tendencies of L1 transfer among the learners of English (Müller, 2004: 1176; Romero Trillo, 2001; Liu, 2013), it is viable to assume that the Norwegian learners of this study may also be affected by this factor. The three discourse markers of this study would typically be perceived as ‘universal’ features that are easy to transfer, as the corresponding items frequently occur in both languages. Such corresponding items in Norwegian, possibly serving as objects of transfer, are *liksom* (*like*), *du vet* (*you know*), *du skjønner* (*you understand*), *vel* (*well*) and *jeg bare* (*I’m like, I just*).

2.4.2 Lack of Input

The Comprehensible Input Hypothesis, presented in Krashen (1985), states that learners acquire a language best by receiving what is referred to as ‘Comprehensible Input’ in the L2, i.e. messages that the learners can understand, but which also contain parts that the learners have not yet acquired. This combination of known and unknown information is referred to as ‘i+1’, and such input can come from both reading and listening (Krashen, 1991: 409). Ortega (2009: 59) also suggests input as a factor affecting how learners acquire a second language. This is further addressed in studies on DMs in particular, where learners have been found to underuse such items due to lack of and unnatural input both on an everyday basis and in classroom situations (Hellermann and Vergun, 2007: 176-177; Fung and Carter, 2007: 433). Based on this assumption, I see sufficient input as an essential factor for learners to succeed in acquiring English discourse markers.

In the acquisition of discourse markers, learners may receive such comprehensive input (‘i+1’) through, for instance, living abroad, travelling, speaking with native speakers, media, or in classroom situations where there is a particular focus on such markers or where the teacher consciously or subconsciously uses these markers when speaking in the target language. However, research reveals that teachers tend not to use DMs when speaking in a classroom situation, and that textbooks often do not direct focus toward this aspect of the English language (Hellermann and Vergun, 2007; Mukherjee and Rohrbach, 2006; Romero Trillo, 2002; Lam, 2009). As already suggested in section 1.1, this appears to be an issue in the Norwegian classroom instruction of English too. Another problem related to the acquisition of DMs, addressed by Gilquin and Paquot (2008: 52), concerns the fact that the little input foreign language learners receive in English typically is restricted to the non-native talk of the teacher and students, and the unauthentic teaching materials used in the

classroom. These gaps in the EFL education often result in, as mentioned above, little and unnatural input of the target language, which further may affect the learners' production of for example discourse markers.

2.4.3 Speaker Confidence

In her comparative study of German EFL speakers and American native speakers, Müller (2004) suggests that EFL speakers' lack of self confidence is a potential factor contributing to the learners' overrepresentation of the DM *well* (2004: 1175). This aspect is also suggested by Holmes (1986), who in her study of *you know* as used by women and men identifies cases where this DM is used to express both speaker-confidence and speaker-uncertainty. As suggested by Müller (2005), speakers may signal uncertainty or lack of confidence by using *well* to search for the right word or to "test" out words and phrases. DMs may also be used in this sense to soften a statement or an argument with the purpose of reducing the face-threat of the utterance. Ortega (2009) argues that this lack of confidence to some extent relates to, and influences, Foreign Language Anxiety, which is defined as "intense feelings of apprehension, tension, and even fear, when [learners] think of foreign languages" (Ortega, 2009: 200), and Willingness to Communicate, which concerns with how willing learners are to initiate, or continue, a conversation in a foreign language. Based on such arguments, one can expect the confidence of the learners in this study to influence their production of DMs.

3 Method

“A helluva lot of words, stored on a computer” (Leech, 1992: 106)

Although his definition of a corpus is somewhat broad and inaccurate, Leech (1992) captures the essence of what a corpus really is: an overwhelmingly large computerized database of words upon which one can study unlimited aspects of language. From a linguistic point of view, however, a corpus needs to contain a certain set of features. Ideally, it must be computer-based, include authentic texts, be representative, and be compiled for the purpose of linguistic analysis. Gries (2009) neatly captures all these criteria in his definition of a corpus presented below.

*[...] the notion of “corpus” refers to a **machine-readable** collection of (spoken or written) texts that were produced in a **natural** communicative setting and the collection of texts is compiled with the intention (1) to be **representative** and **balanced** with respect to a particular linguistic variety or register or genre and (2) to be **analyzed linguistically**. (Gries, 2009: 29)²*

Over the past few decades, since the emergence of large computerized corpora, corpus linguistics has become one of the most central methods in English language research. Although corpora, or large databases of collected words, existed before, the easy access to such computerized and potable databases of texts is what has signaled the breakthrough for corpus linguistics (Johansson, 2008: 33). This change has made corpora more accessible for researchers all over the world, which further has resulted in a boost in corpus research and an increased interest in the field. In short time, corpus linguistics has grown from being dependent on first generation corpora of relatively small sizes such as the Brown Corpus, the Lancaster-Oslo-Bergen Corpus (LOB) and the London-Lund corpus, to include large corpora of over 100 million words such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The corpora have also grown to include more diversified texts from different varieties of English, such as historical texts, texts produced by people of different ages or social backgrounds, and texts produced by learners of English (Johansson, 2008: 35). In addition to other linguistic research methods, such as introspection and elicitation, corpus linguistics has thus enriched the field as a powerful force (Halliday,

² The four features are not highlighted in the original source.

1991: 41-42) and an objective tool (Kilgarriff, 1997: 137) enabling investigations of aspects of language that have never been studied before.

The present study is a corpus-based investigation aimed to test the hypothesis that Norwegian learners of English differ in their use of the three DMs *like*, *well* and *you know* as compared to their British native-speaker peers. This approach to corpus material is classified as corpus-based, and loosely corresponds to Granger's term "hypothesis-driven" (1998b: 15), because it draws on pre-existing theories or ideas about language, and aims to "test a specific hypothesis about the nature of interlanguage" (Barlow, 2005: 344). Since this investigation compares an interlanguage variety (IV), i.e. learners' "separate linguistic system" (Selinker, 1972: 214), with a reference language variety (RLV) using Granger's (1996; 2015) Contrastive Interlanguage Analysis as framework², it is classified as belonging to the particular field of Learner Corpus Research (LCR). The following sections of this chapter will look at LCR, learner corpora and the notion of 'learners' in more detail, before entering the discussion of Contrastive Interlanguage Analysis.

3.1 Learner Corpus Research and Learner Corpora

In the early 1990s, learner corpus research was considered a "new research enterprise" (Leech, 1992). It sprung out from the field of corpus linguistics in the late 1980s or early 1990s with the purpose of contributing to research in Foreign/Second Language Acquisition (F/SLA) and the development of pedagogical tools and methods for foreign language teaching (Granger, 2002; Granger, 2008). The emergence of LCR has enriched the field of corpus linguistics by broadening the scope of studies within English language. Also, this easy access to learner data has increased the interest in learner language profoundly (Hasselgård and Johansson, 2011: 37).

3.1.1 Learner Corpora and Learner Corpus Design

As with corpora in general, learner corpora are also required to contain a certain set of features in order to qualify as a corpus. These characteristics involve a) authenticity, b) representativeness, c) balance, d) digitalization, and e) that they are compiled with the purpose of linguistic analysis. The only way in which learner corpora differ from other corpora, is that they contain data produced by learners of a language (Granger, 2008: 259). Thus, learner corpora can be defined as: "systematic collections of authentic, continuous and

contextualized language use (spoken or written) by L2 learners stored in electronic format” (Callies and Paquot, 2015: 1).

The interlanguage of learners differs greatly depending on factors such as L1 background, knowledge of other foreign languages, and level of proficiency. In order to qualify as a corpus, it is therefore desirable to collect learner data from speakers and speaker contexts based on a set of strict criteria. In this way one avoids large collections of heterogeneous learner data that cannot be considered representative for a specific L1 group or a specific learning context. Granger (2008: 264) divides these criteria into two main branches: 1) **learner variables**: variables that characterize the learners (age, gender, L2 proficiency, learning context etc.), and 2) **task variables**: variables that relate to the language situation, e.g. whether the language is spoken/written, whether it is elicited or natural occurring etc. Within these two main subdivisions, Granger (2008: 263) further divides the factors into a) **general variables** and b) **L2-specific variables**. General variables refer to factors relevant for any corpus compilation, such as age, gender, medium (spoken/written) and genre or text type. L2-specific variables, on the other hand, refer to factors only relevant for learner corpus compilations, such as learning context, L2 proficiency, the type of task learners are engaged in when producing the data (role-play, interview, picture description task etc.), and conditions that may influence learners’ L2 production and proficiency (Granger, 2008: 264). Thus, in order to compile a learner corpus that can yield valid and representative results, Granger (2008) suggests this design, as illustrated in Figure 1 below.

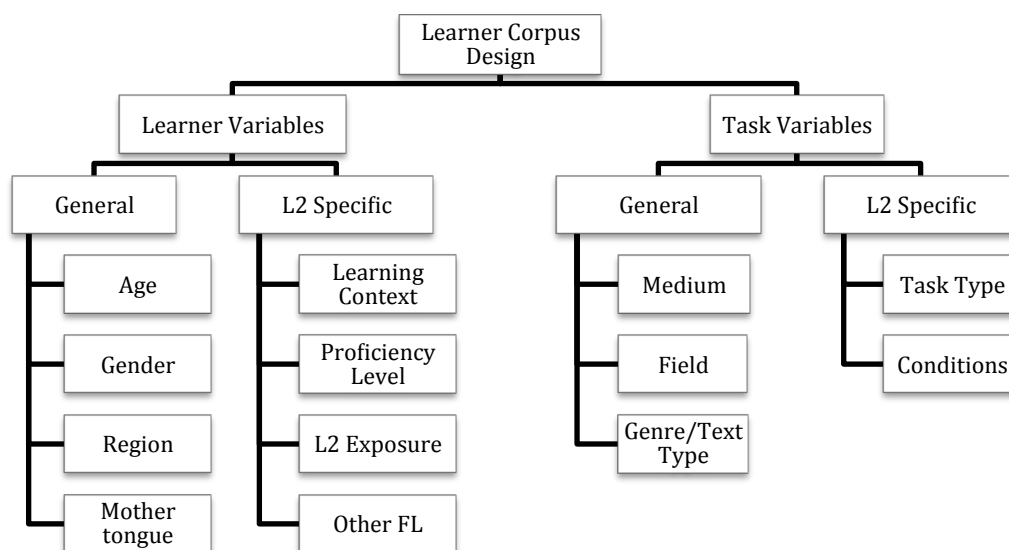


Figure 1: Learner Corpus Design (Granger, 2008: 264)

3.1.2 Learners: who are they?

It may seem straightforward to distinguish learner corpora from other corpora, but it is crucial to clarify who are considered to be *learners* in this context. If we look at the English language, its speakers can be divided into two main branches: native speakers and non-native speakers. The term ‘learners’ may potentially include all non-native speakers of English, coming from both, what Kachru’s (1985) refers to as, the “outer circle”, where “English has achieved a status of official language of education or administration” (Granger, 2008: 260), and the “expanding circle”, i.e. countries where English does not have any official status. As illustrated by Figure 2 below, the group of non-native speakers can be further subdivided into learners of English as a second language (ESL), learners of English as a foreign language (EFL) and learners of English as an official language (EOL) (Granger, 2002: 6). The term EOL comprises speakers of English in countries where English is spoken as an official language, such as India, Singapore and the Philippines. The term ESL includes those who have acquired English in a country where English is commonly spoken, such as immigrants who have moved to an English-speaking country and acquired the language there. Speakers of EFL, on the other hand, include those who have acquired English in countries where English is not commonly spoken, typically through classroom instruction (Granger, 2002: 5). This is the case for countries such as Norway, Germany and Russia. It is desirable to avoid mixing EOL, ESL and EFL, as these three varieties of non-native Englishes cannot be referred to as learners of English in the same sense. For the purpose of this study I will therefore follow Granger’s definition of learners, and describe them as “speakers who learn a language which is neither their first language nor an institutionalized additional language in the country where they live” (Granger, 2008: 260). Thus, only speakers of English as a foreign language coming from Kachru’s (1985) expanding circle are included in this definition, and EOL and ESL speakers are excluded. Figure 2 below illustrates the distinctions between the different non-native speakers of English.

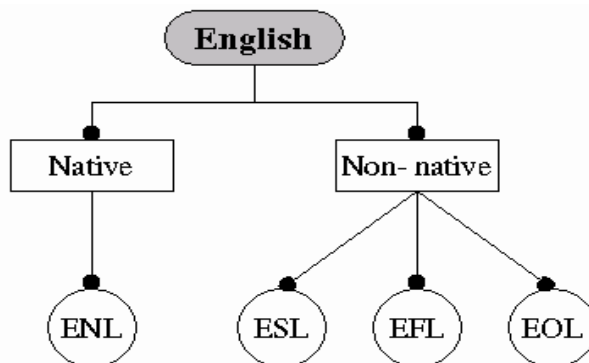


Figure 2: Varieties of English by Granger (Granger, 2002: 8)

3.2 Contrastive Interlanguage Analysis²

Contrastive Interlanguage Analysis (CIA), a comparative framework for analyzing computerized learner language, was introduced by Sylviane Granger (1996) together with the development of the International Corpus of Learner English (ICLE). This model of analysis emerged as a deviation from Contrastive Analysis (CA), which is a model involving “linguistic comparison[s] of [normally] two languages” (Hasselgård and Johansson, 2011: 38). Whereas CA is designed mainly to analyze and compare languages either by putting original texts from two different languages up against each other (OL vs. OL), or by comparing original texts in their source language with translated texts in their target language (SL vs. TL), CIA was developed as a model for analyzing and comparing varieties of the *same* language. Such analyses may either include contrasting a native language with an interlanguage (NS vs. IL), or comparing two interlanguages (IL vs. IL) (Granger, 2009: 18). Yet, Granger (1996) suggests that the CA and CIA models can be combined into what she calls the Integrated Contrastive Model, as illustrated by Figure 3 below. This model proposes that a contrastive analysis of two languages (either OL vs. OL or SL vs. TL) will form a good basis for a contrastive interlanguage analysis in that it gives the researchers predictions about what to expect from the learners of the target language in terms of L1 transfer. This idea is based on the assumption that “individuals tend to transfer the forms and meanings and the distribution of forms and meaning of their native language and culture to the foreign language and culture” (Lado, 1957: 2). Also, if we turn the model around, the CIA may be diagnostic in that it can provide an explanation of the errors detected in the CA. Such explanation can, however, only serve to form hypotheses, as there may be other factors resulting in the errors pointed out by the model too, as emphasized by (Gilquin, 2000/2001).

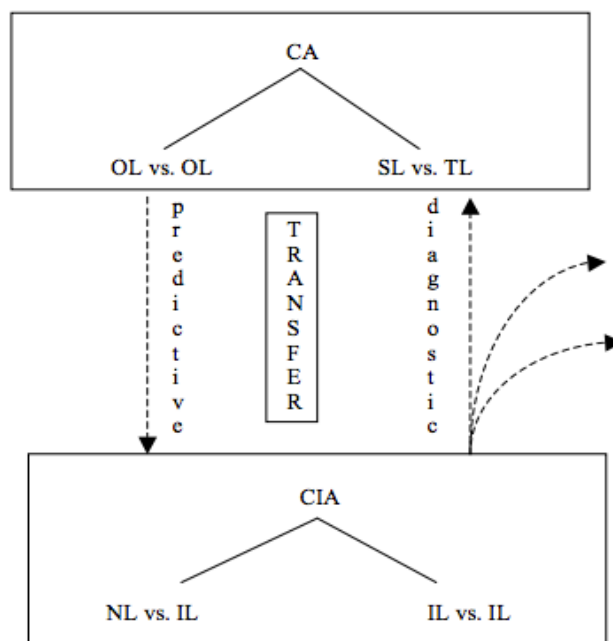


Figure 3: The Integrated Contrastive Model (Gilquin, 2000/2001: 100)

Despite its usefulness and great contribution to the field of learner language research, CIA has received much criticism on methodological and theoretical matters. One of the most criticized aspects of CIA is what Bley-Vroman (1983) and Barbieri (2010: 149-150) refer to as the ‘comparative fallacy’. This criticism is based on the argument that by comparing learner language with native speech, learner language research becomes “trapped” (Barbieri, 2010: 149-150), “hindered” (Bley-Vroman, 1983: 2), and “deficient” (Larsen-Freeman, 2014: 217) and is no longer studied in its own right (Selinker, 2014: 230). This criticism has come especially from scholars within the field of SLA who claim that learners do not necessarily aim to reach a native-like level, and who believe that interlanguage should be considered a language variety in its own right (Hunston, 2002: 211).

Another criticism of CIA is based on the argument that the division of speakers into ‘natives’ and ‘non-natives’ is inconsistent (Brutt-Griffler and Samimy, 2001: 105), as the emergence of English as a Lingua Franca has opened up for a whole new field of ‘native’ Englishes that cannot be compared with for instance British or American English in terms of proficiency and performance levels (Ferguson, 2009: 127). Tan (2005: 128) claims that this division has formed “imperialistic assumptions about the ownership of English”, and asks for a change in focus.

As a counter-argument to this criticism, Granger (2015: 14) claims that many learners do indeed strive to reach a native-like level of proficiency, and that it is natural to look at L1-L2 comparisons in research with the aim of providing pedagogical applications and knowledge that can be used for teaching English as a foreign language. She also highlights the fact that the model has never intended to place any ownership of English to any standard, but acknowledges that it is time to revise the model and make some alternations to the issues on which the criticism has shed light. As a result, Granger (2015: 17) presents the reappraised CIA², illustrated by Figure 4 below. In this developed model the term ‘Reference Language Varieties’ (RLV) is introduced as a replacement for ‘Native Language’, and the term ‘Interlanguage Variety’ (Polanyi and Scha, 1983) as a replacement for ‘Interlanguage’. This change serves as a solution to the problem with the term ‘native language’ in that the term RLV makes it clear “that the corpus does not necessarily need to represent a norm” (Granger, 2015: 17). Also, the addition of ‘varieties’ to the term ‘interlanguage’ highlights “the highly variable nature of interlanguage” (Granger, 2015: 18). Additionally, in order to avoid criticism of the terms ‘underuse’ and ‘overuse’, the more neutral words ‘overrepresentation’ and ‘underrepresentation’ are introduced. In sum, the revised model addresses most of the issues brought up in the criticisms of CIA, and appears as more neutral and less imperialistic than its ancestor. CIA², with its revised terms, is the model applied for the present study.

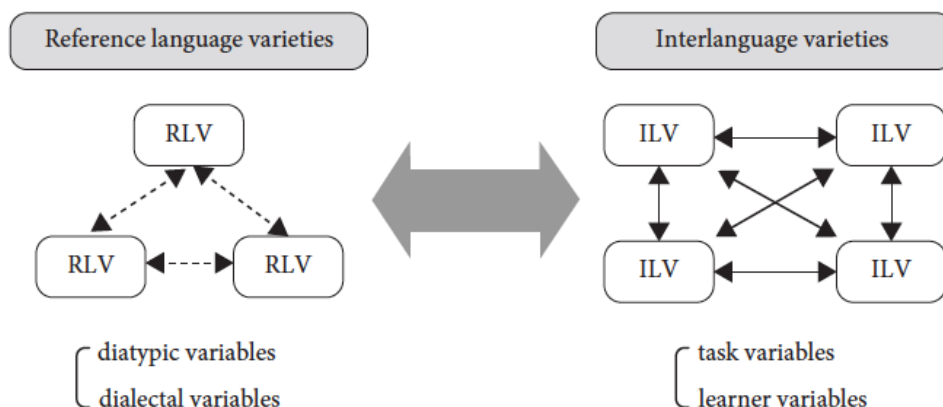


Figure 4: Contrastive Interlanguage Analysis 2 (Granger, 2015: 17)

4 Material

Although the emergence of corpora and learner corpus research has enriched the field and opened up for and increased interest in new areas of studies, there are still many issues and areas of improvement related to using this material that need consideration. This chapter will introduce the two corpora used for this study, LOCNEC and LINDSEI-NO, and discuss some methodological choices and issues related to this material. Thereafter, section 4.2 will present the analysis procedure, explaining how the data was extracted and which methodological considerations and problems I encountered during the process. The framework for the classification of the DMs will be outlined in section 4.3, followed by a brief description of the statistical calculations used for this study.

4.1 The corpora in this study

Aijmer (2002: 39) highlights the importance of the corpus data's authenticity and representativeness for a corpus study to yield valid results. By authenticity it is meant that the material has to be gathered from "the genuine communications of people going about their normal business", and not under "experimental conditions or in artificial conditions of various kinds" (Tognini-Bonelli, 2001: 55-56). This may be a challenge for compilers of learner corpora, as foreign language learners normally do not use their L2 for normal, everyday speech. Rather, the target language is typically only spoken in the classroom, or outside the classroom in informal conversation with other native speakers of learners. For practical reasons, and in order to gather enough data, the second language production is therefore often clinically elicited where the learners under certain degrees of artificial conditions are set up to speak in the target language. Consequently, one can argue that the data in learner corpora is not entirely authentic, and this may cause some people to reject the classification of learner corpora as corpora, and rather define them as 'databases' (Gilquin et al., 2010: 6). However, several scholars speak against this claim, and suggest that learner corpus data can be regarded as highly authentic despite the artificial conditions. After introducing the two corpora of this study, LOCNEC and LINDSEI-NO, a further discussion of the authenticity of the material extracted from these corpora will take place.

4.1.1 LINDSEI and LINDSEI-NO

The Louvain International Database of Spoken English Interlanguage (LINDSEI) was compiled as a spoken counterpart to the International Corpus of Learner English (ICLE), and launched at the Centre of English Corpus Linguistics at Université catholique de Louvain by Sylviane Granger in 1995. Each sub-section of the LINDSEI corpus contains data from 50 interviews of a chosen L1 group (Sinclair, 1996). At the moment of writing there are 14 officially completed sub-corpora representing 14 different L1 groups, and 7 more in progress (Gilquin, 2015). The Norwegian sub-section, LINDSEI-NO, upon which the present study is based, is considered as one of those still in progress. It is completed, but not yet officially published.

LINDSEI-NO is compiled by a team of researchers at Hedmark University of Applied Sciences (HUAS), and is designed to represent Norwegian advanced learners of English. The corpus contains 50 interviews with students of English at HUAS. Among these 50 participants, 35 are females and 15 males, and their age ranges from 19 to 47 (average age is 26). All members have Norwegian as (one of) their first language, and they have all attended primary and secondary school in Norway.

All sub-sections of the LINDSEI corpus are designed in the exact same manner in order to be fully comparable. They are built up of 50 interviews consisting of three tasks lasting for about 15 minutes, yielding approximately 2,000 words of learner language. With 50 such interviews, each sub-corpus thus contains about 100,000 words of learner language (De Cock, 2004: 227). To be specific, LINDSEI-NO contains 122,956 words. The participants are labeled ‘advanced learners of English’, but this proficiency is measured according to external rather than internal criteria, using the criterion that they are all in their third or fourth year of studying English (or related studies) at university level (Gilquin et al., 2010: 10). A drawback of measuring language proficiency through such an external factor is that studying English in the third or fourth year of University not necessarily means that the learners have reached an advanced level. Also, the level of education might differ according to which university the students attend or in which country they study. Consequently, there might be differences in the learners’ proficiency level both within and across the LINDSEI sub-corpora. Tests have been made on some of the subsections of LINDSEI to measure the level of proficiency and the degree of variation, and the participants are found to range between ‘higher intermediate’

and ‘advanced’ proficiency levels, depending to a certain degree on their L1 background. Swedish learners, for instance, score high on the proficiency measure, whereas Bulgarian learners provide much lower scores (Gilquin et al., 2010: 11). Since Swedish and Norwegian are two closely related languages with similar systems of education, I assume the Norwegian learners to be approximately at the same proficiency level as the Swedish learners, and therefore choose to stick to the label ‘advanced learners’ when referring to the proficiency level of this study’s participants.

4.1.2 LOCNEC

The Louvain Corpus of Native English Conversation (LOCNEC) was compiled according to the exact same principles as LINDSEI in order to serve as a comparable native speaker corpus (Gilquin et al., 2010: 65). Like LINDSEI-NO, it consists of 50 interviews including the same three tasks, yielding a total of 170,347 words of reference language. The participants are all British students at Lancaster University in the UK (Gilquin et al., 2010: 65). The majority of the students are undergraduates, and some are postgraduates. Their age ranges from 18 to 30, and the ratio of males and females are 2:3 (30 females and 20 males), almost identical to LINDSEI-NO (Gilquin, 2012).

4.1.3 Evaluation of comparability

Comparisons between learner corpora and native speaker corpora raise a number of methodological and theoretical issues which have to do with regional variety, text types and level of proficiency (Barlow, 2005: 345). It is essential to address and critically assess these matters when choosing which corpora to compare and when discussing the validity of results extracted from these corpora. Comparability is determined by the extent to which the two contrasted corpora are built up by the same variables, such as age, gender, medium of speech and task type (Johansson, 1998: 5; Lawson, 2001: 282). LINDSEI and LOCNEC are compiled following the exact same principles with the aim of ensuring that as many variables as possible are identical. Apart from a few differences in terms of participants’ age and the degree scheme, LOCNEC is considered highly comparable with LINDSEI (De Cock, 2004: 227). In a contrastive study of the kind performed here, it is essential to be able to point out the variables that differ between the two corpora that are compared in order to be able to discuss it as a possible explanation to the observed differences. In the present study, the only variable that differs between the two corpora is the learner variable “mother tongue”.

4.1.4 Evaluation of authenticity and representativeness

The interviews in LINDSEI-NO and LOCNEC include three tasks: i) a warm-up task where the interviewee chooses a set topic and talks freely about it for a few minutes; ii) a free discussion; and iii) a picture description task where the interviewee is asked to describe a picture with which s/he is presented³ (Gilquin et al., 2010: 3). Thus, the learner data is clinically elicited, i.e. the interviewees are presented with a specific task aimed at “getting the informant to produce data of any sort” (Corder, 1976: 69). The data material in the two corpora is therefore not entirely authentic, but collected specifically for research purposes. This has caused some researchers to question the level of authenticity data produced in such a situation can reach (Ellis and Barkhuizen, 2005; Van Lier, 1989; Davies, 1978; He and Young, 1998), arguing that several factors may interfere with, and prevent, the production of authentic data in an oral interview. For instance, factors such as the interviewer and whether s/he knows the interviewee, the setting, and the presence of a tape recorder may all contribute to the artificiality of the situation, causing the interviewee to feel uncomfortable and by that prevent him/her from producing natural, authentic speech (Gilquin et al., 2010: 13; Ellis and Barkhuizen, 2005: 33-34).

However, although many of these arguments challenge the validity of the results obtained by studying learner languages on the basis of corpora such as LINDSEI-NO and LOCNEC, one can still argue that oral interviews to a high degree elicit authentic, natural speech production. Out of several clinical elicitation instruments, oral interviews are among those eliciting the most interactional authentic language production (Ellis and Barkhuizen, 2005: 33). The tasks upon which the interview is built encourage free discussion and leave little room for planning what to say. The only exception here is the picture description task, where the learner has the picture to use as a guide for what to say. However, since it is highly difficult to collect fully authentic learner language, we will have to accept that clinically elicited data is the most authentic and representative learner data available at the time.

4.2 Extracting the data

The data from LOCNEC and LINDSEI-NO was extracted using Wordsmith Tools 6.0 (Scott, 2012). Since the corpora are not tagged, the three search strings that were used in Wordsmith were ‘like’, ‘you know’ and ‘well’. Also the contracted form ‘y’know’ was searched for, but

³ See Appendix 5 for a copy of the picture description task.

this string yielded no results. Subsequently, all hits for the three search strings not functioning as a discourse marker according to the criteria presented in section 2.1.2 were manually sorted out and discarded. The remaining hits, a total of 5,174 cases, were analyzed according to my framework of classification, which is presented in section 4.2.3 below. In ambiguous cases where it was difficult to determine the DM's function or position in the clause, listening to the original sound recordings to hear whether the intonation or force of the utterance could facilitate the decision was useful.

The searches were only performed on B-turns (the interviewees' turns). The reason for this was to make sure that only data from the groups of interest were collected. If the interviewer turns had been included too, the results would be disrupted, as the interviewers might represent other language backgrounds and proficiency levels than the two groups of interest, namely Norwegian advanced learners of English and British English speakers.

4.2.1 Statistical calculations

UCREL's Log-Likelihood and Effect Size calculator⁴ has been used to calculate statistical significance for the differences between the two corpora. Log-Likelihood (LL) as a measure of statistical significance shows to what degree we can claim that a result is not due to chance alone (Levshina, 2015: 129). This is expressed through a P-value which indicates whether the observed differences are significant or not. However, significant results may not necessarily be meaningful in the sense that the two groups do not differ greatly enough for us to reject the H₀. It is for this reason that it is important to also use the Bayes Factor (BIC)⁵ to calculate the effect size of my results. The Bayes Factor is a statistical measure that standardizes results and compare them on a common ground in order to say something about "how strongly different variables are related/associated, or how greatly groups of observations differ from one another" (Levshina, 2015: 129). In other words, effect size can tell us to what extent the relationship between two variables can be interpreted as evidence against the null hypothesis. The null hypothesis in this study will always be that there is no difference between the two speaker groups.

⁴ Accessed at <http://ucrel.lancs.ac.uk/llwizard.html>, date 04.03.2016.

⁵ Also accessed at <http://ucrel.lancs.ac.uk/llwizard.html>, date 04.03.2016.

When interpreting the Log-Likelihood value, Rayson's (1993-2014) scale of evaluation has been followed:

95th percentile; 5% level; $p < 0.05$; critical value = 3.84

99th percentile; 1% level; $p < 0.01$; critical value = 6.63

99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83

99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

When interpreting the effect size, Wilson's (2013) scale of evaluation has been followed:

0-2: Not worth more than a bare mention

2-6: Positive evidence against H_0

6-10: Strong evidence against H_0

> 10: Very strong evidence against H_0

When the BIC-value is negative, the scale indicates that the results are "in favour of" the H_0 instead of "against" it. (Wilson, 2013)

4.3 Framework of classification

In chapter 2, the previous frameworks for categorizing discourse markers by Schiffrin (2001), Redeker (1990; 1991), Aijmer (2002) and Müller (2005) are discussed. As neither of these previous frameworks were found to fully fit with the material of this study, inspiration was taken from all of them in the compilation of a new framework of classification suitable for the discourse markers *well*, *like* and *you know* as used by speaker groups investigated in this study. This framework is presented in Table 6 below. As can be read from this table, the framework is built up of six variables all consisting of a certain number of values. These six variables will be further elaborated on and exemplified below. The pragmatic functions belonging to each functional level, however, will be discussed in more detail and exemplified in chapter 5, and will therefore not be mentioned here.

Table 6: Framework of classification

Variables	Values
Discourse marker	Yes
	No
	Unknown
	A-turn
Position in clause	Pre-front field (theme)
	Post-end field (tail/afterthought)
	Wildcard/different positions (insertion)
Orientation	Prospective
	Retrospective
	Unknown
Textual functions	Transition/topic shift
	Exemplifier
	Rephrase/repair
	Move (back) to story
	Closing a point
	Introducing explanation, justification etc.
	Quotative
	Pausal interjection
	Introducing given information
Interpersonal functions	Approximator
	Attention-getter
	Searching for the right word/phrase
	Indirect/insufficient answer
	Hold the floor
	Take the turn
	Hesitator
	Thinking/considering
	Self-interruption
	Plea for understanding
	Plea for cooperation/confirmation
	Initiate a response
	"See the implication"/unfinished point
	Hedger
	Emphasizer
"Imagine the scene"	
Qualifying functions	Expressing attitudes, feelings or evaluations
	Contributing/qualifying an opinion
	Correcting a misunderstanding
	Refusing to answer a question directly
	Expressing disagreement
	Rejecting a previous statement
	Modifying a previous statement
	Confirming a previous statement

The first variable, ‘discourse marker’, is included to determine whether the target item can be considered a discourse marker or not, following the criteria presented in section 2.1.2. In short, if the item forms part of the syntax and cannot be omitted from the clause without altering the meaning or grammaticality of the utterance, it is not considered a DM. I also added the value ‘A-turn’ here because, despite filtering the data to only include B-turns, some of the DMs occurring in the search turned out to be spoken by the interviewer.

The structure and selection of the remaining five variables mainly draws on Aijmer (2002), who states that DMs can occupy several positions in a clause, be oriented either forward or backward, and serve at three different pragmatic functional levels: the textual, interpersonal and qualifying. In terms of ‘position in clause’, DMs can occupy three different positions. Whenever it occurs at the beginning of an utterance, or at the beginning of a new clause or tone unit, it is analyzed as belonging in the pre-front field. This is illustrated in example (4.1). By contrast, a DM is classified as occurring in the post-end field whenever it occurs at the end of an utterance, clause or tone unit, illustrated by example (4.2). Whenever a DM does not occupy one of these two positions, and rather occurs in the middle of an utterance, clause or tone unit, it is classified as a wildcard. This use is demonstrated by example (4.3) below. The reason why tone unit has been included in these definitions is because the DM sometimes is found to serve as a wildcard although it follows a conjunction indicating a new clause. For cases where this distinction was unclear, listening to the recorded files helped to classify the DMs’ position correctly.

- 4.1 **well** I really wanted to see it because erm a lot of people had said it was really good (E518)
- 4.2 the main thing was to make them feel normal . **you know** (E409)
- 4.3 and and then just little things **like** the cars that they drove around the locals sort of the little Yugo cars you know really [[like er in Germany in East (E1367)

The third variable, ‘orientation’, contains three values: prospective, retrospective and unknown. A DM is considered to be prospective whenever it points forward in the clause or utterance to some new information that will be introduced, as in example (4.1). When it, on the other hand, points backward to old information that has already been introduced or is assumed to be common knowledge to the hearer, as in example (4.2), it is classified as being retrospective. Retrospective DMs typically only occur in the post-end field, whereas prospective DMs can occupy both the pre-front and wildcard position.

The remaining three variables, ‘textual’, ‘interpersonal’ and ‘qualifying functions’, all contain potential values of pragmatic functions assigned to the three DMs *like*, *well* and *you know*. The division of these values into the three functional categories is mainly based on the above-mentioned four scholars’ work. The textual level, for example, includes elements taken from Schiffrin’s (1987) ideational structure and Redeker’s (1991) ‘ideational’, ‘rhetorical’ and ‘sequential structure’, and involves cases where the DMs are mainly used to

create textual cohesion, i.e. helping the hearer to understand how one sequence of an utterance relates to another. This can for instance include cases where the DM is used to introduce an example, as shown in example (4.4) below.

4.4 *cos when they're kids there's so much . things to learn **like** how to . (eh) relate to toher people and . like basic stuff* (N321)

Following Aijmer (2002) and Müller (2005), DMs are in this study classified as functioning at the interpersonal level whenever they serve to signal messages to the hearer that are related to the organization of the conversation and to the meaning that lies behind the utterance. This use is for instance found when a DM serves to indicate to the hearer that the speaker wishes to hold the floor, or that what follows needs to be stressed for a particular reason, as illustrated by examples (4.5) and (4.6) below, respectively.

4.5 *if you wanted to stay up late at night you would be too tired to: <A> [mhm [to study the next day so **well** [sometimes [oh erm ..* (E812)

4.6 *cos like getting up to get the bus [yes [or to hitch on is . **you know** horrible* (E279)

The 'qualifying' level includes cases of DMs commenting on speaker-hearer relations that lie at a higher level than for the interpersonal functions. Although interpersonal functions also comment on the relationship between the two interlocutors, they are mainly centered on the meaning of the utterance. Qualifying functions, on the other hand, are more related to communicative competence in the sense of the appropriateness of the utterance, politeness, face-saving situations, disagreements etc. Example (4.7) demonstrates the use of the DM *well* to correct a misunderstanding. This functional category is similar to Schiffrin's (1987) Action Structure and Aijmer's (2002) category sharing the same name.

4.7 <A> *well there is no computer department . **well** I I there is but I had a little laugh because when I went to Lillehammer* (N116)

In the present study, all DMs are classified as potentially serving within all of these functional levels (textual, interpersonal and qualifying) simultaneously. Depending on whether a DM serves within one, two or three of the functional levels at the same time, its function is referred to as being single, dual or triple. For example, if the DM *like* serves at the textual level as an 'exemplifier' and at the interpersonal level as an 'approximator'

simultaneously, this will be referred to as the dual textual and interpersonal functions ‘exemplifier’ and ‘approximator’. If *like* serves the textual function ‘pausal interjection’, the interpersonal function ‘searching for the right word/phrase’, and the qualifying function ‘negative politeness’, as in example (4.10) below, this usage will be referred to as the triple textual, interpersonal and qualifying functions ‘pausal interjection’, ‘searching for the right word/phrase’ and ‘negative politeness’.

4.10 *when you come from Europe and when yo= travelling to Thailand you're like a . you're like a king .. you know the food is **like** .. (eh) you pay . like . how can I say it you pay like (eh) it's almost . (eh) free . for us (eh) and also the people .. (N288)*

Due to this multi-functional approach to the DMs in this study, the numbers in the tables summing up the results for the functional analysis of each DM (Table 11, Table 16, and Table 21) do not add up to the total number of occurrences for each corpus. They rather indicate how many instances of each DM out of the total occurrences served within the three functional levels in the discourse.

5 Analysis and Results

The following chapter provides a description and an analysis of the use of the discourse markers *like*, *well* and *you know* in representative and comparable samples of a British reference language variety (RLV) and a Norwegian interlanguage variety (IL) of spoken English. The chapter is mainly divided into two parts: a preliminary frequency analysis and an in-depth frequency and functional analysis of each of the three DMs. Starting with the big picture, section 5.1 will present the results from the preliminary frequency analysis of the overall usage of the three discourse markers in the two corpora, where some general tendencies will be pointed out and discussed. Then, the three sections 5.2, 5.3 and 5.4 each present the frequency and functional analysis of the discourse markers *like*, *well* and *you know*, respectively, including a deeper analysis of the pragmatic functions found to be most significantly underrepresented in the learner-material. Both parts of the analysis are quantified in nature, as I am counting the number of instances for each function and comparing the numbers, but the preceding act of analyzing and identifying the different functions leading up to the quantitative results in the latter part is considered highly qualitative. The frequencies of use by the Norwegian learners (NLEs) and British English Speakers (BESs) have been tested for statistical significance using the log-likelihood and effect size calculator⁶, and this is referred to where necessary.

5.1 Preliminary frequency analysis

In this preliminary quantitative analysis the overall use of the three discourse markers *like*, *well* and *you know* in LOCNEC and LINDSEI-NO is compared to detect potential over- or underrepresentation in the NLEs' use. With these results I hope to answer whether there is a difference between the NLEs and BLEs' use of the DMs *like*, *you know* and *well*.

This analysis is inspired by previous studies that report finding both overrepresentation (Müller, 2004; Aijmer, 2011; Unaldi, 2013; Castele and Collewaert, 2013) and underrepresentation (Müller, 2005; Fung and Carter, 2007) of discourse markers in learner language, and which suggest that native speakers tend to use DMs “for a wider variety of pragmatic functions” than learners of English (Fung and Carter, 2007: 410). See section 2.1 for a discussion of previous research. Based on these earlier studies, I have hypothesized that

⁶ Accessed at <http://ucrel.lancs.ac.uk/llwizard.html>, date 04.03.2016. See also section 4.3.1.

the Norwegian learners will differ from the British RLV speakers in their use of the three discourse markers under investigation due to the lack of input of such phenomena in the teaching of English in Norway. Surprisingly, this analysis reveals an overall statistically significant underrepresentation of all three markers in the Norwegian learner data.

5.1.1 Corpus-Driven Frequency Analysis and Quantitative CIA²

Table 7 below gives an overview of the total number of occurrences in LOCNEC and LINDSEI-NO for the search strings ‘like’, ‘well’ and ‘you know’, including instances of non-discourse markers. As can be read from the table, this quick corpus search shows a clear tendency of underrepresentation of these items among the NLEs as compared to the BESs.

Table 7: LOCNEC and LINDSEI-NO total number of occurrences (including non-discourse markers)

DISCOURSE MARKERS	LOCNEC B-TURNS		LINDSEI B-TURNS		LOG-LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
LIKE	1761	103.4	887	72.1	79.07	< 0.0001	66.48
WELL	1158	68	531	43.2	78.69	< 0.0001	66.11
YOU KNOW	690	40.5	147	11.9	227.54	< 0.0001	214.95
TOTAL	3609	211.9	1565	127.2	300.43	< 0.0001	287.84

However, as this data contains noise in the form of irrelevant units, a manual analysis was conducted where all non-discourse marker uses were discarded. Such non-discourse marker cases were similar to examples 2.14-2.16 for *well* (section 2.3.1), example 2.22 for *you know* (section 2.3.1), and examples 2.28-2.32 and 2.34 for *like* (section 2.3.3). The final results for this analysis, after all such cases were removed, are presented in Table 8 below.

Table 8: LOCNEC and LINDSEI-NO discourse marker occurrences only

DISCOURSE MARKERS	LOCNEC		LINDSEI-NO		LOG-LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
LIKE	794	46.6	400	32.5	35.62	< 0.0001	23.03
WELL	681	40	319	25.9	42.51	< 0.0001	29.92
YOU KNOW	632	37.1	142	11.5	195.94	< 0.0001	183.35
TOTAL	2107	123.7	861	69.9	211.95	< 0.0001	199.36

Figure 5 gives an illustration of the findings presented in Table 8, making it easier to compare the two speaker groups. The frequency numbers for this figure are presented in relative frequency per 10,000 words (pttw).

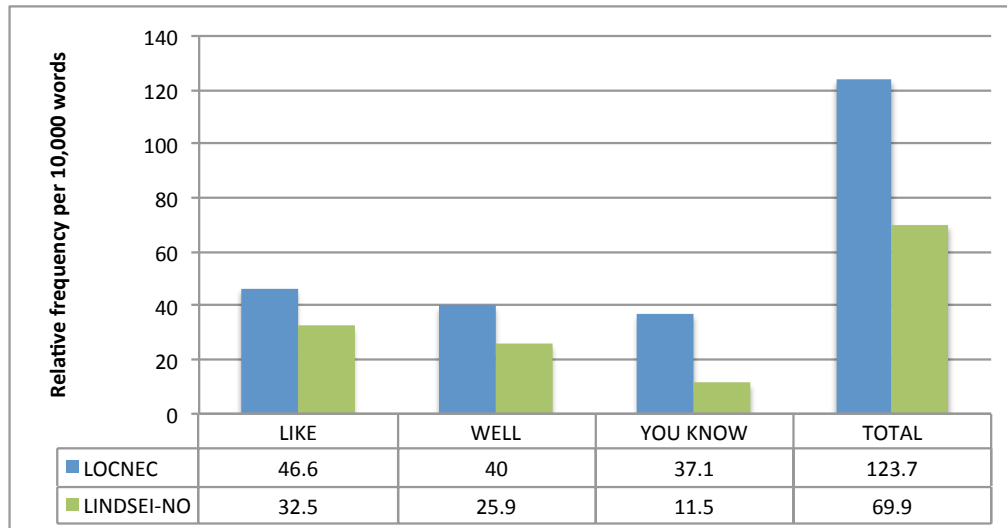


Figure 5: LOCNEC and LINDSEI-NO discourse marker occurrences only

As seen, there is an underrepresentation of all the discourse markers in LINDSEI-NO as compared to LOCNEC. As shown in Table 8, for all three DMs and for their use overall, this underrepresentation is highly significant⁷, and the effect size value tells us that these results are meaningful and serve as strong evidence against a null-hypothesis (H0) saying that the two groups use DMs in the same way. In other words, the above results serve as strong evidence that the NLEs and BESs differ in their use of the discourse markers *like*, *well* and *you know*. The effect size will mainly be commented on when it rejects a statistical significance and rather suggests that the results serve as evidence in favor of the H0.

5.1.2 Individual differences within LOCNEC and LINDSEI

The above results suggest that all NLEs underrepresent the three DMs to the same extent. By calculating the average frequency per speaker, one would for instance expect *like* to occur approximately 15.88 times in each interview in LOCNEC and 8 times in LINDSEI-NO. For *well*, one would expect the DM to occur 13.62 times per interview in LOCNEC and 6.38 times in LINDSEI-NO, and for *you know* it is 12.64 in LOCNEC but only 2.84 in LINDSEI-NO. However, this is just a calculation of average frequency based on the total number of

⁷ The difference between LOCNEC and LINDSEI-NO is statistically significant with more than 99.99% certainty.

occurrences for each marker, and individual differences within each corpus have not been taken into consideration. Figure 6 and 7 below are therefore included to illustrate that such individual variation within the two speaker groups does exist. These two figures present five frequency intervals, and show how many speakers use the three DMs within each of these intervals in the two corpora.

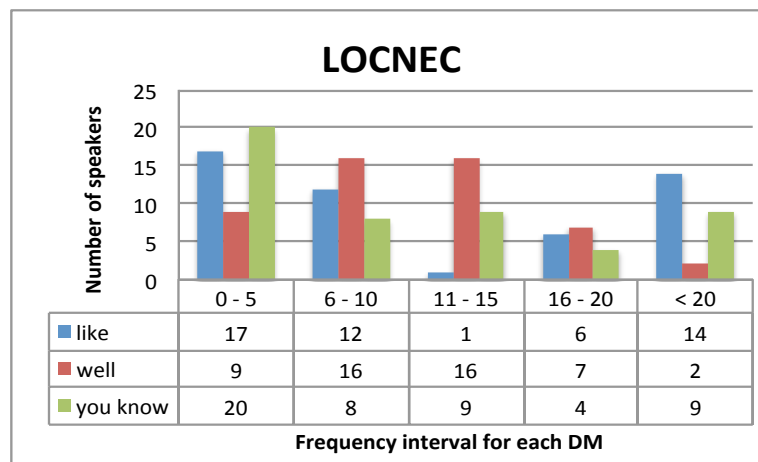


Figure 6: The frequency of *like*, *well* and *you know* for each speaker in LOCNEC

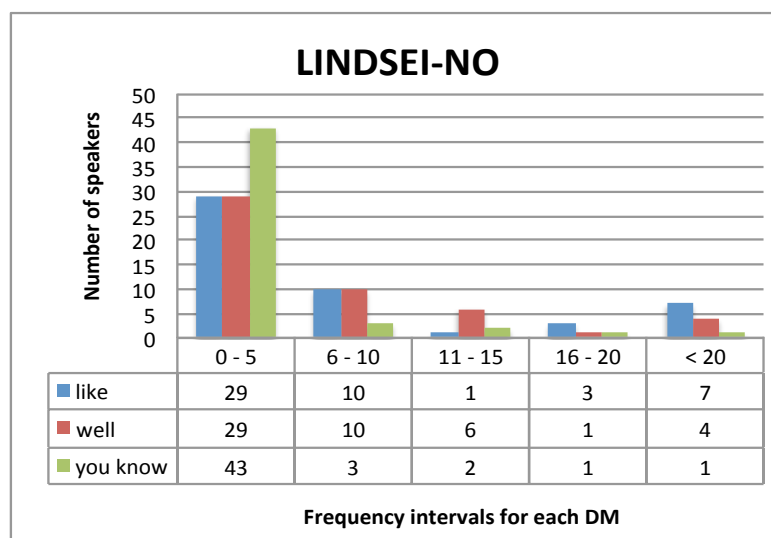


Figure 7: The frequency of *like*, *well* and *you know* for each speaker in LINDSEI-NO

Figure 6 and 7 clearly show that there is great internal variation within each corpus, and that the average frequency does not always mirror the reality. For instance, we find that 29/50 NLEs use *like* less than six times during the entire interview, although the average number suggests that most speakers use the marker 8 times per interview. This indicates that there are a few speakers who contribute to a higher value, such as speaker 31⁸, for example, who uses *like* a total of 69 times and *you know* 26 times during the interview. Speaker 45, who uses *like*

⁸ See appendix 1 and 2 for LINDSEI-NO and LOCNEC metadata.

30 times and *well* 34 times, serves as another example of such a case. The same tendencies are found for the speakers in LOCNEC, although the frequency for each DM here seems slightly more evenly distributed, according to Figure 6.

This observed discrepancy within the two speaker groups makes it difficult to generalize and categorize the speakers' usage as one unified tendency. More sophisticated statistical measurements would have to be employed to map each speaker's detailed use of the three DMs, but this lies beyond the scope of this thesis.

5.2 *Like*

Building on the preliminary frequency analysis, which has revealed that the NLEs' display an overall underrepresentation of the three DMs, this sub-section examines *like* in more detail to investigate why this underrepresentation occurs and whether the result is present for all pragmatic functions. In order to explore these questions, a functional analysis of *like* is conducted. By categorizing each usage of the three DMs according to the framework of classification presented in section 4.3, it becomes possible to discover exactly where the underrepresentation takes place, and further detect any misuses made by the learners.

This section begins with a presentation of the results from the quantitative analysis of the NLEs' and BESSs' use of *like* in terms of utterance position and orientation. Subsequently, the functional analysis, where all instances of this *like* are categorized in terms of textual, interpersonal and qualifying functions, is presented and discussed.

5.2.1 Frequency analysis

Overall frequency

As discussed in section 5.1, the overall frequency analysis uncovers an unexpected underrepresentation of *like* by the Norwegian learners as compared to the British English speakers. The DM *like* is identified a total of 794 times in LOCNEC (46.6 pttw), whereas it only occurs 400 times in LINDSEI-NO (32.5 pttw). The LL-value indicates that this discrepancy is highly significant⁹.

⁹ More than 99.99% certain.

Utterance position of *like*

In the examination of the 1194 cases of *like* in both corpora, it is observed that when occurring in the pre-front field, *like* tends to function as a theme introducing new information. This position is typically found with the textual functions ‘exemplifier’, ‘introducing a shift in topic’, ‘introducing explanation’, ‘introducing further specification’, ‘quotative *like*’ and ‘rephrase/repair’, and for the interpersonal functions ‘attention-getter’, ‘emphasizer’ and ‘hedger’. When occurring as a wildcard, *like* functions as a random insertion that can occur anywhere in the utterance except from at the beginning or the end. This position is mostly associated with the interpersonal functions ‘approximator’, ‘emphasizer’, ‘hedger’ and ‘searching for the right word/phrase’, but also with the textual functions ‘exemplifier’, ‘introducing explanation’, ‘pausal interjection’ and ‘quotative *like*’. Since no cases of the DM *like* occurring in the post-end field have been identified, this utterance position has been removed from the analysis. Table 9 and Figure 8 present the results and comparison of *like*’s utterance position in the two corpora.

Table 9: Utterance position of *like* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Pre-front field	341	20	185	15	10.03	< 0.01	-2.56
Wildcard	453	26.6	215	17.5	26.77	< 0.0001	14.18
TOTAL	794	46.6	400	32.5	35.62	< 0.0001	23.03

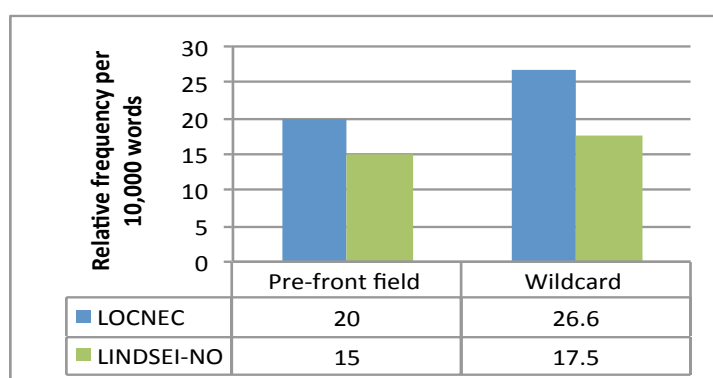


Figure 8: Utterance position of *like* in LOCNEC and LINDSEI-NO

As Table 9 and Figure 8 display, both speaker groups demonstrate the same tendency of using *like* as a wildcard more often than as a theme occurring in the pre-front field. Yet, a

significant underrepresentation of *like* by the Norwegian learners is identified both within the pre-front field¹⁰ and as a wildcard¹¹ as compared to the British native speakers. The effect size value for the results of *like* occurring as a wildcard further supports this significance by showing that the result serves as strong evidence against the H0, but for *like* occurring in the pre-front field, the effect size value indicates that the two groups do not differ greatly enough for us to reject the H0.

Example (5.1) below illustrates the use of *like* as a wildcard serving the dual textual and interpersonal functions ‘exemplifier’ and ‘hedger’. This is a good example of a discourse marker serving both a textual and interpersonal function simultaneously, as discussed in section 4.3.

5.1 *too warm apparently it's a lot a lot poorer down there* <A> *yeah that's true*
 *have to be and you have to be quite careful with **like** southern dialects as well*
<A> *mm I didn't realise there's a lot of dialects mm of course there's (E1546)*

Example (5.2) illustrates the use of *like* occurring in the pre-front field and serving the single textual function ‘quotative *like*’.

5.2 *so . when I got to Sydney airport I was **like** okay .. I'm alone in the world what to do: and I got really scared because my: th= the driver that was gonna pick me up (N796)*

Orientation of *like*

When categorizing *like* in terms of orientation, the following criteria have been followed: if the discourse marker points forward to some new information that is to come, it is considered to be prospective. If, on the other hand, it points backwards to old information that is already mentioned, it is classified as retrospective. There are no occurrences of retrospective *like* in the material, so this variable has been removed from the table and figure below. Table 10 presents the results from the analysis of *like*'s orientation in LOCNEC and LINDSEI-NO, and Figure 9 gives an illustration of the difference between the two English speaker varieties.

¹⁰ More than 99% certain.

¹¹ More than 99,99% certain.

Table 10: Orientation of *like* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Prospective	794	46.6	400	32.5	35.62	< 0.0001	23.03

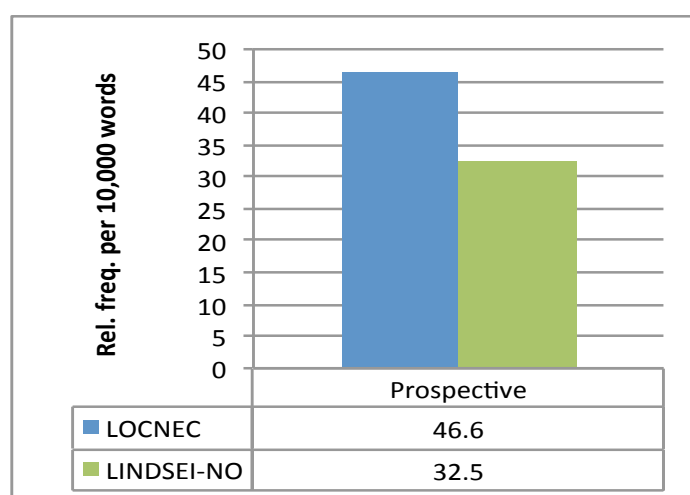


Figure 9: Orientation of *like* in LOCNEC and LINDSEI-NO

The fact that *like* only occurs with prospective orientation in both corpora further strengthens the argument that both speaker groups demonstrate similar usage of the DM *like*. However, there is a statistically significant underrepresentation¹² demonstrated by the learners. Example (5.3) below illustrates the use of *like* with prospective orientation. In this example, *like* serves the single textual function ‘introducing further specification’.

5.3 really really old horror movies . The Cabinet of Dr. Caligari and . others well yeah . (em) . well they're real special (eh) **like** in a g= . almost gothic style .. because (eh) the filming is done in .. well . every other angle than . a straight . (N93)

So far the findings show a highly significant underrepresentation of *like* by the NLEs as compared to the BESs. When looking closer into the orientation and positions in which *like* is used in the utterance, LOCNEC and LINDSEI-NO are found to show quite similar usage. Both speaker groups use *like* only prospectively, and both use *like* slightly more often as a wildcard than within the pre-front field as a theme.

¹² More than 99.99% certain.

These tendencies noted for both speaker groups may indicate that although the DM *like* is underrepresented by the NLEs, the learners still use it for the same pragmatic functions as the BESs. In order to investigate whether this assumption is valid, an in-debt qualitative study of the different functions of *like* in each utterance is essential.

5.2.2 Functional analysis

For this analysis, every case of *like* in LOCNEC and LINDSEI-NO is categorized according to their textual, interpersonal and qualifying functions. The result of this analysis is presented in Table 11 below. The presentation and discussion of these results is divided into three parts, each discussing one of these functional levels. Every section begins with addressing the general results of the use of *like* functioning at these three levels, before looking closer into two or three functions used differently by NLEs as compared to the BESs.

Table 11: Textual, interpersonal and qualifying functions of *like* in LOCNEC and LINDSEI-NO

		LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	EFFECT SIZE
		Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
TEXTUAL FUNCTIONS	Transition/ topic shift	39	2.3	2	0.7	29.88	< 0.0001	17.29
	Introducing explanation	127	7.5	54	4.3	11.29	< 0.001	-1.3
	Introducing further specification	31	1.9	28	2.3	0.74	n.s.	-11.85
	Exemplifier	98	5.8	57	4.6	1.71	n.s.	-10.88
	Quotative 'like'	49	2.9	71	5.8	14.4	< 0.001	1.81
	Pausal interjection	86	5	45	3.7	3.15	n.s.	-9.44
	Rephrase/repair	39	2.3	18	1.5	2.58	n.s.	-10.01
	TOTAL	469	27.7	275	22.9	7.61	< 0.05	-4.98
INTERPERSONAL FUNCTIONS	Approximator	67	3.9	51	4.1	0.08	n.s.	-12.51
	Emphasizer	85	5	45	3.7	2.91	n.s.	-9.68
	Hedger	220	12.9	60	4.8	55.44	< 0.0001	39.85
	Attention-getter	121	7.1	24	2	43.01	< 0.0001	30.51
	Searching for the right word/phrase	58	3.4	43	3.5	0.02	n.s.	-12.57
	Unfinished point ("and so on")	17	1	6	0.5	2.5	n.s.	-10.08
	Hold the floor	20	1.2	4	0.3	7.06	< 0.05	-5.53
	Other	7	0.5	3	0.3	0.61	n.s.	-11.98
	TOTAL	595	35	236	19.2	65.26	< 0.0001	52.67
QUALIFYING FUNCTIONS	Expressing attitudes, feelings or evaluations	22	1.3	31	2.5	5.87	< 0.05	-6.72
	Other	5	0.3	8	0.6	2.02	n.s.	-10.57
	TOTAL	27	1.6	39	3.1	7.85	< 0.05	-4.74

Textual functions of *like*

The textual functions of *like* identified in LOCNEC and LINDSEI-NO are ‘rephrase/repair’ and ‘pausal interjection’, following Schourup’s (1985) classification, ‘quotative *like*’ and ‘exemplifier’, following Jucker and Smith (1998), and ‘introducing explanation’, following Müller (2005). In addition to these functions, several instances of *like* serving to introduce further specification are also discovered and therefore added to the framework of classification. Table 12 below provides an overview of all textual functions of *like* occurring in LOCNEC and LINDSEI-NO with a short definition and example for each.

Table 12: Definitions and examples of the textual functions of *like*

Textual functions	Definition	Example
Introducing explanation	When <i>like</i> is used to explain an argument, statement or a word already introduced to the hearer.	<i>yeah and like if you're under twenty-five you got more points (E74)</i>
Exemplifier	When <i>like</i> is used to introduce an example. Can be paraphrased as "for example". Note that when <i>like</i> forms part of the syntax and cannot be omitted, such as in "A critic like you", it is classified as the prepositional meaning of <i>like</i> and thus excluded from the analysis.	<i>I get s= quite = quite a lot of time to knit and things there and also: . like in breaks . when I'm at school and . things like that (N11)</i>
Pausal interjection	When <i>like</i> is used to fill or indicate a pause in speech.	<i>so we lived in cabins . like (eh) and all the cabins were named Bergen (N41)</i>
Quotative <i>like</i>	When <i>like</i> is used to introduce direct speech, thoughts or feelings.	<i>all my friends were just like wow what a film I mean and I I suppose (E698)</i>
Rephrase/repair	When <i>like</i> is used to start over again after an immediate stop in the middle of an utterance.	<i>you know but there like some of the places it was like thirty pounds just (E1325)</i>
Transition/topic shift	When <i>like</i> is used to indicate a transition or shift in topic.	<i>that is: like I don't find it that good here as my old town (N289)</i>
Introducing further specification	When <i>like</i> is used to introduce more information about a topic already known to the hearer.	<i>I went to school like (eh) a folkehøyskole and (N550)</i>

Based on the results presented in Table 11, Figure 10 below offers a neat illustration of the differences between the NLEs and BESs for each textual function of *like*, presented in relative frequencies per 10,000 words (pttw).

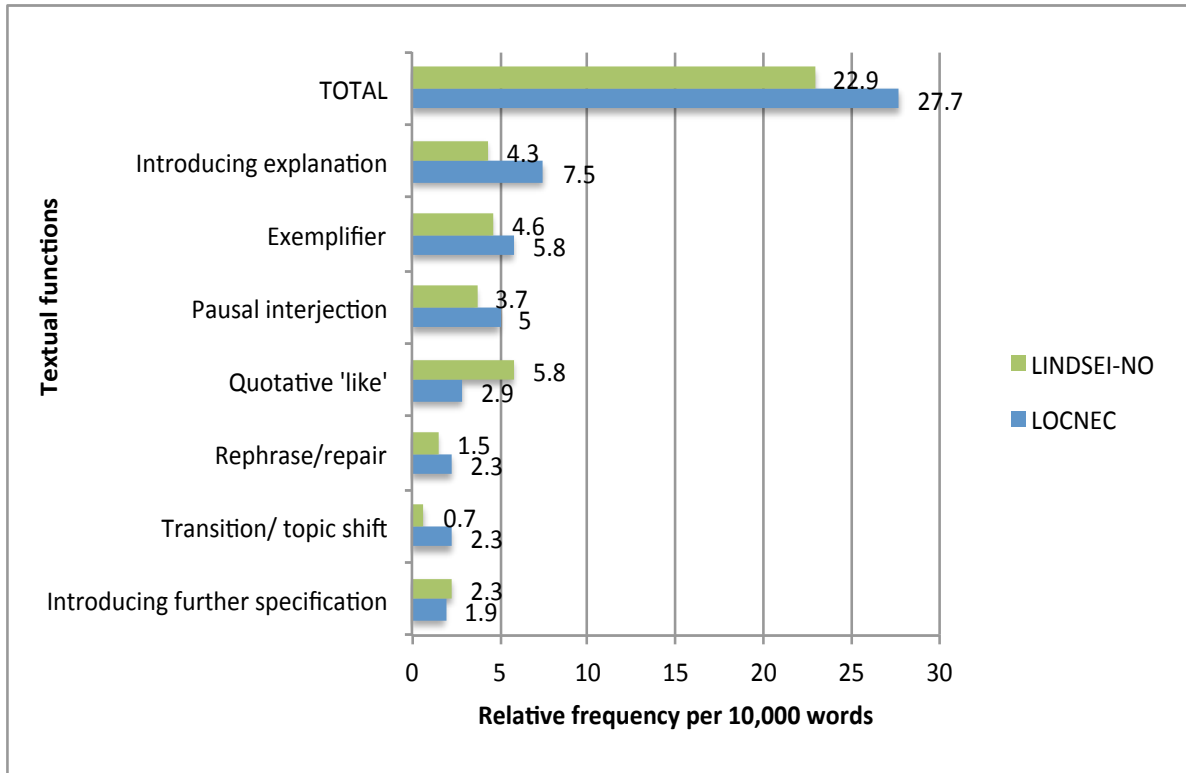


Figure 10: Textual functions of *like* in LOCNEC and LINDSEI-NO

As illustrated by Figure 10, there is an underrepresentation of textual *like* in general by the NLEs as compared to the BESs. The statistical calculations presented in Table 11 further underline that this result is significant¹³, but the effect size value tells us that the difference is not strong enough for it to serve as evidence against the H0. In other words, although the LL-value suggests that these results are statistically significant, the effect size requests more data in order to be able to fully trust the results to be significant.

Table 11 shows that *like* is underrepresented in almost every function by the NLEs, apart from ‘introducing further specification’, which yields quite similar frequencies for both speaker groups, and ‘quotative *like*’, which surprisingly is used more frequently by the NLEs than the BESs. The latter finding contrasts with Müller (2005), who discovers an underuse of ‘quotative *like*’ among German learners of English.

Figure 10 shows that LOCNEC has ‘introducing explanation’ as the most frequent textual function, followed by ‘exemplifier’, ‘pausal interjection’, and ‘quotative *like*’. In LINDSEI-NO, on the other hand, *like* is used most frequently for the textual function ‘quotative *like*’,

¹³ More than 95% certain.

followed by ‘exemplifier’, ‘introducing explanation’ and ‘pausal interjection’. These results differ from Müller (2005), who finds that ‘quotative *like*’ is the second biggest textual function among the American native speakers, and the fourth biggest among the German learners, and that ‘introducing explanation’ is the third biggest function among both the Americans and the Germans.

The least common textual function in LOCNEC is ‘introducing further specification’, whereas ‘transition/topic shift’ is the least common function in LINDSEI-NO. As for where LOCNEC and LINDSEI-NO differ the most, according to the statistical calculations presented in Table 11, ‘transition/topic shift’, where a significant underrepresentation is identified by the NLEs, and ‘quotative *like*’, which is the only textual function found to be overrepresented by the learners, stand out. These two functions will be subject to a closer examination in the two following sections.

Transition/ Topic Shift:

As Table 11 and Figure 10 above show, *like* functioning to indicate a transition or a shift in topic is underrepresented in LINDSEI-NO as compared to LOCNEC. Surprisingly, only two instances of ‘transition/topic shift’ are found in LINDSEI-NO, and these are presented in the two examples below. Example (5.6) shows a case where the speaker is talking generally about term papers and then uses *like* to change the topic to a more detailed elaboration of one of the papers. In example (5.7), the speaker is first talking about where he is from, before using *like* to change the topic to a discussion of where the skiing facilities are best.

5.6 do the Nordic part . at some point I have two term papers actually one on (em) . linguistics . and one on . literature and . **like** I have to do .. (eh) I was going to (em) .. (eh) do a term paper on (eh) Jan Erik Vold . and his: prose .. (em) for the (N21)

5.7 here well not o= cross-country skiing I don't know I'm I'm not from here I'm from another town in Norway . (eh) that is: **like** I don't find it that good here as my old town to cross-country ski and I know they don't have many pistes or (N289)

When investigating the British RLV cases of *like* as a transition-marker more closely, I notice that this use of the DM repeatedly serves the interpersonal function ‘attention-getter’ simultaneously, which is used for the purpose of appealing for attention or to indicate that the hearer needs to listen because what follows is important. This dual textual and interpersonal functional use is illustrated by example (5.8) below, where the speaker is talking about an

impressive and extraordinary event and indicates to the hearer that he needs to pay attention because an important point in the story is about to be told.

5.8 *on the track in front of us oh they stopped and turned to charge the minibus [and the [oh the driver turned it round and **like** you've never seen [anything move so quickly in your life [and you said it wasn't dangerous but and he turned this mini (E50)*

As becomes clear from the two examples above, such dual functional use of *like* is absent in LINDSEI-NO, indicating that the Norwegian learners have not discovered this typical use of the DM. A possible explanation for this underrepresentation can be lack of input. This use of *like*, as most uses of DMs, is considered typically spoken and colloquial, and it is therefore not likely that this function has been given any profound attention in any textbook or classroom situation. Rather, one can assume that those speakers who have acquired this use of the DM have spent some time abroad or gained input through other means such as media or English-speaking relatives and friends. This assumption will be further discussed in section 5.5.

Quotative 'like':

As mentioned before, 'quotative *like*' is the only textual function that is overrepresented in the Norwegian learners' production of *like* as compared to the British English speakers. A closer examination of the different cases of *like* serving this function in LOCNEC and LINDSEI-NO reveals that the two speaker groups mostly display similar usage. They both use 'quotative *like*' with prospective orientation and in the pre-front field only, and some form of the verb *to be* typically directly precedes the DM. However, some differences are particularly marked. For instance, the BESs tend to use the DM phrase *just like* (n=13, 0.7 pttw), whereas this usage only occurs once in LINDSEI-NO (0.1 pttw). The NLEs, on the other hand, are found to use the phrase *go like* (n=2, 0.2 pttw), which is absent in LOCNEC. It also appears that the NLEs use 'quotative *like*' to avoid complicated explanations more often than the BESs (LOCNEC n=10, 0.5 pttw; LINDSEI-NO n=17, 1.3 pttw). It is easier to quote what someone would say in a given situation than to explain the circumstances using difficult words and complex syntactic structures. In example (5.9) below, the speaker is trying to explain the differences between a given country and Norway. The speaker's point is that people from this other country display a more positive and relaxed attitude towards unforeseen situations and changes, whereas Norwegians would react in a pessimistic and

stressful way to the same incident. Instead of explaining this explicitly, the Norwegian speaker finds it easier to quote how a Norwegian would react.

5.9 (eh) even when you talk to people they're more like everything's gonna be okay we can fix this in Norway everybody would be **like** no no no this is not gonna be okay the world is gonna end . so it's just . more relaxed you don't get the shoulders up a (N28)

This approach of using ‘quotative *like*’ to avoid complex explanations of feelings and arguments involving advanced syntactic formulations may serve as an explanation to why the this function is overrepresented in LINDSEI-NO. Moreover, this overrepresentation can also be a result of more input of this function than of the other functions of *like*. Although ‘quotative *like*’ is highly colloquial in nature, and therefore not likely to have been given any attention in textbooks and classrooms, it is likely that learners get much input of this use through reality shows, movies or social media. A third option is that Norwegians transfer the equivalent phrase *jeg bare*, which is an informal way to introduce direct speech, thoughts or feelings in Norwegian.

Interpersonal functions of *like*

The interpersonal functions of *like* identified in LOCNEC and LINDSEI are ‘hedger’, following Andersen (2001), ‘attention-getter’, following Underhill (1988), Meehan (1991), Andersen (2001) and Fuller (2003), ‘approximator’, following Jucker and Smith (1998), ‘emphasizer’, following Helt and Foster-Cohen (1996), ‘hold the floor’, following Schourup (1985), and ‘searching for the right word/phrase’ (Macmillan, 2002). In addition to these six interpersonal functions, which have already been established in previous research, cases where *like* is used to indicate an ‘unfinished point’ were also observed. Below follows an overview of all interpersonal functions uncovered in my data, with a short definition and example of each.

Table 13: Definitions and examples of the interpersonal functions of *like*

Interpersonal functions	Definitions	Examples
Hedger	When <i>like</i> is used in the sense of "similar to" or "kind of".	<i>I had to quit and I start working in a: . like a surf shop (N1)</i>
Attention-getter	<i>Like</i> used to plea for the hearer's attention. This function is often used to signalize that there is a shift or that something important is about to come.	<i>do it the way that she'd appreciate it because like when <?> he did the real thing (E136)</i>
Emphasizer	When <i>like</i> is used to highlight a particular word/phrase. Often used to make a point, express attitude etc.	<i>I just checked it was like totally random I just checked Hamar and (N809)</i>
Approximator	When <i>like</i> is used to indicate an approximate time, amount of something, age etc.	<i>I was probably like three or four years old (N467)</i>
Searching for the right word/phrase	When <i>like</i> is used for the speaker to gain time to think of what to say next or search for a word or a phrase.	<i>so we just we had these like erm . bags with ready-made meals (E1329)</i>
Hold the floor	When <i>like</i> indicates that the speaker intends to continue his/her turn.	<i>they were<?> quite interesting . and like . the fact that everything is on campus (E46)</i>
Unfinished point	In the sense of "and so on" or "you know what I mean".	<i>her friends saying this is a picture of her and like but it's not really a true (E51)</i>

This section will first discuss the overall use of *like* for interpersonal functions in LOCNEC and LINDSEI. Thereafter, a closer examination of the three interpersonal functions ‘hedger’, ‘attention-getter’ and ‘approximator’ will be conducted, with the purpose of exploring possible explanations for the discrepancy between the two speaker groups. I have chosen these three functions because the analysis has revealed a highly significant underrepresentation of *like* by the NLEs for the first two functions, and a strong similarity between the two speaker groups for the third function.

Figure 11 below illustrates the relative frequency pttw for each interpersonal function of *like* in both corpora. As this figure shows, there is an overall underrepresentation of *like* serving interpersonal functions in LINDSEI-NO as compared to LOCNEC. The statistical calculations presented in Table 11 indicate that this result is highly significant¹⁴, and the effect size tells us that it further serves as strong evidence against the H0. This finding is in line with Müller (2005), who also discovers an underuse of interpersonal functions of *like* among the German learners. However, the present study’s results differ from Müller’s results when it comes to ‘emphasizer’ *like*. Whereas Müller (2005) reports a highly significant

¹⁴ More than 99.99% certain.

underuse of this function by the German learners, the NLEs of this study demonstrate a much more reference-like usage. A possible explanation to this difference is that the learners of this study are compared with British speakers of English whereas the German learners in Müller’s study are compared with American speakers, who may use *like* with this function more often than the British.

Moreover, Figure 11 shows that the three most frequently used interpersonal functions of *like* in LOCNEC are ‘hedger’, ‘attention-getter’ and ‘emphasizer’, whereas they in LINDSEI-NO are ‘hedger’, ‘approximator’ and ‘emphasizer’, respectively. This result contrasts with Müller’s (2005), who reports that ‘attention-getter’ denotes the largest group of interpersonal functions for both speaker groups, although underrepresented by the German learners. *Like* serving as an ‘attention-getter’ is highly underrepresented by the learners in this study’s data too.

Furthermore, Table 11 shows that there are only 4 instances of *like* used to ‘hold the floor’ in LINDSEI-NO (0.3 pttw), whereas LOCNEC has 20 (1.2 pttw), and only 6 instances of ‘unfinished point’ in LINDSEI-NO (0.5 pttw), whereas there are 17 instances in LOCNEC (1 pttw). These numbers suggest that only a few of the Norwegian learners actually master these two functions.

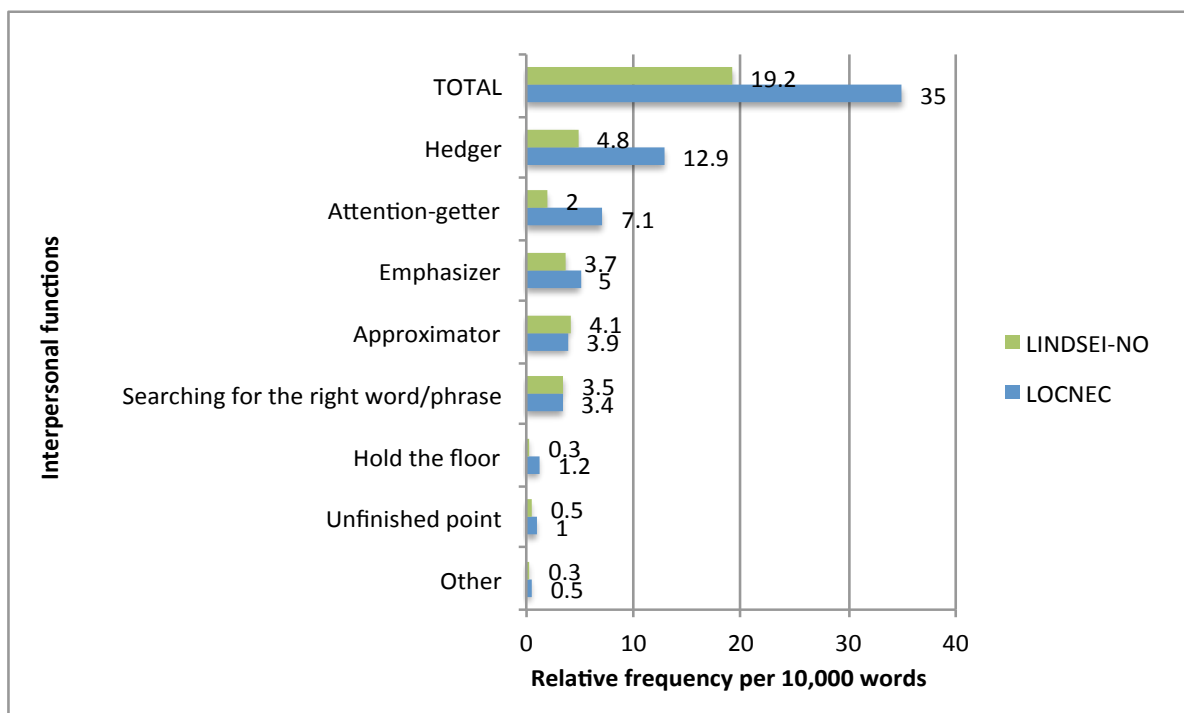


Figure 11: Interpersonal functions of *like* in LOCNEC and LINDSEI-NO

Hedger:

Table 11 presents a total of 220 instances of *like* functioning as a ‘hedger’ in LOCNEC (12,9 pttw) and 60 instances in LINDSEI-NO (4.8 pttw). As already mentioned, this function of *like* is significantly¹⁵ underrepresented in the learner data. If we compare the LL-value for this function with the other functions in Table 11, we notice that this in fact is the function of *like* that is most strongly underrepresented in LINDSEI-NO as compared to LOCNEC.

With regard to the immediate co-text of *like* functioning as a ‘hedger’, some differences can be observed. The most common collocate of *like* serving this function is any form of the verb *to be* (LOCNEC n=44, 2.6 pttw; LINDSEI n=21, 1.7 pttw) for both speaker groups. This construction is illustrated below in example (5.10). Furthermore, the construction *sort of/kind of like*, as in example (5.11), occurs 40 times in LOCNEC (2.3 pttw). Surprisingly, this construction is only identified 4 times in LINDSEI-NO (0.3 pttw), indicating that the learners have not fully acquired this use of the DM. Another frequent construction of the ‘hedger’ *like* in LOCNEC is *just like* (n=21, 1.3 pttw), as illustrated by example (5.12) below, but this structure is entirely absent in LINDSEI-NO.

5.10 *to go on a honeymoon . to Hawaii whenever that time . comes <A> yeah (eh) yeah well I'm not a . big fan of . cities I'm not **like** nature kind of guy . especially when I travel but (eh) . for a big city well . Los Angeles was: awesome and (eh) (N434)*

5.11 *great big sort of sticks built out over onto onto the river <A> oh yes so they'd be literally living over the river <A> mhm sort of **like** reflecting the tribes of south east Asia and things . (E896)*

5.12 *but we've been up a few times . I've got a friend who . er I'm in college with er she's got a car so we can often go just **like** walking up there it's quite it's quite nice <A> are you staying on campus yeah <A> do you like it (E672)*

In sum, the above observations indicate that there are certain uses of *like* serving the interpersonal function “hedger” that learners have not acquired, and certain uses that can be considered misuses. It is difficult suggest explanations for this underrepresentation and misuse judging only by this study’s data, but it can possibly be affected by non-linguistic factors such as lack of input and transfer from the learners’ mother tongue. With regard to transfer, it is possible that the learners transfer the Norwegian hedger *på en måte* into the

¹⁵ More than 99,99% certain.

DMs *kind of* and *sort of* instead of into *like*. A comparative study of the BESs and NLEs' use of these two DMs would reveal whether this assumption is valid and can serve as an explanation for the observed underrepresentation of the 'hedger' *like*.

Attention-getter:

With regard to "attention-getter", Table 11 shows that this interpersonal function of *like* occurs 121 times in LOCNEC (7,1 pttw) and 24 times in LINDSEI-NO (2 pttw), again revealing an underrepresentation by the learners. The statistical calculations presented in Table 11 suggest that this result is highly significant¹⁶, which is further supported by the effect size value indicating that it serves as strong evidence against the H0.

In order to investigate how, and perhaps why, this function is underrepresented in the learners' speech production, a closer examination of the total 145 cases belonging to this functional category has been conducted. This analysis has uncovered that this use of *like* typically is underrepresented in the learner-data when the DM serves the dual interpersonal and textual functions 1) 'attention-getter' and 'topic shift' (LOCNEC n=37, 2.2 pttw; LINDSEI-NO n=4, 0.3 pttw), illustrated in example (5.13) and (5.15) below, 2) 'attention-getter' and 'exemplifier' (LOCNEC n=24, 1.4 pttw; LINDSEI-NO n=6, 0.5 pttw) as in example (5.16), and 3) 'attention-getter' and 'introducing explanation' (LOCNEC n=35, 2 pttw; LINDSEI-NO n=3, 0.2 pttw), illustrated in example (5.14). The analysis further reveals that the BESs use 'attention-getter' *like* more often in constructions where it is preceded by conjunctions, such as *and like* (LOCNEC n=18, 1 pttw; LINDSEI-NO n=1, 0.1 pttw), *so like* (LOCNEC n=4, 0.2 pttw; LINDSEI-NO n=0), *but like* (LOCNEC n=7, 0.4 pttw; LINDSEI-NO n=0), and *because/cos like* (LOCNEC n=9, 0.5 pttw; LINDSEI-NO n=1, 0.1 pttw) than the NLEs. Example (5.13) – (5.16) illustrate such cases.

5.13 *on the track in front of us oh they stopped and turned to charge the minibus [and the [oh the driver turned it round and **like** you've never seen [anything move so quickly in your life [and you said it wasn't dangerous but and he mini (E50).*

5.14 *been (eh) a .. a to= a topic . I . like a language I enjoy . and I I (eh) wanted to explore the . literature more .. because **like** Hamlet is something you you think you= . you're going to do but you never get around to it you know ... yeah now I w= I' (N31)*

¹⁶ More than 99,99% certain.

5.15 *the Danish I think yeah yeah mhm Danish around here that kind of thing [anyway [yeah yeah all that . so yeah so anyway so **like** I di= I did that mhm and then I moved er I moved away again mhm er went to do golf green keeping oh yeah yeah oh you did (E1143)*

5.16 *to be cos when we were there a: . the things that the rides were magnificent you went on but there wasn't enough of them and **like** you got easily bored of it and just walked around the same shops all the time [it was nice but not for to: long yeah ye (E72)*

These observations indicate that the BESs' use of *like* with the interpersonal function 'attention-getter' often occurs when there is a shift, or where the speaker moves to introduce new information in the form of a new topic, an explanation or an example. In such cases the speaker wishes to signalize to the hearer that he or she needs to pay attention because something important is to come. This can be either new information or just an important point. Furthermore, these shifts typically occur in subordinate clauses or in a new sentence, and it is therefore natural that these uses of *like* are preceded by conjunctions such as the ones listed above. The in-depth analysis has revealed that the NLEs have not fully acquired these uses of *like* to express the interpersonal function "attention-getter". More input and instruction would perhaps serve to reduce the underrepresentation.

Approximator:

As already mentioned above, the interpersonal function 'approximator' revealed a quite similar distribution in both corpora. In LOCNEC a total of 67 hits of *like* are identified as "approximator" (3.9 pttw), and in LINDSEI-NO 51 cases are identified serving the same function (4.1 pttw).

Based on these findings, a qualitative analysis of every instance of *like* serving this function has been conducted, with special focus on what type of measurement the DM is referring to in such constructions. This analysis uncovers that both speaker groups use the 'approximator' *like* most frequently to refer to time (LOCNEC n=27, 1.6 pttw; LINDSEI-NO n=18, 1.5 pttw). This can either be a point in time or duration of time, and is illustrated by example (5.17) below. The second most frequent entity of reference is quantity (LOCNEC n=24, 1.4 pttw; LINDSEI-NO n=11, 0.9 pttw), as illustrated by example (5.18). Within this category are cases of *like* referring to number of items, people etc. Furthermore, 'approximator' *like* is by both speaker groups also used to refer to age (LOCNEC n=1, 0.1 pttw; LINDSEI-NO n=4,

0.3 pttw) and temperature (LOCNEC n=2, 0.1 pttw; LINDSEI-NO n=4, 0.3 pttw), although the numbers of occurrences are very low for approximator *like* used with reference to these measurements.

5.17 *own I wanna travel and see the all the places . because when you first settle down and you get kids you're kinda stuck . for **like** . ten years . unless you wanna drag a kid along yeah kids and with home school home schooling and y (N130)*

5.18 *lege County College oh County oh yes. I think there's a . a good atmosphere [there in that college [yeah .. yeah we've got **like** . twenty-nine people in the kitchen [and yeah <A> [twenty-nine .. [that's [so that's a lot yeah . [so we had to get to (E480)*

So far, although some differences in frequency are present, the two speaker groups seem to display similar usage of the ‘approximator’ *like*. However, further observations reveal some discrepancy between the two groups. The NLEs are found to use the ‘approximator’ *like* to refer to age (n=4, 0.3 pttw), distance (n=5, 0.4 pttw) and price (n=7, 0.6 pttw), whereas the BESs demonstrated no use of *like* with these categories. On the other hand, the BESs use *like* to refer to weight (n=6, 0.4 pttw) and size (n=3, 0.2 pttw) whereas this category is absent in LINDSEI-NO. Example (5.19) below illustrates the use of ‘approximator’ *like* used to refer to price.

5.19 *yeah . Scandinavian people yeah yeah yeah I'm going back this summer maybe no only for a few weeks you know . you pay **like** .. ten thousand Norwegian . just for the plane and when you arrive everything is .. almost free (N454)*

Although the frequencies of the above mentioned observations are too low to make any statistically significant claims, these observations suggest that some differences do exist between the two speaker groups in how they use ‘approximator’ *like*. These results may be due to chance, as each interview only lasts approximately 15 minutes and thus limits the number of topic discussed, but they may also be explained by factors such as cultural differences, and differences in fields of interest. Furthermore, transfer may also serve as an explanatory cause for the above results, as the NLEs possibly rather would use words more similar to the Norwegian equivalents *kanskje (maybe)* and *sånn (ca.) (like (about))* to express the same function.

Qualifying functions of *like*

The analysis of qualifying functions of *like* yielded few results both in terms of total amount of hits and number of functions. Only 66 out of the total 1194 cases of *like* in LOCNEC and LINDSEI-NO were analyzed as serving a qualifying function. Surprisingly, 39 of these 66 cases are spoken by the Norwegian learners (3.1 pttw), whereas the remaining 27 are produced by British English speakers (1.6 pttw). This overrepresentation of qualifying *like* by the NLEs is found to be statistically significant¹⁷. However, looking at the effect size value presented in Table 11, we see that this result cannot be considered meaningful, as the difference is not considered strong enough to reject the H0.

Figure 12 below presents the different qualifying functions of *like* identified in the two corpora, presented in relative frequency pttw. The figure demonstrates that *like* is mainly used to ‘express attitudes, feelings or evaluations’ by both the BESs and NLEs. Other functions identified are ‘negative politeness’ and ‘positive politeness’, but these are infrequent in the material and have therefore been placed in the category ‘other’. This section will therefore only examine the two speaker groups’ use of *like* serving the qualifying function ‘expressing attitudes, feelings or evaluations’.

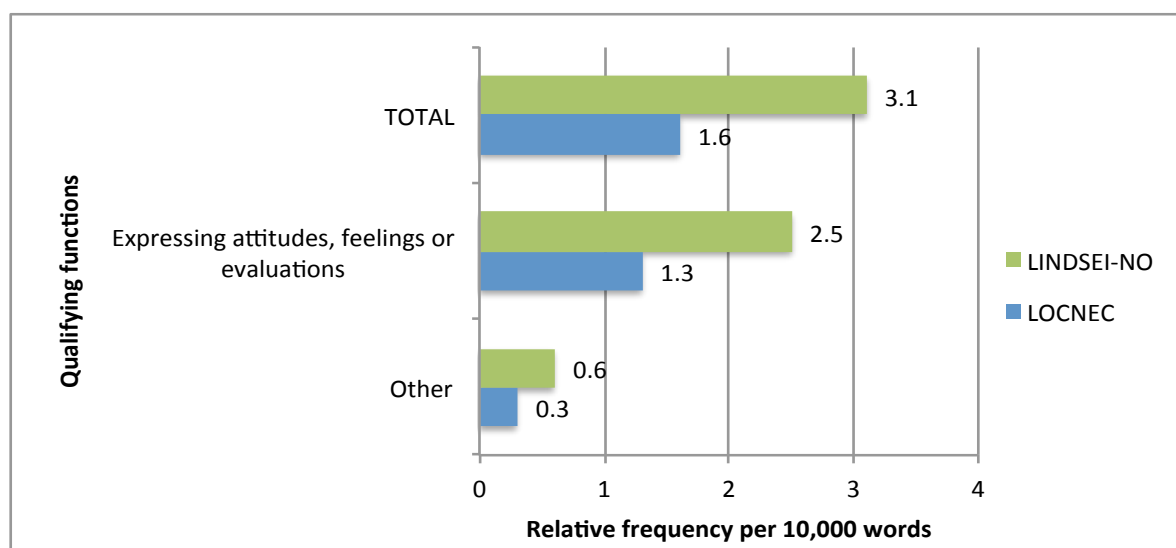


Figure 12: Qualifying functions of *like* in LOCNEC and LINDSEI-NO

¹⁷ More than 95% certain.

Expressing attitudes, feelings or evaluations:

A closer examination of the 66 cases of *like* functioning to ‘express attitudes, feelings or evaluations’ reveals that there are hardly any cases where this use of *like* serves a dual qualifying and textual function. On the other hand, these instances of *like* almost always occur as dual qualifying and interpersonal functions. The four most frequent interpersonal functions in this context are: 1) ‘approximator’ (LOCNEC n=4, 0.2 pttw; LINDSEI-NO n=6, 0.5 pttw), 2) ‘attention-getter’ (LOCNEC n=1, 0.1 pttw; LINDSEI-NO n=1, 0.1 pttw), 3) ‘emphasizer’ (LOCNEC n=15, 0.9 pttw; LINDSEI-NO n=19, 1.5 pttw) and 4) ‘hedger’ (LOCNEC n=2, 0.1 pttw; LINDSEI-NO n=2, 0.2 pttw). Such uses of *like* are illustrated in examples (5.20)-(5.23) below, respectively.

5.20 *like .. ten thousand Norwegian . just for the plane and when you arrive everything is .. almost free you don't pay . I pay **like** (eh) . twenty Norwegian crowns for a . for a good . good dinner for me and my girlfriend and that's cheap ... and also t (N453)*

5.21 *there trying to find Mona Lisa yeah I found her but I I thought that the painting would be bigger . (eh) she was so little . **like** you were standing five metres . away from her and she was just this tiny picture and . oh (eh) . when you see her in a b (N328)*

5.22 *I did English literature A-level oh yes so . we had set texts and our set texts of Shakespeare was King Lear uhu which is **like** the longest play he ever wrote <A> [oh yes [and also the most depressing because everyone dies in the[i:] end so I thought (E599)*

5.23 *holds the door before<?> you if they walk into you oh sorry love you know what I mean . here it's just like everyone's just **like** [stony-faced <A> [yes yes and when you go to the supermarket mm everyone bumps into you and they they never say sorry no . (E641)*

As becomes clear from the above results, all these uses of *like* are represented in both corpora with approximately the same relative frequency. This indicates that the overrepresentation of this function of *like* in the learner-data cannot be explained by misuse, as both speaker varieties display similar tendencies. Instead, the above results indicate that the NLEs master this function of *like* quite well. This can be explained by satisfactory input or by transfer from the Norwegian equivalent *liksom*, which is used for approximately the same purposes. The observed overrepresentation of this function in LINDSEI-NO can for instance be explained cultural differences. However, as this aspect lies beyond the scope of this thesis, this assumption will have to be further investigated in future studies.

5.3 *Well*

The following section presents the analysis of the second discourse marker of this study, *well*. *Well* is classified as a reception marker (Jucker and Smith, 1998: 197) and as a sharing device (Svartvik, 1980: 168), and previous researchers have found one of its main functions to be to create coherence (Schiffrin, 1987). These three characteristics mainly belong within the textual level, but *well* is also found to serve interpersonal and qualifying functions such as ‘taking and yielding the turn’ and ‘marking disagreements’. This section investigates how the Norwegian IL speakers and British RLV speakers behave when using *well* in spoken language, and to what extent they differ in their usage.

This section has the same structure as section 5.2 for *like*. First, in sub-section 5.3.1, the results from the quantitative frequency analysis will be presented and discussed. Then, section 5.3.2 presents the results from the functional analysis where some of the most significant results will be singled out for a closer examination.

5.3.1 Frequency analysis

Overall usage in LOCNEC and LINDSEI-NO

As disclosed in Table 8 in section 5.1, *well* is the second most frequent discourse marker of the three markers under study in both LOCNEC and LINDSEI-NO. It occurs 681 times in LOCNEC (40 pttw) and 319 times in LINDSEI-NO (25,9 pttw). Similar to the results for *like*, *well* is also generally underrepresented in LINDSEI-NO as compared to LOCNEC. This difference is found to be highly significant¹⁸, and the effect size value indicates that the result is meaningful and serves as strong evidence against the H0. Thus, the results suggest with a high degree of scientific certainty that *well* is underrepresented in the learner-data as compared to the British RLV material.

Utterance position of *well*

Table 14 below illustrates the distribution of the utterance position of *well* in LOCNEC and LINDSEI-NO, and Figure 13 presents a graphical illustration of the difference between the two corpora. The definitions of the three positions presented in section 5.2.1 also apply here: when occurring in the pre-front field, *well* typically functions as a theme to introduce new information. This position is typically used for the textual functions ‘introducing explanation,

¹⁸ More than 99,99% certain.

justification, conclusion etc’, ‘move to story’, ‘quotative *well*’, ‘rephrase/repair’ and ‘response marker’ and the interpersonal functions ‘attention-getter’, ‘hesitator’, ‘take the turn’ and ‘thinking/considering’. When occurring as a wildcard, *well* serves as a random insertion that can occur anywhere except from at the beginning or the end of the utterance. This position is mostly associated with the textual functions ‘pausal interjection’ and ‘rephrase/repair’, and the interpersonal functions ‘hesitator’ and ‘searching for the right word/phrase’. There are no cases of *well* occurring in the post-end field. This third position is therefore excluded from Table 14 and Figure 13 below.

Example (5.24) illustrates the use of *well* in pre-front field, where the discourse marker serves the dual textual and interpersonal functions ‘response marker’ and ‘take the turn’. Example (5.25) illustrates the use of *well* in wildcard position, where it serves the dual textual and interpersonal function ‘pausal interjection’ and ‘hesitator’.

- 5.24 <A> (mm) yeah okay . let’s switch topics a little bit and . tell me do you have any hobbies . beyond .. movies yes of course . **well** I’m a scout actually <A> a scout yeah <A> okay what do you do a girl guide .. I’ve been (eh) a scout for ten years (mhm) (N114)
- 5.25 just . look at the reactions that the movie got when it came out (mm) because: one of the last things that happens is that . **well** . the bikers get shot from . by some . rednecks in a pickup truck <A> (mhm) and (eh) the reaction in . Alabama and Texaswas (N395)

Table 14: Utterance position of *well* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	BAYES
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Pre-front field	615	36.1	281	22.9	42.38	< 0.0001	29.79
Wildcard	66	3.9	38	3.1	1.26	n.s.	-11.33
TOTAL	681	40	319	25.9	42.51	< 0.0001	29.92

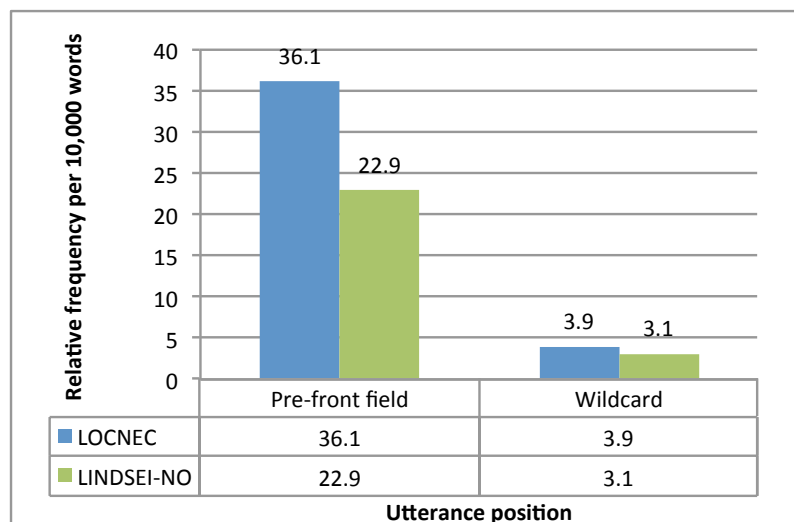


Figure 13: Utterance position of *well* in LOCNEC and LINDSEI-NO

As Table 14 demonstrates, most cases of *well* in both LOCNEC and LINDSEI-NO occur in the pre-front field. The difference between the NLEs and the BESs in their use of *well* in the pre-front field is highly significant¹⁹, and the effect size value for this difference indicates that the result is meaningful and serves as strong evidence against the H0. The results for *well* used as a wildcard, on the other hand, display a much smaller difference between the two speaker groups, and which is not significant. The effect size tells us that this result is not strong enough to serve as evidence against the H0.

Orientation of *well*

When categorizing *well* in terms of orientation, the same definitions as the ones applied for the analysis *like* are used: if the discourse marker points forward to some new information that is to come, it is considered prospective. If, on the other hand, it points backwards to old information that is already mentioned, it is classified as retrospective. Similarly to the results for the DM *like*, there are no occurrences of retrospective *well* in the material, and this orientation is therefore not included in Table 15 and Figure 14 below. Example (5.26) illustrates the use of *well* with prospective orientation. In this example, *well* has the textual function ‘introducing further specification’.

5.26 A in primary [education <A> [uhu yeah what does that involve erm . well you you pick two subjects .. I do <?> English er **well** linguistics and English and religious studies oh yes and erm as well as those two subjects you do a lot of (E338)

¹⁹ More than 99.99% certain.

Table 15 presents the results and statistical calculations for the orientation of *well* in LOCNEC and LINDSEI-NO. Figure 14 gives a graphical illustration of how the two speaker groups differ from each other, presented in relative frequency per 10,000 words.

Table 15: Orientation of *well* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG-LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Prospective	681	40	319	25.9	42.51	< 0.0001	29.92

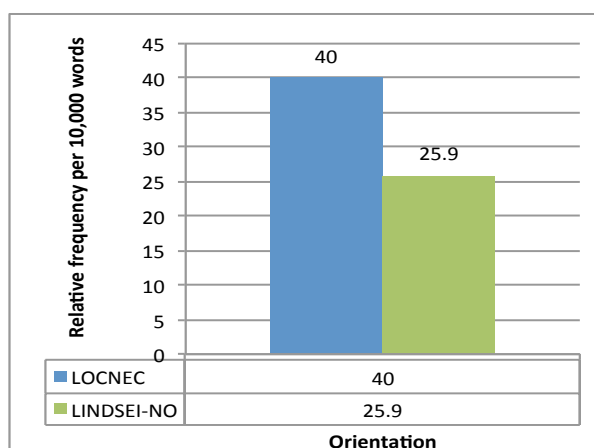


Figure 14: Orientation of *well* in LOCNEC and LINDSEI-NO

As the results in Table 15 show, the NLEs display a highly significant²⁰ underrepresentation of *well* with prospective orientation as compared to the BESs.

5.3.2 Functional Analysis

Table 16 below presents the results of the functional analysis of *well*, offering an overview of the distribution of textual, interpersonal and qualifying functions in LOCNEC and LINDSEI-NO. The table shows that the general underrepresentation of the DM *well* in the NLE's speech production, as reported in section 5.3.1, also is registered for most pragmatic functions of *well*. Exceptions here are the interpersonal functions 'thinking/considering' and 'attention-getter' and the qualifying function 'refusing to answer a question directly', where the learners are found to overrepresent *like*, and the functions 'move (back) to story', 'indirect/insufficient answer', 'expressing attitudes, feelings or evaluations' and 'contributing/qualifying an opinion', where the two speaker groups show similar frequencies. The functions that show the strongest underrepresentation are 'quotative *well*',

²⁰ More than 99.99% certain.

‘rephrase/repair’, ‘approximator’, ‘take the turn’, ‘expressing disagreement’, ‘modifying a previous statement’, and ‘confirming a previous statement’. These, together with some of the functions suggesting overrepresentation of *well* in the learner data, will be subject to a more in-depth investigation later in this section.

Table 16 further demonstrates that the NLEs use *well* most often for textual functions (n=279, 22.3 pttw), followed by interpersonal functions (n=190, 15.6 pttw) and qualifying functions (n=82, 6.8 pttw). The BESs have the same order, but exhibit a much higher relative frequency of all three functions. The difference between the two speaker groups is most convincing for the qualifying functions, indicated by the effect size value.

Table 16: Textual, interpersonal and qualifying functions of *well* in LOCNEC and LINDSEI-NO

		LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	EFFECT SIZE
		Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
TEXTUAL FUNCTIONS	Response marker	192	11.3	115	9.4	2.54	n.s.	-10.05
	Introducing explanation, justification, conclusion etc.	125	7.4	77	6.4	0.95	n.s.	-11.38
	Quotative 'well'	53	3.1	7	0.1	26.54	< 0.0001	13.95
	Rephrase/repair	65	3.8	30	2.4	4.31	< 0.05	-8.28
	Move (back) to story	47	2.7	30	2.4	0.28	n.s.	-12.31
	Other	27	1.5	20	1.6	0.01	n.s.	-12.58
	TOTAL	509	29.8	279	22.3	11.53	< 0.01	-1.06
INTERPERSONAL FUNCTIONS	Approximator	11	0.6	1	0.1	6.81	< 0.01	-5.78
	Attention-getter	17	1	17	1.4	0.9	n.s.	-11.69
	Searching for the right word/phrase	28	1.6	11	0.9	3.15	n.s.	-9.43
	Indirect/ insufficient answer	17	1	12	1	0	n.s.	-12.59
	Self-interruption	50	2.9	28	2.3	1.18	n.s.	-11.41
	Take the turn	56	3.3	18	1.5	10.05	< 0.01	-2.54
	Hold the floor	31	1.8	13	1.1	2.88	n.s.	-9.71
	Hesitator	78	4.6	48	3.9	0.76	n.s.	-11.82
	Thinking/ considering	30	1.8	37	3	4.79	< 0.05	-7.8
	Other	4	0.3	5	0.4	0.68	n.s.	-11.91
	TOTAL	322	18.9	190	15.6	9.01	< 0.01	-3.57
QUALIFYING FUNCTIONS	Correcting a misunderstanding	2	0.1	2	0.2	0.11	n.s.	-12.48
	Refusing to answer a question directly	5	0.3	7	0.6	1.3	n.s.	-11.28
	Expressing disagreement	6	0.4	0	0	6.52	< 0.05	-6.07
	Rejecting a previous statement	6	0.4	1	0.1	2.52	n.s.	-10.07
	Modifying a previous statement	86	5	28	2.3	15.04	< 0.001	2.45
	Confirming a previous statement	22	1.3	4	0.3	8.54	< 0.01	-4.05
	Contributing/qualifying an opinion	9	0.5	6	0.5	0.09	n.s.	-12.57
	Expressing attitudes, feelings or evaluations	51	3	34	2.8	0.13	n.s.	-12.46
	TOTAL	187	11	82	6.8	14.98	< 0.001	2.39

Textual functions of *well*

Among the textual functions discussed in previous research, (see section 2.3.1), most have been identified in my data. Following Svartvik (1980), cases of *well* serving to indicate a topic shift are present. However, as there are too few instances of this use of *well*, this function has been placed in the ‘other’ category together with ‘pausal interjection’. The textual function ‘move to story’ (Müller, 2004) is identified in both LOCNEC and LINDSEI-NO. So is Müller’s (2004) function ‘introducing a conclusion’. However, I do not find this category to cover all relevant cases in my data, and I have therefore found it necessary to create a broader category including the functions ‘introducing explanation’, ‘introducing justification/clarification’, ‘introducing conclusion’, and ‘introducing further specification’. This category has been labeled ‘introducing explanation, justification, conclusion etc.’.

The two functions ‘quotative *well*’ and ‘rephrase/repair’ identified in the present analysis correspond to Svartvik’s (1980: 174) categories “[indicate] beginning of speech” and “editing marker for self-correction”. In addition to these textual functions, the pragmatic function “response marker” is also included, following Lakoff’s (1973: 463) emphasis on *well* used for responses to questions. In several instances the speaker appears to use *well* to indicate that a response to a question or previous statement will come. In most of these cases, the speaker seems to sense some sort of insufficiency in his/her answer, corresponding to how Lakoff (1973) defines this category. Table 17 below presents definitions and examples of each of the textual functions identified in LOCNEC and LINDSEI-NO for this study.

Table 17: Definitions and examples of the textual functions of *well*

Textual function	Definition	Example
Move (back) to story	When <i>well</i> is used to move back to the main topic after a digression, an input from the hearer or other types of shifts and stops	<i>when he was told his father had just died . things like that mhm .. well sometimes when I watched that film (E586)</i>
Rephrase/repair	When <i>well</i> is used to start over again after an immediate stop in the middle of the utterance	<i>I guess she wanted (eh) ... well . guess she felt she (N145)</i>
Quotative <i>well</i>	When <i>well</i> is used to quote direct speech, thoughts or feelings	<i>just looking around me thinking well .. I'm in a foreign country I've got .. absolutely no money (E904)</i>
Introducing explanation, justification, clarification, conclusion, and further specification	When <i>well</i> is used to introduce an explanation, justification, clarification or modification to a previous statement, a conclusion, or further specification of something already known to the hearer.	<i>I don't normally go home very much erm . well I go home for the hol= holidays (E422)</i>
Response marker	When <i>well</i> is used to introduce a response to a previous question or statement.	<i>[A] then you must have a lot of background knowledge [B] well as I said I like history and (N2539)</i>

Figure 15 below illustrates the use of textual functions of *well* in LOCNEC and LINDSEI-NO presented in relative frequencies per 10,000 words. As this figure shows, *well* serving textual functions is strongly underrepresented in LINDSEI-NO as compared to LOCNEC, and this difference is present in almost every textual function. The exceptions here are ‘move (back) to story’ and ‘other’. The strongest underrepresentation is found in ‘quotative *well*’, and thereafter follows ‘rephrase/repair’. The LL and P-value in Table 16 suggest that the difference between the two speaker groups for the former function is highly significant²¹, and the effect size confirms that the result serves as strong evidence against the H0. For the latter function, the difference is again statistically significant²², but the effect size rejects the meaningfulness of this result and indicates that it rather serves as evidence in favor of the H0. Due to the limited scope of this thesis, it is not possible to examine every textual function in detail. Therefore, these two functions, which reveal the strongest difference between the NLEs and BESs, are chosen.

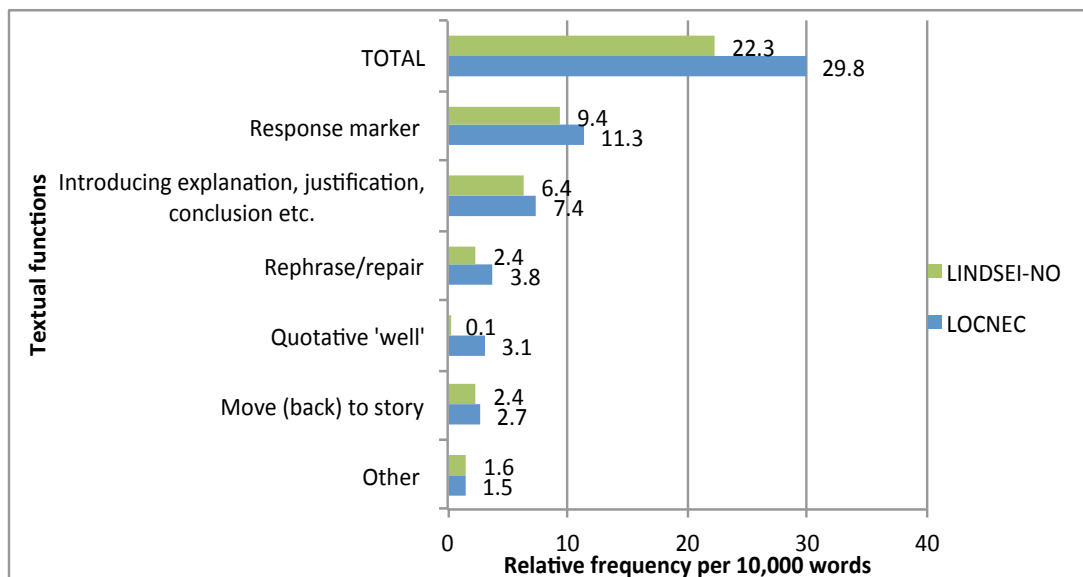


Figure 15: Textual functions of *well* in LOCNEC and LINDSEI-NO

Quotative 'well':

As discussed above, ‘quotative *well*’ is the function that is most underrepresented in LINDSEI-NO as compared to LOCNEC. There are 53 instances of *well* serving this function in LOCNEC (3.1 pttw), whereas only 7 instances are attested in LINDSEI-NO (0.1 pttw). The examination of the collocation patterns of ‘quotative *well*’ uncover that the DM typically is preceded by verbs indicating the actions of saying (LOCNEC n=15, 0.9 pttw) and thinking

²¹ More than 99.99% certain.

²² More than 95% certain.

(LOCNEC n=19, 1.1 pttw). This is illustrated by example (5.27) below. Surprisingly, this result only applies to LOCNEC. Among the 7 hits of quotative *well* found in LINDSEI-NO, only one is immediately preceded by a verb denoting such actions (0.1 pttw). Two more are preceded by the string *thought/thinking that* (0.2 pttw), which is demonstrated by example (5.28) below. There are no instances of such a string in LOCNEC.

5.27 *even I'll have to just drink [this quick [yeah I mean I talked to one of the students a first year student who and he said **well** I know that er I have three hours to have a certain am= amount of alcohol [[and so I just drink and drink and drink (E785)*

5.28 *she can't have a lot of self-insight .. well they're probably just smiling and nodding and thinking that . **well** it doesn't look like you but good for you (mhm) .. yeah . okay . thank you (N396)*

These observations indicate that the learners have not fully acquired how 'quotative *well*' typically is used in context (e.g. preceded by verbs of saying or thinking). This can possibly be explained by non-linguistic factors such as lack of input and transfer from the Norwegian translation equivalent *vel*. As 'quotative *well*' is a colloquial expression, it is unlikely that this function of *well* has been given any profound attention in classroom instructions. Moreover, the Norwegian *vel*, which also is a DM serving to organize discourse and modifying the force of an utterance, is, to my knowledge, more rarely used together with introductions of direct speech than its English counterpart.

Rephrase/repair:

'Rephrase/repair' is the second textual function that is highly underrepresented in LINDSEI-NO as compared to LOCNEC. It occurs 65 times in LOCNEC (3.8 pttw) and 30 times in LINDSEI-NO (2.4 pttw). An in-depth investigation of the total 95 cases of *well* used to mark 'rephrase/repair' uncovers that the two speaker groups use this function similarly in terms of orientation, utterance position and collocations. *Well* is used exclusively with prospective orientation and mostly in the pre-front field (LOCNEC n=56, 3.3 pttw; LINDSEI-NO n=28, 2.3 pttw), and in terms of collocations, no particular patterns were identified for either speaker group. Furthermore, most of the cases only serve the single textual function 'rephrase/repair' (LOCNEC n=21, 1.2 pttw; LINDSEI-NO n=25, 2 pttw), and the remaining cases serve the following dual interpersonal functions: 1) 'self-interrupting' (LOCNEC n=26, 1.5 pttw; LINDSEI-NO n=7, 0.6 pttw), 2) 'attention-getter'

(LOCNEC n=2, 0.1 pttw; LINDSEI-NO n=3, 0.2 pttw), 3) ‘hesitator’ (LOCNEC n=4, 0.2 pttw; LINDSEI-NO n=1, 0.1 pttw), and 4) ‘thinking/ considering’ (LOCNEC n=4, 0.2 pttw).

Some of the above-mentioned characteristics are illustrated by example 5.29 and 5.30 below. Example (5.29) shows the use of *well* with the textual function ‘rephrase/repair’ where the DM has prospective orientation and occurs in a wildcard position. In this example the speaker starts by saying “England”, but then stops, goes back, and uses *well* to correct the statement to “Great Britain”. In example (5.30) *well* serves the textual function ‘rephrase/repair’ and the interpersonal function ‘self-interruption’, and the marker has prospective orientation and occurs as a wildcard. In this example the speaker is about to say something like “little less people”, but then interrupts him/herself and corrects it to “fewer people”.

5.29 to Europe [em [or <?> the continent well here it's Europe too I mean it it's funny because ... it's always as if England **well** Great Britain was not part of Europe when we're talking <A> yeah .. that's true .. em . really I'm I'm not bothered I'd (E333).

5.30 When she was able to connect with each and one of us not . (eh) necessarily as a group but (mm) when we were a little le= **well** fewer people (mm) s= . <A> so you identified with the Tolkien reader already then yeah seventeen yes (laughs) yeah well I (N255)

Based on previous research suggesting that learners tend to be more hesitant in their speech and therefore search more often for what to say than native speakers (Müller, 2004: 1175), I had expected to uncover an overrepresentation of *well* serving this function in the learner-data. This was not the case, and the in-depth investigation revealed no noteworthy difference between the two speaker groups. These findings indicate that the underrepresentation of ‘rephrase/repair’ *well* in LINDSEI-NO cannot be explained by factors such as misuse or lack of knowledge within this area. Instead, it might be that the Norwegian IL speakers use other means for marking ‘rephrase/repair’ than the discourse marker *well*, which are not investigated in this study. Such means may include the markers *em* and *eh*, which are frequently used for the pragmatic functions ‘searching for the right word/phrase’, ‘hesitator’ and ‘rephrase/repair’. However, these are only speculations, and further research is needed to answer these questions.

Interpersonal functions of *well*

A total of 512 cases of *well* are classified as serving interpersonal functions the data of this study. Among these, 322 are produced by the BESs (18.9 pttw) and 190 by the NLEs (15.6 pttw). The learners thus display an overall underrepresentation of *well* for interpersonal functions as compared to their British peers. The P-value presented in Table 16 indicates that this result is statistically significant²³. However, the effect size value tells us that the difference is not solid enough to reject the H0.

Figure 16 below provides an overview of all interpersonal functions of *well* discovered in both corpora, presented in relative frequency per 10,000 words. The most frequently used interpersonal function of *well* in LOCNEC is ‘hesitator’, followed by ‘take the turn’ and ‘self-interruption’. ‘Hesitator’ is also the most frequent function in LINDSEI-NO, but the following frequent functions are ‘thinking/considering’ and ‘self-interruption’. Furthermore, the Figure shows that the two functions of *well* ‘thinking/considering’ and ‘attention-getter’ are overrepresented in LINDSEI-NO as compared to LOCNEC.

The two functions yielding the most significant and meaningful results, and thus the two functions that stand out as most interesting in terms of differences between the BESs and NLEs, are ‘thinking/considering’ and ‘take the turn’. These will be subject to an in-depth examination in the following sub-sections.

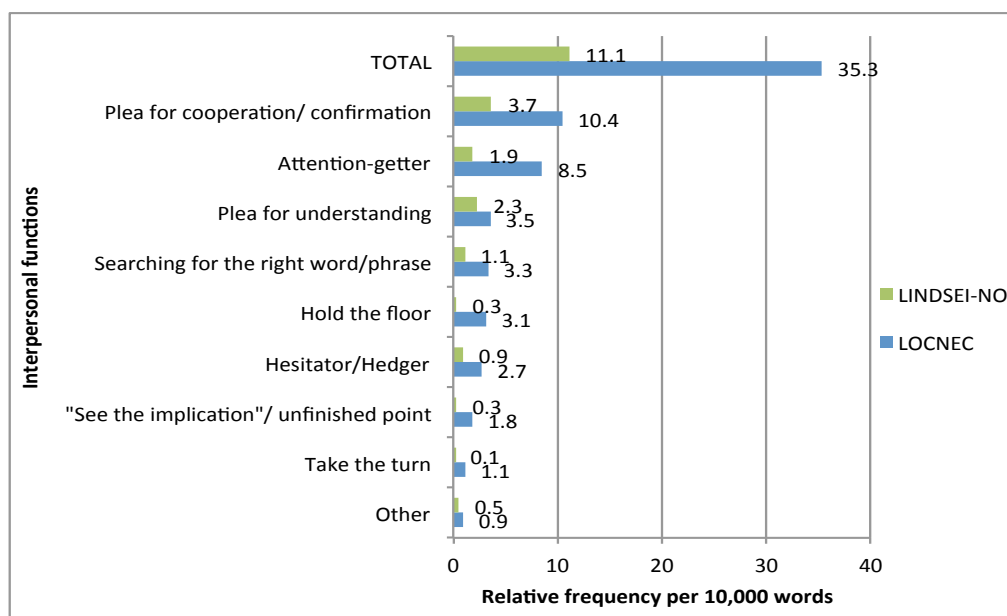


Figure 16: Interpersonal functions of *well* in LOCNEC and LINDSEI-NO

²³ More than 95% certain.

Thinking/considering:

A total of 67 cases of *well* used to indicate the interpersonal function ‘thinking/considering’ are identified in this study. Among these, 30 are produced by the BESs (1.8 pttw) whereas 37 are produced by the NLEs (3 pttw). Together with ‘quotative *like*’, this is one of few functions in this study found to be overrepresented in LINDSEI-NO as compared to LOCNEC. Yet, the effect size rejects this result as evidence against the H0 due to the low number of occurrences in each corpus.

An examination of these 67 cases reveals that both speaker groups use ‘thinking/considering’ *well* exclusively with prospective orientation, and that the utterance position is either pre-front field (LOCNEC n=31, 1.8 pttw; LINDSEI-NO n=32, 2.6 pttw) or wildcard (LOCNEC n=1, 0.1 pttw; LINDSEI-NO n=5, 0.4 pttw).

So far both speaker groups display similar usage. However, in terms of dual functions expressed by this use of *well*, some differences are particularly marked. The most frequently combined dual textual and interpersonal function for this use of *well* is ‘response marker’ and ‘thinking/considering’. This structure occurs 13 times in LOCNEC (0.8 pttw) and 20 times in LINDSEI-NO (1.6 pttw). In other words, the NLEs seem to express the need to think or consider what to say more often than the BESs when giving a response to a question or statement. Apart from this use, the NLEs also use *well* with the dual textual and interpersonal functions ‘pausal interjection’ and ‘thinking/considering’ (n=5, 0.4 pttw), whereas the BESs use it for the textual and interpersonal functions ‘rephrase/repair’ and ‘thinking/considering’ (n=4, 0.2 pttw).

In example (5.31) below, *well* is preceded by *eh*, which is found to be very typical for this function. In this example, *well* serves the dual functions of ‘response marker’ (textual) and ‘thinking/considering’ (interpersonal). The speaker is signaling that he is going to present an answer to the question asked by the interviewee, but that he needs time to consider what to answer first.

5.31 *(eh) the painter do as she he's told and . so the customer's (eh) happy yeah .. and then what do her friends think (eh) .. well . I'm not: (eh) sure they are very . impressed by the painting . to be honest yeah they don't look like they're very (N161).*

Example (5.32) illustrates the use of *well* with the dual textual and interpersonal functions ‘pausal interjection’ and ‘thinking/considering’. As already mentioned, this use of *well* is only found twice in LOCNEC and 4 times in LINDSEI-NO. In the example, the speaker is listing countries, and uses *well* to indicate that she needs a pause to come up with more countries to add to the list. This is an example of *well* occurring as a wildcard, whereas sentence (5.31) illustrates *well* occurring in the pre-front field.

5.32 <A> yeah French . yeah mhm I think it's a good language French yeah well in Europe you can go to France Belgium yeah and **well** . even some other countries yeah . French is quite [usef= useful I think <A> [yeah yeah yes useful if (E17).

In sum, this examination has revealed that the NLEs use the interpersonal function of ‘thinking/considering’ *well* more frequently than the BESs, and that the two speaker groups differ to some extent in how they use *well* for this function. The fact that the Norwegian learners are non-native speakers, and therefore need more time to think of what to say than the native speakers, may serve to explain this difference between the two speaker groups. I do not, however, believe L1 transfer from the Norwegian cognate *vel* to serve as an explanation to the difference, as this DM is not considered a marker of ‘thinking/considering’, especially not by the younger generations. The Norwegian markers *tja*, *nja*, and *em* are more likely to be used for this function than *vel*.

Take the turn:

The second interpersonal function of *well* that will be further examined is ‘take the turn’, of which the NLEs demonstrate a less frequent use than the BESs. This function is identified 74 times LOCNEC and LINDSEI-NO, where 56 cases come from LOCNEC (3.3 pttw) and 18 from LINDSEI-NO (1.5 pttw). The P-value indicates that this difference is highly significant²⁴, but the effect size tells us that this result is not strong enough, thus indicating positive evidence in favor of the H0.

Nevertheless, in the search for an explanation for the underrepresentation of ‘take the turn’ *well* in the NLE data, this function has been closer examined in terms of the DM’s orientation, utterance position, other pragmatic functions and immediate co-text. This analysis reveals that *well* functioning as a turn-taker is exclusively used with prospective

²⁴ More than 99% certain.

orientation in the pre-front field by both speaker groups. *Well* is furthermore found to serve the single interpersonal function ‘take the turn’ 25 times, whereas 40 cases of *well* serve the dual textual and interpersonal functions ‘response marker’ and ‘take the turn’ (LOCNEC n=26, 1.5 pttw; LINDSEI n=14, 1.1 pttw). Moreover, *well* serving this function is typically preceded by discourse markers such as *erm* and *eh* (LOCNEC n=10, 0.6 pttw; LINDSEI-NO n=2, 0.2 pttw), *okay* (LOCNEC n=4, 0.2 pttw; LINDSEI-NO n=1, 0.1 pttw), *right* (LOCNEC n=5, 0.3 pttw; LINDSEI-NO n=0), and *yeah* (LOCNEC n=8, 0.5 pttw; LINDSEI-NO n=4, 0.3 pttw). These words can all be categorized within Aijmer’s function ‘topic-changer’, defined to indicate “a boundary or a ‘frame’ between two discourse units” (2002: 57). As can be read from the frequency numbers for each of these markers, all are generally underrepresented in LINDSEI-NO, suggesting that the learners do not fully master this use of the DM *well*. However, the numbers are very low for each category, so more data is required to be able to make general claims about the two speaker groups’ use of *well* for this function.

With regard to transfer, the Norwegian DM *men* (*but*) serving the same function may cause the learners to choose the marker *but* in favor of *well* to serve the interpersonal function ‘turn-taking’. It would be interesting to conduct a similar analysis on the DM *but* to explore this hypothesis, but this lies beyond the scope of this thesis.

Example (5.33) below illustrates the use of *well* with the dual textual and interpersonal function ‘response marker’ and ‘take the turn’. Here the interviewer asks how old the interviewee’s children are. The interviewee begins her answer with *well*, indicating that she will give an answer to the question and that it is her turn to speak now. Example (5.34) shows the same construction, where the interviewer tells the interviewee that he can phrase it in whichever way he wants. Thereafter the interviewee takes the floor by using the DM *well*, and continues his elaboration. The last example demonstrates the use of *well* with the single interpersonal function ‘take the turn’, whereas the first example illustrates the dual textual and interpersonal function ‘response marker’ and ‘take the turn’.

5.33 <A> *how old are they hm how old are they your children* (*eh*) **well** *my youngest boy is nine years old and I have got two girls they are fourteen and sixteen . (N163)*

5.34 *would you like me to sort of speech bubble it sort of <?> put [words into their mouths* <A> [*oh just .. the way you want to* *erm well in the first picture the[i:] artist is erm drawing the[i:] the sitter and then er the sitter the sit= it’s the (E405)*

In sum, this analysis has uncovered that the turn-taker *well* is slightly underrepresented in LINDSEI-NO compared with LOCNEC. The BESs show strong tendencies of using this interpersonal function of *well* together with discourse markers or particles classified as ‘topic-changers’. The NLEs, on the other hand, do not make use of these collocations to the same extent. This discrepancy, together with possible transfer from Norwegian causing the NLEs to use other particles and markers to mark a turn, may be possible explanations for the observed underrepresentation of this use of *well* in the learner-data.

Qualifying functions of *well*

The BESs and NLEs are found to differ more in their use of *well* for qualifying functions than for textual and interpersonal functions. First of all, we see from Table 16 that the NLEs use *well* with qualifying functions much less frequently than the BESs. This observation is in line with Aijmer (2011: 248) who reports that learners typically use DM for attitudinal (qualifying) functions less often than native speakers. Moreover, the results presented in Table 16 show that whereas ‘modifying a previous statement’ is the most frequently used qualifying function in LOCNEC, ‘expressing attitudes, feelings and evaluations’ is the most common function in LINDSEI-NO. The qualifying functions of *well* in general display underrepresentation in LINDSEI-NO as compared to LOCNEC, and so does the individual qualifying functions ‘modifying a previous statement’, ‘confirming a previous statement’ and ‘rejecting a previous statement’. The function ‘expressing disagreement’ is entirely absent in LINDSEI-NO, whereas it is found 6 times in LOCNEC. The relative frequency of each of the qualifying functions discovered in the two corpora is illustrated in Figure 17 below, and an explanation and exemplification of each function is presented in Table 18.

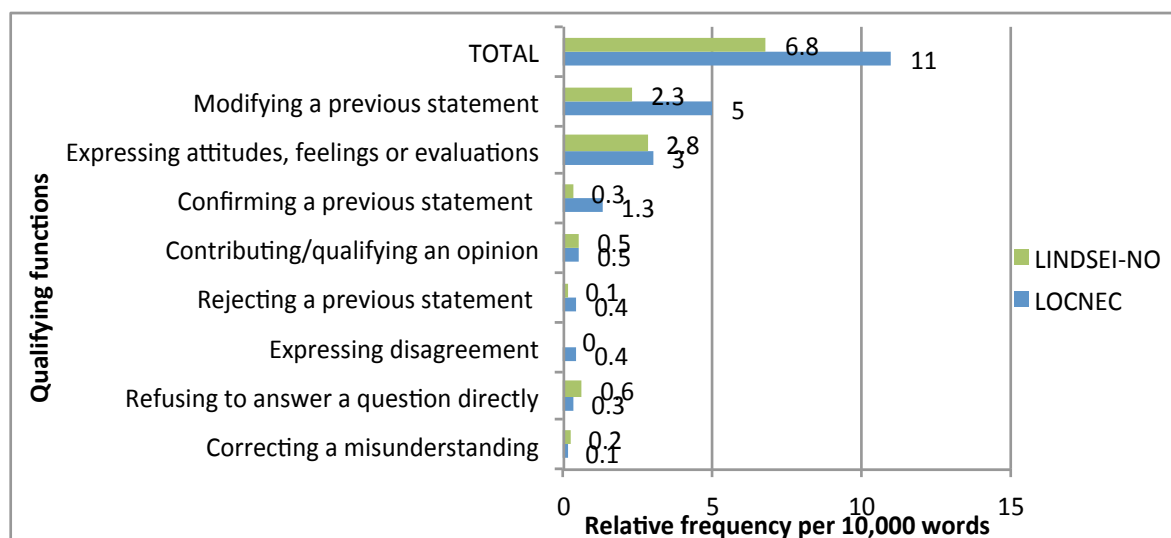


Figure 17: Qualifying functions of *well* in LOCNEC and LINDSEI-NO

Table 18: Definitions and examples of the qualifying function of *well*

Qualifying functions	Definitions	Examples
Modifying a previous statement	<i>Well</i> used to modify a previous statement made either by the speaker or the hearer. Often negative politeness as the speaker uses <i>well</i> to avoid sounding too direct.	<A> you might not have time for anything else because you're a full time student yes yeah well (eh) full time . my course has just . I had sixteen . hours a week (N502)
Expressing attitudes, feelings or evaluations	<i>Well</i> introducing an evaluation or an expression of a feeling or attitude.	I went to Praha that's two couple of years ago and . that people they weren't happy .. and .. well . they seemed depressed in a way (N15)
Confirming a previous statement	<i>Well</i> introducing a confirmation of a previous statement.	there's a sort of gap between them and you I guess well there is a difference yeah a big difference (E464)
Contributing/qualifying an opinion	<i>Well</i> introducing a qualifying opinion to a previous statement or topic.	so I'm not too fond of (eh) samskipnaden but (eh) no well . I can't blame them (N290)
Rejecting a previous statement	<i>Well</i> introducing a rejection of a previous statement. Often negative politeness as the speaker here often uses <i>well</i> to avoid sounding too direct.	 it is a critical course yes and although [I I [<A> you're not really interested in that well n= no it's not that I'm not interested (E869)
Expressing disagreement	<i>Well</i> used to introduce a disagreement to a previous statement. Often negative politeness as the speaker here often uses <i>well</i> to avoid sounding too direct.	<A> cos if you study you tend to meet people from your country again well not especially it depends (E9)
Refusing to answer a question directly	When the speaker is avoiding to answer a question directly following the most obvious intended answer given by the question	<A> so how was the English of the kids you were working with well .. (eh) first of all we noticed that many of the teachers: spoke a lot of Norwegian (N447)
Correcting a misunderstanding	<i>Well</i> introducing a correction to a previous statement. This use of <i>well</i> is often considered negative politeness, as the speaker tries to avoid to insult the hearer.	<A> if you had a choice would you really live on campus or would you prefer living well I did have a choice and I chose to live on campus (E559)

This section will look closer into the two qualifying functions of *well* ‘modifying a previous statement’ and ‘confirming a previous statement’, which are significantly underrepresented in the NLE data of this study. Through an in-depth investigation I hope to be able to point to differences and similarities between BESs and the NLEs, and discuss possible reasons for the observed discrepancy.

Modifying a previous statement:

A total of 114 cases of *well* serving the qualifying function “modifying a previous statement” are identified in my data. Among these, 85 are produced by the BESs (5 pttw), whereas 28 are produced by the NLEs (2.3 pttw). This discrepancy is found to be highly significant.

Looking closer at the 114 cases, I find that this function of *well* is used exclusively with prospective orientation and in the pre-front field. It typically follows a question, and therefore also serves the textual function ‘response marker’ (LOCNEC n=45, 2.6 pttw; LINDSEI-NO n=17, 1.4 pttw). It is often also used with the interpersonal function ‘hesitator’ (LOCNEC n=5, 0.3 pttw; LINDSEI-NO n=5, 0.4 pttw), indicating that the speaker tries to be careful when modifying what the other person just said to avoid insulting him. Example (5.35) illustrates this use of *well* where the DM serves the dual textual and qualifying functions ‘response marker’ and ‘modifying a previous statement’. Here, the speaker uses *well* to introduce a modification of the interlocutor’s previous statement saying that he or she gets to travel a lot. This is a misconception, according to the speaker, and must therefore be modified.

5.35 *and . that was it so I really wanna go back there as well <A> (mm) but it sounds like you get to do a lot of travelling . (um) **well** basically I’ve been travelling to the same place because all our friends are still in Spain where we lived for a year (N425)*

Example (5.36) shows the use of *well* serving both the textual function ‘introducing clarification’ and the interpersonal function ‘modifying a previous statement’ simultaneously. In this case the speaker is modifying his own statement by first saying that she has got two younger siblings, but then modifying it by clarifying that her sister is only a year younger. This example demonstrates that this function of *well* can also be used to modify a speakers’ own statements.

5.36 *work environment mhm you know I’m one of four children <A> oh yes so it’s and perhaps they’re younger . I’ve got two younger **well** my sister’s only a year younger but she’s got her boyfriend as well [uhu [so that <?> he practically lives with us mhm (E1157)*

The examination of the 114 cases of *well* functioning to ‘modify a previous statement’ has not revealed any noteworthy differences between the two speaker groups except for the fact that this qualifying function of *well* is underrepresented in LINDSEI-NO as compared to LOCNEC. In searching for an explanation for this finding, one may therefore have to turn to what Müller (2005) refers to as “non-linguistic factors”. These can be factors such as age, gender, relationship between the two speakers, native language etc. It is arguably the case that non-native speakers use *well* with the qualifying function ‘modifying a previous

statement' less frequently than the native speakers because they tend to be more hesitant and careful in their speech (Müller, 2004). This has to do with the confidence the speakers feel when speaking in a second language, and with the fact that they want to avoid insulting the hearer – especially if they do not know the person that well. It is furthermore possible that the NLEs show transfer from their native language. In Norwegian it is common to use other markers such as *nja* and *tja* to express modification of a previous statement in addition to *vel*. It may therefore be the case that the NLEs use other means for expressing this function than the DM *well*. As these factors are difficult to measure within the scope of this thesis, these points will remain ideas and suggestions for future studies.

Confirming a previous statement:

Another qualifying function yielding different and interesting results for the two speaker groups is 'confirming a previous statement'. As presented in Table 18 above, this use of *well* occurs when the speaker uses the discourse marker to confirm a previous statement expressed by either the speaker or the hearer. This use of *well* occurs a total of 26 times in the data of this study, where 22 hits are identified in LOCNEC (1.3 pttw) and only 4 come from LINDSEI-NO (0.3 pttw). The difference is considered highly significant²⁵, but the effect size value of -4.05 indicates that the results cannot be considered strong enough, and rather serve as positive evidence in favor of the H0.

A closer examination of how the 26 cases of *well* are used by the two speaker groups uncovers that all cases are used with prospective orientation and occur exclusively in the pre-front field. Furthermore, this function of *well* most commonly occurs with the dual textual and qualifying functions 'response marker' and 'confirming a previous statement' (LOCNEC n=19, 1.1 pttw; LINDSEI-NO n=4, 0.3 pttw), where the speaker through a response confirms something said by the interlocutor. This is illustrated by example (5.37) and (5.38) below.

5.37 <A> (mm) .. because you're also a full time student yeah <A> so so in other words you're very busy **well** it's: enough to do yeah . <A> and then you but you want to become a a teacher which is why you're here yeah . that's why I'm (N108)

5.38 <A> so you want to work there yeah so you must be keeping contact in contact with them **well** I've seen erm one of the guys around in Manchester and he said oh you must come back and I was intending to do that over (E879)

²⁵ More than 99% certain.

The in-depth analysis of the DM *well* used to confirm a previous statement has not yielded any noteworthy results in terms of differences between the BESs and NLEs. Apart from the fact that this function of *well* is strongly underrepresented in the learner-data, both groups appear to use the DM similarly for this function. This means that there are no obvious linguistic factors that can explain the difference in frequency. As was the case with the function ‘modifying a previous statement’, it is also here more likely that factors such as age, gender, native language, relation to the hearer etc. may serve to explain the discrepancy. It may for instance be that Norwegian IL speakers rather prefer to use markers and particles such as *right*, *yes* and *agree* to confirm a previous statement, as the Norwegian equivalents *stemmer*, *ja* and *enig* typically are used for this purpose. These are only speculations, and further research is needed to examine this in more depth.

5.4 *You know*

The third, and final, part of the analysis is concerned with the discourse marker *you know* and how the British RLV speakers and Norwegian IL speakers use it. As in the two previous subsections on *like* and *well*, this section starts with a quantitative analysis, where the two speaker groups’ use of *you know* in terms of overall frequency, utterance position and orientation is compared. Thereafter, the next section presents the qualitative functional analysis of *you know*, discussing the use of this DM serving at the textual, interpersonal and qualifying functional level.

5.4.1 Frequency analysis

The preliminary analysis for this study is presented in Table 8 in section 5.1, where it is reported that a total of 632 hits of the DM *you know* are identified in LOCNEC (37.1 pttw), and 142 in LINDSEI-NO (11.5 pttw). The statistical calculations indicate that this difference is highly significant²⁶. Entering the detailed analysis of the DM *you know* with this background information together with findings from previous research, I expect to uncover differences between the Norwegian learners and the British English speakers in their use of the DM *you know*, and discuss possible reasons to why the learners underrepresent this discourse marker so significantly.

²⁶ More than 99,99% certain.

Utterance position of *you know*

The first factor that is examined is the positioning of *you know* in the utterance. In contrast to *like* and *well*, *you know* is frequently found in the post-end field with retrospective orientation in both LOCNEC and LINDSEI-NO. The numbers of instances *you know* occurs in terms of these three positions are presented in Table 19.

Table 19: Utterance position of *you know* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG-LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Pre-front field	406	23.8	94	7.6	121.35	< 0.0001	108.76
Wildcard	25	1.5	7	0.6	5.72	< 0.5	-6.87
Post-end field	201	11.8	41	3.3	69.52	< 0.0001	56.93
TOTAL	632	37.1	142	11.5	195.94	< 0.0001	183.35

You know occurring in the pre-front field is mostly associated with the textual functions ‘move (back) to story’, ‘introduce explanation’, ‘introduce given information’, and ‘quotative *you know*’ and the interpersonal functions ‘hold the floor’, ‘plea for understanding’, ‘attention-getter’ and ‘hesitator’. *You know* functioning as a wildcard is most frequently found with the textual function ‘pausal interjection’ and the interpersonal functions ‘searching for the right word/phrase’ and ‘plea for cooperation/confirmation’. *You know* occurring in the post-end field is typically connected with the textual functions ‘introduce given information’ and ‘closing a point’ and the interpersonal functions ‘plea for cooperation/confirmation’, ‘plea for understanding’, and ‘attention-getter’.

The three examples below illustrate the use of *you know* in the three different utterance positions. Example (5.39) shows a case where the DM *you know* occurs in the pre-front field, serving the dual textual and interpersonal function of ‘move (back) to story’ and ‘attention-getter’. Example (5.40) illustrates *you know* occurring as a wildcard, and in this case *you know* serves the dual textual and interpersonal function ‘pausal interjection’ and ‘hesitator’. Example (5.41) shows the use of *you know* in the post-end field, serving the dual textual and interpersonal functions ‘closing a point’ and ‘see the implication’. Although the DM in example (5.40) and (5.41) is preceded by a conjunction suggesting the beginning of a new clause and thus the position of the DM as pre-front, listening to the sound recordings revealed that these cases illustrate the use of *you know* with wildcard and post-end position.

- 5.39 yeah but it was pretty awesome yeah yeah <A> no . that's true .. but **you know** when she was like talking in one of our . in one of our ears (eh) in one of my ears and . I was listening to (N21)
- 5.40 (em) maybe at first . but if . if you pursue the relationship . it will last . <A> yeah (mhm) yeah . maybe it's because **you know** . because it was so cold in Norway people had to stay inside and so they became more . introvert= (N17)
- 5.41 article writing [something not more like for a women's magazine something that is interesting but doesn't take so long and . **you know** <A> yeah [ah yeah oh [erm .. I worked in a petshop for four years erm . I worked in a freezer centre for a year (E25)

Figure 18 below provides a comparative illustration of the utterance position of *you know* in LOCNEC and LINDSEI-NO.

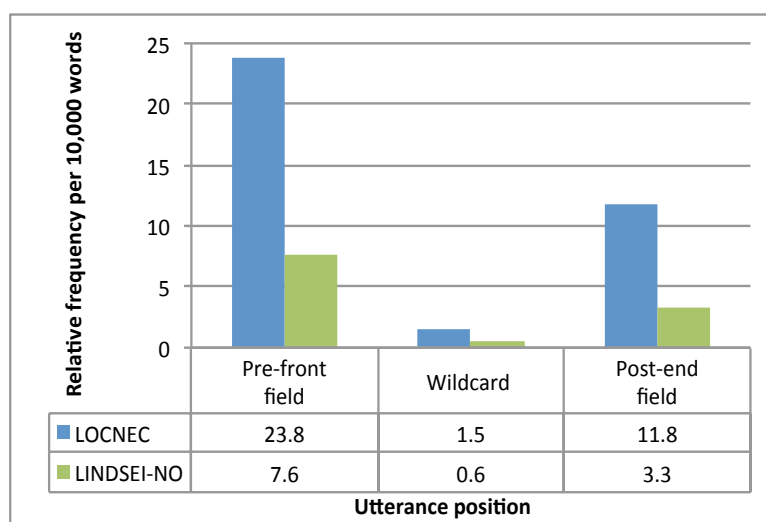


Figure 18: Utterance position of *you know* in LOCNEC and LINDSEI-NO

Figure 18 demonstrates that both speaker groups use *you know* most frequently in the pre-front field. This use of *you know* is found 406 times in LOCNEC (23.8 pttw) and 95 times in LINDSEI-NO (7.6). Looking at the statistical calculations for these results, we see that the differences between the NLEs and BESs in terms of the DM's position in the pre-front field and post-end field are highly significant²⁷. Concerning *you know* occurring in the wildcard position, on the other hand, the difference between the two speaker groups is not as strong. For this position, the Log-Likelihood value is only 5.72, which yields a p-value of 0.5. Although this initially is a significant result, the effect size tells us that the result is not strong enough to reject the H0, and rather serves as positive evidence in favor of it. More data would perhaps yield stronger results.

²⁷ More than 99.99% certain.

In sum, the analysis of the utterance position *you know* in LOCNEC and LINDSEI-NO has yielded significant results confirming that the NLEs use *you know* occupying the pre-front and post-end field less frequently than their British peers. As described above, the different positions in the utterance are associated with different textual and interpersonal functions. The functional analysis will examine whether this underrepresentation of the orientation of *you know* in LINDSEI-NO also is mirrored by one or more of these pragmatic functions.

Orientation of *you know*

The second factor examined in this frequency analysis is the orientation of the DM *you know* in the utterance. Prospective *you know* typically occurs in the pre-front field or as a wildcard, whereas retrospective *you know* always occurs in the post-end field. The most typical textual functions associated with prospective *you know* are ‘move (back) to story’, ‘introduce explanation’ and ‘introduce given information’, and the most typical interpersonal functions are ‘attention-getter’, ‘hesitator’, ‘hold the floor’ and ‘plea for cooperation/confirmation’. For prospective *you know*, the most typical textual functions are ‘closing a point’ and ‘introduce given information’, whereas the most typical interpersonal functions are ‘attention-getter’ and ‘plea for cooperation/confirmation’.

Example (5.42) shows the use of *you know* with prospective orientation serving the dual textual and interpersonal functions ‘introducing explanation’ and ‘plea for understanding’. In this sentence, *you know* serves to introduce the part where the speaker says that she only knows a few songs. She pleads for the interlocutor to understand what she means when she says she has been playing some guitar. Example (5.43) illustrates *you know* used with a retrospective orientation serving the dual textual and interpersonal functions ‘introduce given information’ and ‘plea for cooperation/confirmation’. Here, the speaker refers to his friends in Madrid, of which he assumes that the hearer knows. He uses *you know* to indicate meta-knowledge and indirectly pleads for a confirmation from the hearer that his assumption is correct.

5.42 *yeah yeah oh well I: have been playing some guitar but . I'm not a: . guitarist just (eh) **you know** . know a few songs but (eh) I've als= always been singing . in choirs and . every family gathering there is (eh) . (N49)*

5.43 *yeah oh that's that's funny cos I know I know my friends in Madrid **you know** they live in the city or live near and they work go to the[i:] university (E342)*

Table 20 below presents an overview of the distribution of all instances of *you know* as a DM in LOCNEC and LINDSEI-NO in terms of orientation. Figure 19 provides a graphical comparison of the two speaker groups with regard to the orientation of *you know*.

Table 20: Orientation of *you know* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG-LIKELIHOOD	P-VALUE	EFFECT SIZE
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words			
Prospective	436	25.6	101	8.2	130.23	< 0.0001	117.64
Retrospective	196	11.5	41	3.3	65.96	< 0.0001	53.37
TOTAL	632	37.1	142	11.5	195.94	< 0.0001	183.35

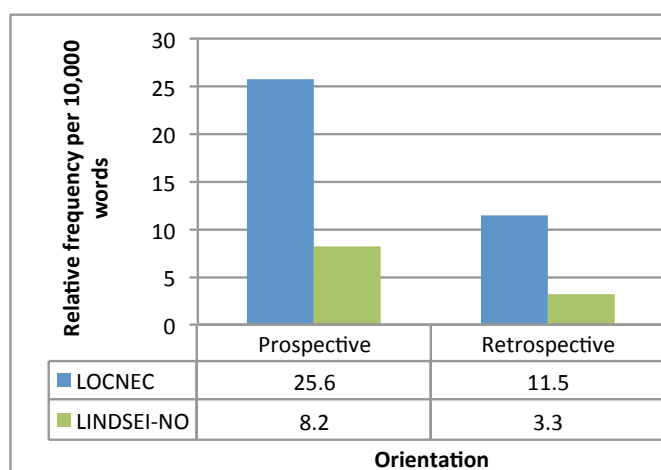


Figure 19: Orientation of *you know* in LOCNEC and LINDSEI-NO

The results in Table 20 indicate that both speaker groups use *you know* most frequently prospectively (LOCNEC $n=436$, 25.6 pttw; LINDSEI $n=101$, 8.2 pttw). The difference between the NLEs and BESs is found to be highly significant²⁸, and the effect size further confirms that this result serves as strong evidence against the H0.

Also with the retrospective use of *you know* a difference between the two speaker groups is detected. This orientation of *you know* is found 196 times in LOCNEC (11.5 pttw) and 41 times in LINDSEI-NO (3.3 pttw), which again suggests an underrepresentation in the NLE data. Looking at the P-value and effect size value for this result in Table 20, we see that this result is highly significant²⁹ and serves as strong evidence against the H0.

²⁸ More than 99.99% certain.

²⁹ More than 99.99% certain.

5.4.2 Functional analysis

The frequency analysis presented in section 5.4.1 has yielded noteworthy results indicating that *you know* is generally significantly underrepresented in LINDSEI-NO as compared to LOCNEC. The frequency analysis presented above has shown that this tendency is particularly convincing in the pre-front and post-end fields and when the DM is used prospectively. Table 21 below presents a full overview of the results from the functional analysis of *you know*.

Table 21: Textual, interpersonal and qualifying functions of *you know* in LOCNEC and LINDSEI-NO

	LOCNEC		LINDSEI-NO		LOG- LIKELIHOOD	P-VALUE	EFFECT SIZE	
	Raw frequency	Per 10,000 words	Raw frequency	Per 10,000 words				
TEXTUAL FUNCTIONS	Transition/ topic shift	19	1.1	4	0.3	6.35	< 0.05	-6.24
	Introducing explanation, justification, clarification, conclusion, consequence and further specification	117	6.8	34	2.8	25.19	< 0.0001	12.6
	Introducing given information	37	2.2	20	1.6	1.11	n.s.	-11.48
	Rephrase/repair	34	2	7	0.6	11.64	< 0.001	-0.95
	Move (back) to story	74	4.3	4	0.3	55.82	< 0.0001	43.23
	Closing a point	10	0.6	0	0	10.87	< 0.001	-1.72
	Other	42	2.5	18	1.5	3.64	n.s.	-8.95
	TOTAL	333	19.5	87	7.1	84.63	< 0.0001	72.04
INTERPERSONAL FUNCTIONS	Attention-getter	144	8.5	23	1.9	62.61	< 0.0001	50.02
	Searching for the right word/phrase	56	3.3	14	1.1	15.14	< 0.0001	2.55
	Plea for understanding	59	3.5	29	2.3	2.98	n.s.	-9.61
	"See the implication"/ unfinished point	30	1.8	4	0.3	14.93	< 0.001	2.34
	Take the turn	19	1.1	1	0.1	14.45	< 0.001	1.86
	Hold the floor	53	3.1	4	0.3	35.59	< 0.0001	23
	Hesitator/Hedger	45	2.7	11	0.9	12.54	< 0.001	-0.05
	Plea for cooperation/ confirmation	177	10.4	45	3.7	46.76	< 0.0001	34.18
	Other	14	0.9	4	0.5	3.1	n.s.	-9.49
	TOTAL	597	35.3	135	11.1	183.67	< 0.0001	171.08
QUALIFYING FUNCTIONS	Expressing attitudes, feelings or evaluations	122	7.2	22	1.8	47.72	< 0.01	35.13
	Other	2	0.2	0	0	2.17	n.s.	-10.42
	TOTAL	124	7.4	22	1.8	49.23	< 0.0001	36.64

Textual functions of *you know*

The textual functions identified for *you know* in LOCNEC and LINDSEI-NO are as follows: ‘Introducing explanation, justification, clarification etc.’ (Holmes, 1986; Fox Tree and Schrock, 2002), ‘introducing given information’ (Östman, 1981; Schiffrin, 1987), ‘rephrase/repair’ (Holmes, 1986; Müller, 2005), ‘transition/ topic shift’ (Erman, 2001; Fox Tree and Schrock, 2002), ‘closing a point’ (Fox Tree and Schrock, 2002) and ‘pausal interjection’ (Östman, 1981). In addition, some textual functions that have not been listed for *you know* in previous research, were also identified, viz. ‘move (back) to story’, ‘exemplifier’, and ‘quotative *you know*’. However, the latter two yielded too few results to be included in the analysis, and have been placed in the category ‘other’ together with the number of occurrences of ‘pausal interjection’ and some miscellaneous cases. Table 22 provides a short definition and an example of each textual function.

Table 22: Definitions and examples of the textual functions of *you know*

Textual function	Definition	Example
Introducing explanation, justification, clarification etc.	When <i>you know</i> introduces an explanation, justification, clarification, further specification, conclusion or consequence/result.	<i>my friend Belinda twisted her ankle so oh she then had to go to hospital but you know for the first few hours it was really good fun (E96)</i>
Move (back) to story	When <i>you know</i> is used to move back to the topic the speaker initially talked about.	<i>that's not my way of saying things yeah so but you know we're parrots yeah and we just copy what we hear (N25)</i>
Introducing given information	When <i>you know</i> introduces what the speaker believes that the hearer already knows something about.	<i>before I'd done a number of eh musicals like you know Grease Oliver .. Little Shop of Horrors (E331)</i>
Rephrase/repair	When the speaker uses <i>you know</i> to interrupt an already started utterance in order to repair something and start over again.	<i>and we'd drive from you know we'd take the ferry from Oslo to Kiel (N61)</i>
Transition/topic shift	When <i>you know</i> introduces a transition or a new topic.	<i>I just wanna go to bed . please you know when I came I was like okey I have jetlag (N100)</i>
Closing a point	When <i>you know</i> indicates that the argument or point made by the speaker is complete.	<i>and then I arr= arrived and it's not quite so wonderful after all you know (E8)</i>
Other	Includes cases of "quotative <i>you know</i> " where <i>you know</i> is used to indicate direct speech, feelings or thoughts, "pausal interjection", where <i>you know</i> is used to stall for time, and "exemplifier" where <i>you know</i> is used to introduce an example.	<i>I rang them up and you know you've not been in touch and oh we've been told not to take on anymore students (E34) // maybe it's because . you know because it was so cold in Norway (N17) // it's . really easy to travel around you can you know hire (eh) a cab driver for a day (N28)</i>

Figure 20 provides an illustration of the frequency of *you know* per 10,000 words in LOCNEC and LINDSEI-NO. As the figure shows, the overall use of *you know* with textual functions is much lower in LINDSEI-NO than in LOCNEC. Looking at Table 21 above, we see that the difference between the two English varieties in terms of the overall use of *you know* with textual functions is highly significant³⁰. According to the effect size value, these results serves as strong evidence against the H0, i.e. the two speaker groups differ greatly in their use of *you know* with textual functions.

Concerning the different textual functions identified in LOCNEC and LINDSEI-NO, Table 21 presents ‘introducing explanation, justification, clarification etc.’ as the most frequently used textual function in LOCNEC, followed by ‘move (back) to story’, ‘introduce given information’ and ‘rephrase/repair’. In LINDSEI-NO, on the other hand, the most frequently used textual functions for *you know* are ‘introduce explanation, justification, clarification etc.’, ‘introduce given information’, and ‘rephrase/repair’. The textual function of *you know* most underrepresented in LINDSEI-NO as compared to LOCNEC is ‘move (back) to story’, followed by ‘introducing explanation, justification, clarification etc.’. A closer investigation of the two latter functions will be conducted in the following two sections.

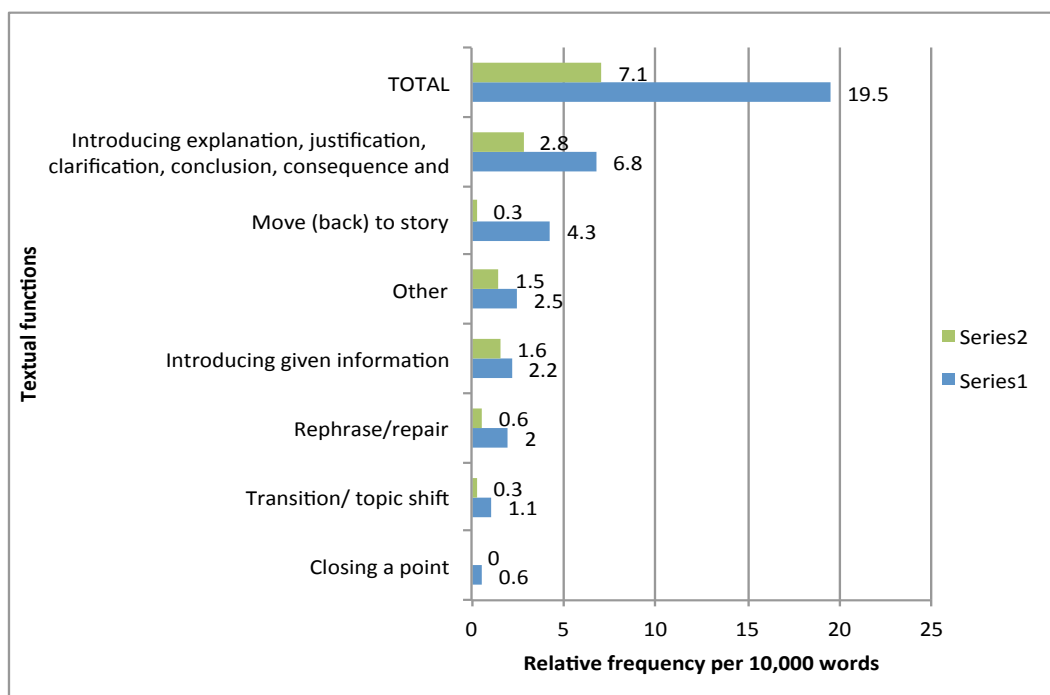


Figure 20: Textual functions of *you know* in LOCNEC and LINDSEI-NO

³⁰ More than 99,99% certain.

Move (back) to story:

As described in Table 22 above, the textual function ‘move (back) to story’ is used when the speaker feels the need to draw the attention back to the main topic. This is typically necessary in situations where the hearer has made a comment or a response, or where there has been need for an explanation or other types of digressions. Example (5.44) below shows a case where the speaker is making a comment about her own way of pronouncing English words, which has little to do with her general argument. The speaker therefore uses *you know* to direct the conversation back to the main topic.

5.44 *of voice you know intonation thing (eh) and I oh why w= why do I say that . that's not my way of saying things yeah so but **you know** we're parrots yeah and we just copy what we hear around us yeah (uhu) ... okay (mm) may I just (N25)*

Among the total 78 hits of *you know* functioning to ‘move (back) to story’, only 4 were uttered by the NLEs (0.3 pttw), leaving 74 for the BESs (4.3 pttw). This discrepancy is considered highly significant³¹.

In order to investigate why the NLEs use this textual function of *you know* less frequently than the BESs, the 78 cases have been examined in terms of collocation patterns and dual interpersonal functions. What stands out in this analysis, is that the use of *you know* to indicate the textual function ‘move (back) to story’ is most typically preceded by conjunctions such as *and* (LOCNEC n=17, 1 pttw; LINDSEI-NO n=0), *but* (LOCNEC n=6, 0.4 pttw; LINDSEI-NO n=3, 0.2 pttw) and *so* (LOCNEC n=7, 0.4 pttw; LINDSEI-NO n=0). In terms of dual interpersonal functions, the cases of *you know* serving the textual function ‘move (back) to story’ also serve the interpersonal functions ‘hold the floor’ (LOCNEC n=39, 2.3 pttw; LINDSEI-NO n=1, 0.1 pttw) and ‘attention-getter’ (LOCNEC n=19, 1.1 pttw; LINDSEI-NO n=2, 0.2 pttw). Example (5.45) below illustrates a case where *you know* serves a dual function as the textual function ‘move (back) to story’ and the interpersonal function ‘hold the floor’, and where the DM is preceded by the conjunction *so*.

5.45 *missed they flight back to England because they just didn't realize <A> yeah yeah erm .. so **you know** it's it's difficult to see some of the things sometimes . but it's a nice place it's really nice it's beautiful (E509)*

³¹ More than 99.99% certain.

To sum up, there is little evidence in LINDSEI-NO suggesting that the NLEs have acquired the use of *you know* for the textual function “move (back) to story”. The in-depth investigation has revealed strong patterns for this use of *you know* in LOCNEC where it is typically preceded by certain conjunctions and where it typically serves a dual textual and interpersonal function. In order for the learners to acquire this use of *you know*, it may be beneficial for them to get more input of such cases so that they can develop an understanding of how it is used.

Introducing explanation (and justification, clarification, conclusion, consequence etc.):

The functional analysis yielded a total of fourteen different textual functions for *you know*. In order to shorten this list and make it more comprehensible, it was decided to combine the functions ‘introducing explanation’, ‘introducing justification’, ‘introducing clarification’, ‘introducing conclusion’, ‘introducing consequence/result’ and ‘introducing further specification’ into one group. ‘Introducing explanation’ is the textual function that yielded the highest frequency numbers among these six (LOCNEC n=91, 5.3 pttw, LINDSEI-NO n=23, 1.9 pttw), whereas the remaining five functions generated results below 12 hits in both corpora. Combining them into one function did not cause any dramatic changes to the statistical results, and was therefore considered a better solution.

As presented in Table 22 above, this functional group is defined as cases of *you know* where the DM is used to introduce clauses that convey additional information in terms of explanations, clarifications, justifications and further specifications, or where the DM is used to close a point by introducing a conclusion, consequence or result. In other words, this is a broad category that contains several different uses of *you know*. Examples (5.46)-(5.48) below illustrate some of the varieties mentioned above. (5.46) presents a case where *you know* is used to introduce an example. In this situation *you know* also serves the interpersonal function “hesitator” at the same time. Example (5.47) illustrates the use of *you know* to introduce a clarification, and example (5.48) shows the use of *you know* to introduce further specification.

5.46 *it's kind of a hard decision .. because: **you know** you take them out of their (em) . environment the close: .. yeah (N16)*

5.47 *erm I was probably going to teach English [**you know** as an assistant <A> [oh yeah in a French school mhm but er it [didn't work out (E161)*

5.48 and (eh) have the . cars and bikes from . the same time period . **you know** (eh) . late sixties and early eighties (N98)

The 114 (LOCNEC n=91, 5.3 pttw; LINDSEI n=23, 1.9 pttw) cases of *you know* used to introduce an explanation show a tendency of being preceded by conjunctions such as *because/cos* (LOCNEC n=9, 0.5 pttw; LINDSEI-NO n=2, 0.2 pttw), *but* (LOCNEC n=5, 0.3 pttw; LINDSEI-NO n=1, 0.1 pttw) and *and* (LOCNEC n=2, 0.1 pttw; LINDSEI-NO n=0). *You know* serving this function is also sometimes preceded by other discourse markers such as *like* (LOCNEC n=8, 0.5 pttw; LINDSEI-NO n=0), *just* (LOCNEC n=4, 0.2 pttw; LINDSEI-NO n=0) and *I mean* (LOCNEC n=3, 0.2 pttw; LINDSEI-NO n=0). The remaining hits are mainly preceded by pause fillers such as *erm*, *er*, *em*, and *mhm*. The numbers for these collocations are very low, but they indicate that the British speakers display a slightly more stable and systematic usage of *you know* for this textual function than the Norwegian learners.

Looking at the interpersonal functions that combine with these cases of *you know*, we can observe that the NLEs typically use *you know* with the dual textual and interpersonal functions ‘introducing explanation’ and ‘plea for understanding’ (n=7, 0.6 pttw) and ‘plea for cooperation/confirmation’ (n=6, 0.5 pttw). Although such dual textual and interpersonal functions are also identified in LOCNEC, the BESs demonstrate more diversity here, with *you know* serving the textual and interpersonal functions ‘introducing explanation’ and ‘attention-getter’ (n=17, 1 pttw), ‘searching for the right word/phrase’ (n=14, 0.8 pttw), ‘hesitator’ (n=6, 0.4 pttw), ‘see the implication’ (n=6, 0.4 pttw) and ‘take the turn’ (n=4, 0.2 pttw). This observation indicates that the learners are not familiar with all possible ways of using *you know* to introduce an explanation. It is arguable that input and explicit teaching of this usage may increase the learners’ ability, and thereby reduce the overall underrepresentation of *you know* with this function in the learner data.

Interpersonal functions of *you know*

In the functional analysis of *you know*, a total of 13 interpersonal functions have been identified in LOCNEC and LINDSEI-NO. These are ‘approximator’, ‘emphasizer’, ‘hedger’, ‘attention-getter’, ‘searching for the right word/phrase’, ‘“imagine the scene”’, ‘plea for understanding’, ‘“see the implication”/unfinished point’, ‘take the turn’, ‘hold the floor’, ‘hesitator’, and ‘plea for cooperation/confirmation’. The two functions ‘approximator’ and

'emphasizer' yield frequencies below 10 in both corpora, and have therefore been placed in the category 'other'. It has been decided to merge the two functions 'hesitator' and 'hedger', as the cases belonging to these categories appeared to be quite similar. Thus, a total of 9 interpersonal functions have been identified for the DM *you know* in LOCNEC and LINDSEI-NO, which are defined and exemplified in Table 23 below.

Table 23: Definitions and examples of interpersonal functions of *you know*

Interpersonal function	Definition	Example
Attention-getter	When <i>you know</i> appeals for the hearer's attention, indicating that what comes next is important.	<i>because sometimes you know they add things</i> (E528)
Searching for the right word/phrase	When <i>you know</i> is used to stall for time while the speaker is searching for what to say or how to phrase it.	<i>there are four different ones and . One is particularly good with . (eh) . you know . alternative (eh) gas like . ethanol</i> (N43)
Plea for understanding	When <i>you know</i> is used to ask for the hearer to understand what the speaker means.	<i>I miss: Australia I wanna og back but you know . being a student and the tickets (em) . a bit expensive</i> (N22)
"See the implication" / unfinished point	When the speaker stops in the middle of a sentence and uses <i>you know</i> to indicate that it is superfluous to continue either because the hearer knows the rest, or because the speaker does not quite know how to phrase it. The hearer is thus left to think for himself what the speaker intended to say.	<i>and then (eh) fall came and . you know . (eh) it's a lot . a lot of things that's more fun to pend money on than buying tires</i> (N9)
Take the turn	When <i>you know</i> is used to regain the floor/turn.	<i>what happened to her in the film is untrue like she falls in love with Wallis and [yes but [you know you know they have to add [some romance</i> (E354)
Hold the floor	When <i>you know</i> is used to indicate that the speaker is not finished yet and that something more is to come.	<i>I decided to carry on with it and you know .. unless you can do it properly so I thought I'd do a year abroad</i> (N12)
Hesitator/hedger	When <i>you know</i> is used with a face-saving intention to indicate that the speaker is not entirely sure about what s/he is about to say, or when the speaker doesn't want to be entirely exact about the details. Can be paraphrased with "kind of" and "don't take it literally".	<i>cars and bikes from . the same time period . you know (eh) . late late sixties and early eighties</i> (N98)
Plea for cooperation/confirmation	When the speaker is asking the hearer to confirm that s/he follows the argument, knows what the speaker is talking about, or agrees with what the speaker says. Often used with face-saving intentions, i.e. the speaker wants to be sure that the hearer agrees with an argument before s/he continues.	<i>I mean Brad Pitt's normally sort of quite light headed you know sort of light weight but he is brilliant</i> (E225)
Other	Cases where <i>you know</i> functions as an emphasizer, approximator or an indicator of "imagine the scene". The latter function is used to make the hearer envision a situation.	<i>I think another reason was I mean everyone . you know everyone's been to see it (E188) // where's the rest of the you know a hundred and . eighty five people (N123) // bying a while chicken and carrying it home by its feet you know</i> (N195)

Figure 21 below demonstrates how often the two speaker groups use *you know* with interpersonal functions, presented in relative frequency per 10,000 words. What the figure undoubtedly reveals is an underrepresentation of all interpersonal functions in the NLE data. As Table 21 above demonstrates, interpersonal *you know* is used 597 times (35.3 pttw) in LOCNEC. In LINDSEI-NO, on the other hand, it is only found 135 times (11.1 pttw). The LL and effect size calculations indicate that this discrepancy is highly significant³², and that it serves as strong evidence against the H0.

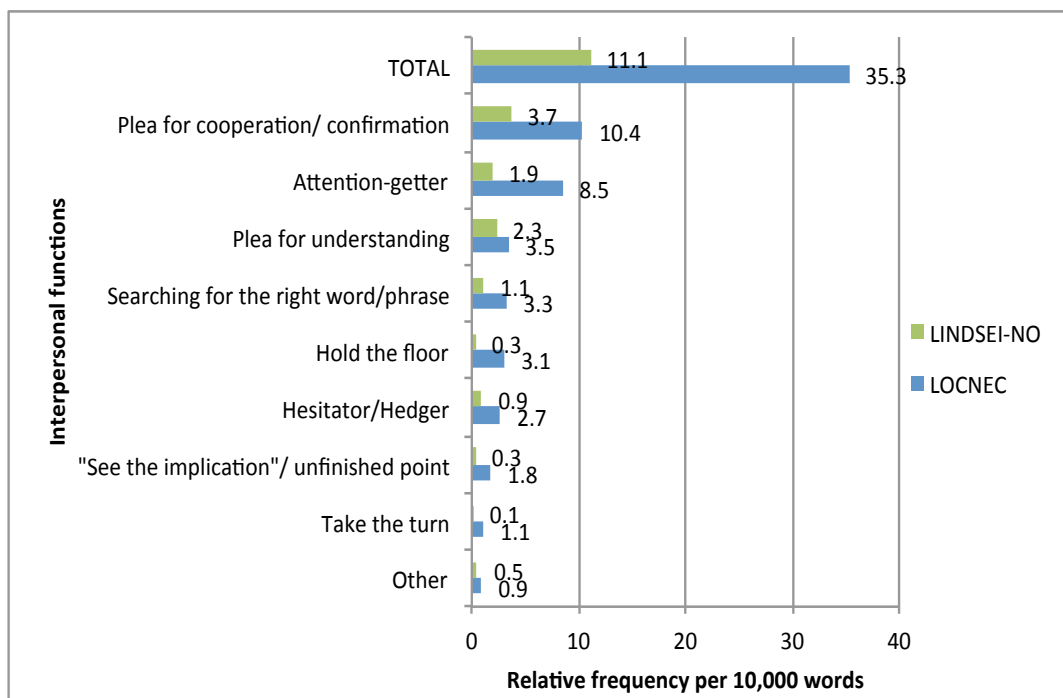


Figure 21: Interpersonal functions of *you know* in LOCNEC and LINDSEI-NO

The three interpersonal functions yielding the most interesting results in terms of underrepresentation in the learner data, ‘attention-getter’, ‘plea for cooperation/confirmation’ and ‘hold the floor’, have been selected for further investigation.

Attention-getter:

y'know marks the speaker as an information provider, but one whose successful fulfillment of that role is contingent upon hearer attention. (Schiffrin, 1987: 290)

As pointed out, *you know* is a hearer-oriented discourse marker. Consequently, it is often used to include the hearer in the conversation by appealing for his or her attention or

³² More than 99.99% certain.

feedback, or to offer him or her the turn. One way of doing this is to use *you know* as an ‘attention-getter’. As shown in Table 21 and Figure 21, ‘attention-getter’ is the second most frequent interpersonal function in LOCNEC only preceded by ‘plea for cooperation/confirmation’, and the third most frequently used interpersonal function in LINDSEI-NO. Surprisingly, it is the most underrepresented interpersonal function in the NLE data. As a matter of fact, it is the most underrepresented function of *you know* overall, followed by the textual function ‘move (back) to story’. With this in mind, I find this use of *you know* worth a closer look in order to reveal how the two speaker groups use this DM differently.

You know as an “attention-getter” is identified 144 times in LOCNEC (8.5 pttw), whereas it is only found 23 times in LINDSEI-NO (1.9 pttw). The result from the statistical calculations presented in Table 21 show that this discrepancy is highly significant³³, and that the result serves as strong evidence against the H0.

The examination of these cases reveals that, in LOCNEC, the DM most often serves the dual textual and interpersonal functions ‘move (back) to story’ and ‘attention-getter’ (n=18, 1.1 pttw), or ‘introducing explanation, justification and further specification’ and ‘attention-getter’ (n=23, 1.3 pttw). These uses are illustrated in examples (5.48) and (5.49) below, respectively. This observation corresponds to Schiffrin’s (1987: 284) finding that *you know* is frequently found in narratives where the speaker is telling a story and uses the discourse marker to highlight important points. Example (5.49) illustrates a case where the speaker is telling a story, and uses *you know* to introduce an explanation containing important information, and therefore also requiring the hearer’s attention. Example (5.50) displays a similar usage of *you know*. Here the speaker attempts to move on with the story after a short digression, and uses the DM to gain the hearer’s attention. Surprisingly, this dual textual-interpersonal function where *you know* serves as an ‘attention-getter’ and indicator of ‘move (back) to story’ only occurs once in LINDSEI-NO.

5.49 when he got to: when he he yeah got out of the building mm of the the police .. [thing <A> [yeah when . cos **you know** the camera was staying on him a bit too long yeah it was [yeah [there was something wrong with him (E116)

³³ More than 99.99% certain.

5.50 *really well with them .. and er a nice house it's always a bit of a tip but . but you know I'm I'm really enjoying it <A> mhm but **you know** there's hassles with getting on busses and yeah I've gotta walk across that bridge (E83)*

An examining the 167 cases reveals that *you know* is most often used prospectively (LOCNEC n=85, 5 pttw; LINDSEI-NO n=20, 1.6 pttw), and less frequently with retrospective orientation (LOCNEC n=56, 3.3 pttw; LINDSEI-NO n=3, 0.2 pttw). As can be seen from the numbers presented here, the retrospective use of *you know* functioning as an 'attention-getter' is almost entirely absent in LINDSEI-NO, suggesting that the learners have not fully acquired the use of *you know* with this orientation.

The orientation of *you know* is closely related to the DM's utterance position. Since the 'attention-getter' *you know* was found less frequently with retrospective orientation, it is also found to occur less frequently in the post-end field (LOCNEC n=59, 3.5 pttw; LINDSEI-NO n=3, 0.2 pttw) than in the pre-front field (LOCNEC n=81, 4.8 pttw; LINDSEI-NO n=19, 1.5 pttw). LOCNEC also discloses three cases where the 'attention-getter' *you know* occurs as a wildcard (0.2 pttw), whereas this position is absent in LINDSEI-NO. Example (5.51) below illustrates the use of 'attention-getter' *you know* with retrospective orientation and post-end field position. Example (5.52) illustrates the same use of *you know* with prospective orientation and wildcard position.

5.51 *we couldn't walk anywhere . we I (eh) noticed a sign . (eh) it was a big sign . beware pedestrians could come here **you know** so it was it was quite st= strange . I think because in Norway we we usually take care of the pedestrians (N69)*

5.52 *[oh yes .. yeah yeah because sometimes **you know** they add things or a lot was true but also what they said about erm Edward being married to Elanor of Aquitaine (E528)*

In sum, the study of 'attention-getter' *you know* uncovers some discrepancies between the BESs and NLEs, which may partly explain why the NLEs use *you know* for this interpersonal function less frequently than the BESs. In addition to lack of knowledge, which most probably comes from lack of input, the underrepresentation of this use of *you know* in LINDSEI-NO may also be caused by the learners' lack of confidence in speech. Although *you know* typically serves to soften the force of the utterance, this particular use of the DM requires speaker confidence. As suggested by previous research, learners often lack such confidence (Müller, 2004: 1175), and this may be an explanatory factor to these results.

Plea for cooperation/confirmation:

As was the case in Östman (1981) and Müller (2005), my data reveals frequent uses of *you know* to plea for cooperation or confirmation from the hearer. As can be read from Table 21 above, this function occurs 177 times in LOCNEC (10.4 pttw), but only 45 times in LINDSEI-NO (3.7 pttw). This difference is considered highly significant³⁴.

As presented in Table 23, this use of *you know* is defined a speaker's request to the hearer for confirmation that he or she follows the argument or agrees with what the speaker says. In other words, *you know* is with this function used to establish a situation where both the speaker and hearer know about the other person's knowledge about a topic. This information is what Schiffrin (1987: 274) refers to as meta-knowledge.

A deeper analysis of the 222 cases of *you know* functioning as a 'plea for cooperation/confirmation' uncovers that this use often combines with the textual function 'introduce given information' (LOCNEC n=28, 1.6 pttw; LINDSEI-NO n=1, 0.1 pttw). As the number of hits suggest, this dual interpersonal and textual function almost exclusively occurs in LOCNEC. The BESs' frequent use of *you know* with the dual interpersonal and textual functions 'plea for cooperation/confirmation' and 'introduce given information' may be explained by the fact that when a speaker introduces given information, it is natural that he also desires a confirmation from the hearer that s/he actually is familiar with the topic under discussion. By receiving such a confirmation, the interlocutors have established a common ground of knowledge, and the speaker can continue his or her argument. This function of *you know* is defined as very central by previous researchers such as Östman (1981) and Schiffrin (1987). Example (5.53) below illustrate such usage of *you know*.

5.53 filling documents or [that kind of thing <A> [yeah and em . work on the computer typing stuff and .. sorting out little **you know** tourist packages and that <A> oh yeah mm well it was a few years ago yes about two years ago (E332)

You know is furthermore found to also serve the dual interpersonal and qualifying function 'plea for cooperation/confirmation' and 'expressing attitudes, feelings or evaluations' (LOCNEC n=59, 3.4 pttw; LINDSEI-NO n=10, 0.8 pttw). A closer examination of these cases reveals that *you know* is commonly used as a face-saver, which is in line with what

³⁴ More than 99.99% certain.

Holmes (1986) discovers in her data: when expressing a feeling, evaluation or attitude, the speaker often invites the hearer to cooperate and confirm the argument so that the speaker may avoid stepping on the hearer's toes. If the hearer confirms an attitude, feeling or evaluation expressed by the speaker, the speaker feels safe to continue the argument. This use of *you know* is illustrated in example (5.54) below. In this example the hearer confirms the speaker's statement saying that it can get quite depressing being in the same place all the time. This statement serves as a sign to the speaker that he can continue his argument without stepping on the hearer's toes.

5.54 *it can get a bit depressing **you know*** <A> *yes* *being in the same place all the time [. and* <A> [*yes it's true that it's very small here [and you (E124)*

Hold the floor:

Although 'hold the floor' has not been emphasized as an important interpersonal function for *you know* by previous researchers, this function is found quite frequently in LOCNEC in this study. As Table 21 shows, the 'hold the floor' use of *you know* occurs 53 times in LOCNEC (3.1 pttw) but only 4 times in LINDSEI-NO (0.3 pttw). This difference is considered highly significant³⁵.

A closer examining of these 57 cases uncovers that this use of *you know* very often serves the dual textual and interpersonal functions 'move (back) to story' and 'hold the floor' (LOCNEC n=40, 2.3 pttw; LINDSEI-NO n=1, 0.1 pttw), as illustrated by example (5.55) below. This observation suggests that the interpersonal function 'hold the floor' is most typically used after a digression or an interruption from the hearer where the speaker feels the need to direct the hearer's focus back to the initial topic. The speaker thus feels the need to simultaneously indicate to the hearer that he is not finished, and that something more is about to come.

5.55 *mhm very claustrophobic indeed cos after three years you think . you get the feeling that you know everybody and . **you know** I know everybody in my my college and everything at times nice to be living off but the convenience (E35)*

Another observation made here is that 'hold the floor' *you know* sometimes is preceded by the conjunctions *and* (LOCNEC n=17, 1 pttw; LINDSEI-NO n=2, 0.2 pttw), *but* (LOCNEC

³⁵ More than 99.99% certain.

n=5, 0.3 pttw; LINDSEI-NO n=0), *so* (LOCNEC n=5, 0.3 pttw; LINDSEI-NO n=0), and *because* (LOCNEC n=2, 0.1 pttw; LINDSEI-NO n=0). The DM is furthermore exclusively used with prospective orientation, occurring either in the pre-front field or as a wildcard. This indicates that *you know* often is followed by a subordinate or coordinate clause when used with the interpersonal function ‘hold the floor’. Example (5.56) below illustrates this. Here the speaker is using *you know* to indicate that the story is not finished yet, and what follows is a coordinate clause introduced by the conjunction *and*.

5.56 *gonna do with this but you know he never came back so I have no idea what it is so we still have that hat a= back home and **you know** and it's just sitting on top of the shelves like . it's fun to s= to story to tell people and they're like . whoa (N8).*

To sum up, the examination of *you know* used to ‘hold the floor’ has yielded strong indications that this function is typically used together with 1) the textual function ‘move (back) to story’ and with 2) conjunctions such as *and*, *but*, *so*, and *because* to introduce a clause. According to the frequency results from the functional analysis, the NLEs display little awareness of this use of *you know*.

Qualifying functions of *you know*

Compared to *like* and *well*, *you know* is not that frequent and does not serve that many different qualifying functions. In fact, the only qualifying functions identified for *you know* in this study are ‘expressing attitudes, feelings or evaluations’, ‘contribute an opinion’ and ‘express disagreement’. However, the latter two only occur once each in the two corpora, and have therefore been placed in the ‘other’ category. This leaves us with only one qualifying function for *you know*: ‘expressing attitudes, feelings or evaluation’. This function is very broad, and includes cases where the speaker is referring to his/her own experiences, or providing a subjective evaluation of something. Example (5.57) from LOCNEC illustrates such a case where the speaker is expressing desire.

5.57 *So although I never noticed it my parents didn't particularly like it .. erm and going to school with Chinese . erm .. **you know** I wanted to go and play at friends' houses yeah and you know they were always doing maths tuition French (E177)*

As the results presented in Table 21 indicate, ‘expressing attitudes, feelings or evaluations’ occurs 122 times in LOCNEC (7.2 pttw) and 22 times in LINDSEI-NO (1.8 pttw). The

statistical calculations reveal that this difference is highly significant³⁶, and that this result serves as strong evidence against the H0. In other words, there is a significant underrepresentation of *you know* serving qualifying functions in general and of *you know* serving the particular qualifying function ‘expressing attitudes, feelings or evaluations’ in LINDSEI-NO as compared to LOCNEC. This is illustrated in Figure 22 below.

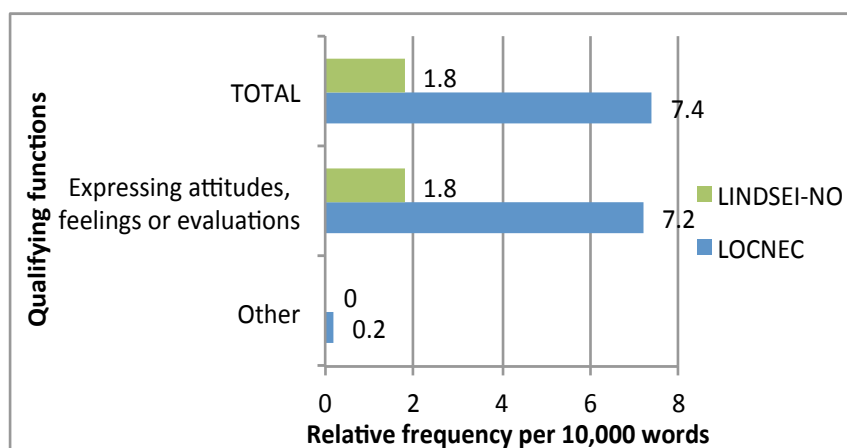


Figure 22: Qualifying functions of *you know* in LOCNEC and LINDSEI-NO

An in-depth examination of the total 144 cases of *you know* used to ‘express attitude, feelings or evaluations’ reveals that this qualifying function is frequently used simultaneously with the interpersonal functions ‘plea for understanding’ (LOCNEC n=25, 1.5 pttw; LINDSEI-NO n=7, 0.6 pttw) and ‘plea for cooperation/confirmation’ (LOCNEC n=58, 3.4 pttw; LINDSEI-NO n=10, 0.8 pttw). This indicates that the speaker often uses *you know* for positive politeness, as Holmes (1986) states, with the purpose of distancing him/herself from face-threatening situations. Such situations may occur when the speaker expresses an opinion, evaluation or attitude, but through the use of *you know* he or she thus modifies the strength of the utterance.

Example (5.58) below shows a case where *you know* serves the triple textual, interpersonal and qualifying functions ‘introducing explanation’, ‘plea for cooperation/confirmation’ and ‘expressing attitude, feelings or evaluation’. In this case the speaker uses *you know* when expressing an opinion or evaluation of Chinese people. In order to avoid sounding racist and offensive, *you know* is used to modify the statement and as a plea for a confirming feedback from the hearer signaling that no offence is taken.

³⁶ More than 99.99% certain.

5.58 *My parents didn't like it particularly but . not being racist in the slightest but Chinese tend to be very sort of em .. like **you know** we go to the same shop for seven years . and you'd thought seven years later the man would make eye (E327)*

Example (5.59) presents a case where *you know* serves the triple textual, interpersonal and qualifying function 'introducing explanation', 'plea for understanding' and 'expressing attitudes, feelings or evaluations'. This is a frequently occurring triple function of *you know* in LOCNEC (n=19, 1.1 pttw), but it only appears twice in LINDSEI-NO (0.2 pttw). In the example the speaker is expressing evaluation of a Norwegian course, and uses *you know* to introduce the explanation for his evaluation and to plea for understanding from the hearer. This use of *you know* can also be seen as face-saving in the sense that the speaker wants to justify why he did not find the course fun.

5.59 *difference between be= it being a B language . course and Norwegian last year was very . (em) detailed and (eh) theoretical . **you know** not as much fun as I thought it would be because . I really like Norwegian too and the Norwegian old literature .. (N125)*

In sum, the examination of the qualifying function 'expressing attitudes, feelings or evaluations' reveals that this use of *you know* usually occurs together with the interpersonal functions 'plea for understanding' and 'plea for cooperation/confirmation' and the qualifying function 'introducing example' in the British RLV data. The NLEs display a less frequent use of *you know* with the qualifying function 'expressing attitudes, feelings or evaluations' when it is used together with these interpersonal and qualifying functions.

5.5 Summary and discussion of results

The aim of this study has been to test and explore the extent to which the British English speakers and Norwegian learners of English differ in their use of the DMs *like*, *well* and *you know*, and to what extent these differences can be explained by factors such as input and L1 transfer. In order to explore these aspects, the following four research questions were asked (see also section 2.1.1):

RQ1: Are there differences between Norwegian advanced learners of English and British native speakers of English in their use of the discourse markers *like*, *you know* and *well* in spoken language?

RQ2: Do the Norwegian learners use these discourse markers for different pragmatic functions than the native speakers?

RQ3: To what extent can the observed discrepancy between the two speaker groups be explained by factors such as lack of input, transfer and anxiety?

RQ4: Are there any other factors that appear to have affected the learners' use of discourse markers?

In the following section these questions will be discussed and answered based on the findings presented in the previous sections of chapter 5.

5.5.1 Reference language and interlanguage variety practice

Quantitative differences

RQ1 is answered by the preliminary analysis presented in section 5.1. This analysis uncovers a clear difference between the NLEs and the BESs, suggesting a highly statistical underrepresentation of all three DMs in the NLE data. These results are presented in detail in Table 7, and as this table shows, the difference between the two speaker groups is strongest with the DM *you know*, and thereafter follows *well* and *like*.

This finding is in line with Fung and Carter (2007) and Müller (2005), who also uncover an underuse in the learner production. Yet, the results appear as quite unexpected, as it contrasts with the majority of studies on learners' use of DMs, typically reporting that non-native speakers overuse DMs (Müller, 2004; Aijmer, 2011; Unaldi, 2013; Buysse, 2012). The present results are especially unforeseen for the DM *well*, as this marker has been profoundly

studied and typically found to be overused by non-native speakers (Aijmer, 2011; Müller, 2005). It is also striking that the present results differ so greatly from Aijmer's (2011), who has conducted a similar study of the DM *well* on the Swedish sub-section of LINDSEI, since one would assume Norwegian and Swedish speakers behave quite similarly in their use of this marker.

Qualitative differences

Like

The results for *like* show that both speaker groups display similar usage of this DM in terms of orientation and utterance position. The functional analysis of this DM reveals that *like* occurs most frequently with textual functions in both corpora, followed by interpersonal and qualifying functions, respectively. According to the statistical calculations presented with the results in Table 10, the strongest underrepresentation of *like* lies within the group of textual functions for the Norwegian learners. Although some functions are found to yield very low frequencies in LINDSEI-NO ('transition/topic shift', 'unfinished point' and 'hold the floor',) none of the pragmatic functions of *like* are entirely absent in any of the two corpora. This suggests that both speaker groups demonstrate a similar functional scope. The findings for *like* are thus mostly in line with Müller (2005), who finds that German learners underuse *like* for all pragmatic functions. However, whereas the German learners are found to underuse quotative *like*, the Norwegian learners of this study use this function more frequently than the BESs. An explanation to this difference between the Norwegian and German learners may be that Norwegians receive more input of colloquial English including such uses of *like* from the media than Germans, because movies and TV shows often are dubbed in Germany whereas they in Norway are broadcast in English with Norwegian subtitles.

Well

Apart from the general underrepresentation of the DM *well* in LINDSEI-NO as compared to LOCNEC, the functional analysis of this DM reveals quite similar tendencies by both speaker groups in terms of utterance position, orientation, and usage of this DM at the textual and interpersonal level. This pattern is illustrated in Figures 9 to 12. Exceptions here are the textual function 'quotative *well*', which is strongly underrepresented in the learner data, and the interpersonal functions 'thinking/considering' and 'attention-getter', which are used more frequently by the NLEs than the BESs. These results mostly contrast with Müller (2005), who discovers that the German learners overuse all functions of *well* except from

‘rephrasing/correcting’, ‘quotative *well*’ and ‘contributing an opinion’, as well as a much more diverse usage of this DM for the different pragmatic functions (see Figure 3 in Müller 2005: 138).

In terms of *well* serving at the qualifying level, this study reveals a greater difference between the BESs and NLEs than for the textual and interpersonal levels, as shown by Figure 13 and Table 15. Both speaker groups demonstrate an overall lower frequency of *well* functioning at this level, but the Norwegian learners’ usage is significantly lower than the British. This observation is in line with Aijmer (2011), who discovers that learners use DMs more often for speech management functions than for other pragmatic functions. Moreover, within this category, several observations are made confirming the difference between the two compared populations: the learners demonstrate no use of *well* to express disagreement and only one instance of *well* to reject a previous statement. Also, a statistically significant underrepresentation by the NLEs is identified for the qualifying functions ‘modifying a previous statement’ and ‘confirming a previous statement’. This observations is also in line with Aijmer’s claim that native speakers use *well* more often to soften disagreements than non-natives (2011: 251).

You know

You know is the discourse marker that is most underrepresented in LINDSEI-NO as compared to LOCNEC in this study. The functional analysis reveals that the NLEs use *you know* less frequently than the BESs with all pragmatic functions, but that ‘introducing given information’ and ‘plea for understanding’ are the two functions closest to the reference variety norm. Similar to what Müller (2005) discovers for German learners, the Norwegian learners exhibit lack of knowledge of *you know* used to close a point, mark transition/topic shift, move back to story, take the turn, “see the implication” and to hold the floor. All these uses of *you know* occur with very low frequency in LINDSEI-NO, suggesting that the learners’ functional scope is much less comprehensive than the native speakers’ for this DM.

In sum, and as an answer to the second research question, the results of the qualitative analysis reveal a difference between the British reference language speakers and the Norwegian interlanguage speakers in the production of *like*, *well* and *you know*, but suggest that the extent of this discrepancy varies for each marker. The most reference language-like usage is found with the DM *like*, whereas the learners’ use of the DM *you know* is found to

differ most extensively from the British English reference variety. Although the NLEs display lack of knowledge of some pragmatic functions, few of the differences are identified as misuses. Rather, it seems that the learners have not acquired all possible usage of these markers, i.e. which words collocate with these DMs and which functions these DMs can serve. Possible reasons for this lack of knowledge will be discussed in the following sections.

5.5.2 Factors contributing to the learners' underrepresentation of DMs

As mentioned repeatedly in the analysis, lack of input, transfer from Norwegian and reduced self-confidence stand out as possible explanations to several of the cases where the three DMs are either under- or overrepresented in the Norwegian learner data.

Lack of input, especially through textbooks and classroom instructions, is marked as one of the most central factors that may explain the results of this study (Lam, 2009; Hellermann and Vergun, 2007; Fung and Carter, 2007). Especially in cases where the BESs display a systematic usage of the DM for a particular pragmatic function, as seems to be the case for “quotative *well*”, “attention-getter” *like*, and *you know* used to introduce an explanation, more native input and perhaps explicit classroom instruction of such patterns would probably serve to reduce the discrepancy between interlanguage and reference speaker norms. Hellermann and Vergun (2007) state that “[t]he more time students have spent in the US and the more contact they have had with the target language culture, the more likely it seems that they will use discourse markers in their pragmatic usages” (2007: 169). This statement is also supported by Müller (2005). But since not every student has the opportunity to take a semester abroad, learners are left to acquire this aspect of the English language through input elsewhere. This is where textbooks and classroom instructions become essential. As already discussed in section 1.1 of this thesis, both Norway and other countries demonstrate deficient focus on DMs in English education, and this lack of input affects the learners' output. This is a challenge addressed by several scholars (Lam, 2009; Hellermann and Vergun, 2007; Fung and Carter, 2007), and the present study's results serve only to strengthen the force of this argument.

However, while at the same time acknowledging the necessity for satisfactory input, I do not believe, based on the results of this study, that input alone explains the underrepresentation of the three DMs in the NLE data. A quick examination of the relationship between the

frequency of the three DMs and the 50 LINDSEI-NO-participants' stay in an English-speaking country discloses a weak correlation³⁷. Among the participants, there are speakers who demonstrate a frequent use of the DMs and who also have spent some time abroad, such as speaker 10, 31 and 37. Yet, there are also speakers who have lived abroad for a long time, and thus received sufficient input of the target forms, but who still demonstrate low frequency of the use of the three DMs (e.g. speaker 13, 24 and 46). Moreover, some of the speakers exhibiting frequent use of the three DMs report not having lived in an English-speaking country before. It is of course possible that these speakers have received input in other ways, such as through video games, the media or English-speaking friends and relatives, which is not included in the metadata. In sum, although the metadata does not provide sufficient information about every aspect of the participants' reference language input, this preliminary examination suggests that input alone cannot be the only factor affecting the use of DMs. Other factors, such as age (Erman, 2001), gender (Holmes, 1986; Macaulay, 2002), motivation (Ortega, 2009) and perhaps cultural norms must also be taken into consideration. Studies also mention acculturation into the L2 culture and desire to speak English fluently as possible factors affecting learners' use of DMs (Hellermann and Vergun, 2007; Aijmer, 2011). Investigating the effect these factors have on the acquisition of DMs lies beyond the scope of this thesis, and would need to be further investigated in future studies.

Another central factor mentioned throughout the analysis, which may affect learners' acquisition of *like*, *well* and *you know*, is transfer from Norwegian. All three DMs have Norwegian counterparts (*liksom* and *jeg bare* for *like* and *I'm like*, *vel* for *well*, *du skjønner* and *du vet* for 'you know', etc.), which arguably may have contributed to the under- and overrepresentation of certain DMs for certain functions in the NLE data. For example, it is argued above that overrepresentation of quotative *like* in LINDSEI-NO may partly be explained by transfer from the Norwegian markers *liksom* and *jeg bare*, which are frequently used for the same function (Hasund, 2002). As mentioned, this study's results for *well*, however, are in fact quite unexpected if one considers transfer as a central factor effecting learners' acquisition, as they generally uncover contrasting results to Aijmer (2011). Since her study is based on the Swedish section of the LINDSEI corpus, and uncovers that learners' use *well* more frequently than the native speakers, one would expect Norwegian learners to

³⁷ See appendix 1 for metadata for LINDSEI-NO.

demonstrate similar tendencies. Although this observation may indicate that Swedish and Norwegian speakers differ in their use of *vel*, it also suggests that transfer alone cannot be seen as the only factor affecting learners' production of DMs either.

The analysis has furthermore revealed cases where factors such as the need for time to think about what to say and lack of self confidence when speaking in a second language may contribute to the under- and overrepresentation of certain pragmatic functions in the NLE data. The function for *like* 'searching for the right phrase', for instance, is used frequently relative to other the other pragmatic functions by the NLEs compared to the BESs, indicating that this use is more central for the learners than for their British peers, which may relate to the fact that learners more often search for words than native speakers, as reported in Müller (2005). This argument is further strengthened by the results for the interpersonal function of *well* 'thinking/considering', which is also overrepresented by the NLEs of this study, and by the interpersonal function 'hesitator', which is used relatively often by the learners as compared to the other functions (although still underrepresented as compared to the BESs). Moreover, the learners display a less frequent use of functions such as the 'take the turn' *well*, 'hold the floor' *well*, and 'attention-getter' *you know*. These are all functions requiring a certain degree of self confidence as they often involve an interruption followed by a request for attention, and this observation further strengthens the claim that learners typically demonstrate lower self confidence and more hesitancy when speaking in a foreign language.

In sum, it has in this section been argued that the discrepancies between the BESs and NLEs uncovered in this study are products of several factors affecting how learners acquire discourse markers. Input, transfer from Norwegian and lack of self-confidence are all found to be central factors contributing to the overall underrepresentation in the NLE data, but they cannot serve as the whole explanation. In addition, this discussion has introduced other factors such as age, gender, cultural differences and a desire to become part of the L2 culture (motivation) as possible explanatory factors. Hopefully, these, among others, will in the future be subject to further investigations of possible elements affecting and contributing to learners' acquisition of discourse markers.

6 Concluding Remarks

The aim of this study was to investigate whether, how and why Norwegian learners differ from British speakers of English in their use of the three discourse markers *like*, *well* and *you know*. In exploring these questions, a corpus-based contrastive interlanguage analysis of spoken material using the two corpora LOCNEC and LINDSEI-NO was conducted.

This analysis has discovered a general underrepresentation of the three discourse markers *like*, *well* and *you know* in the language produced by the Norwegian learners of English as compared to British English speakers at the same educational level. This difference is evident in most pragmatic sub-functions within the textual, interpersonal and qualifying levels, with some exceptions. Although both speaker groups demonstrate an almost similar functional scope for all three DMs, i.e. they use the DMs for approximately the same pragmatic functions, the NLEs reveal a much less frequent use of certain pragmatic functions and a virtually nonexistent use of others when compared to the BESs. In the discussion of these results, I have examined lack of input, especially from textbooks and classroom instruction, transfer from Norwegian translation equivalents, learners' low confidence when speaking in their second language, and cultural differences as some of the possible factors affecting the learners' acquisition of the target discourse markers. The findings suggest that these factors to some extent affect the learners' production of the DMs. However, what is evident from the analysis is the fact that neither of these factors alone can serve to explain the discrepancy between the Norwegian learners and British native speakers. Further research is needed to determine the extent to which each factor plays a role, and whether other elements also should be taken into consideration.

6.1 Strengths and limitations of this study with suggestions for future studies

It turned out to be very beneficial for the purpose of this study to be able to access such large amounts of corpus material spoken by both learners and native speakers of English. This large amount of data has enabled my analysis to yield significant results, revealing strong tendencies of the usage of the three DMs by both speaker groups. However, as discussed in Chapter 4, the accessed corpus material also has its limitations. For this particular study, the most prominent shortcomings of the material used are the issues of authenticity and

individual variation among the participants in each corpus. As discussed in section 4.1.4, the fact that both LINDSEI-NO and LOCNEC are built up of interviews consisting of three different tasks where one is a picture description task, limits the degree of authenticity of the language produced. The fact that this is an unnatural situation for the interviewees, where they have to speak to an unfamiliar person, may affect the way in which they speak in terms of the degree of formality, accent, personal feelings etc. Relevant for this particular issue is Macaulay's report that "[s]peakers are more likely to use *you know* in conversation with an acquaintance than in interviews with a stranger" (2002: 765). Thus, we might expect the frequencies uncovered in LOCNEC and LINDSEI-NO to be lower than if the speakers had been speaking to a friend or an acquaintance. I realize the difficulty of compiling a learner corpus consisting of entirely authentic data, as learners seldom naturally speak in their L2 unless they are abroad or talking to someone with a different mother tongue. However, it would, if possible, be interesting for future studies to look at the same discourse markers and the same populations but from data produced in conversations of two acquaintances.

Concerning individual variation, the examination presented in section 5.1.2 shows that there is great variation within each of the two corpora on how frequently the participants of this study use the three DMs. In order to be able to generalize the results to a greater degree, future studies would benefit from using more sophisticated statistical measures to calculate and control for this variable.

Another strength of this study is the fact that I was able to use my own introspection about the Norwegian language as a comparison to English. This has enabled me to apply the contrastive interlanguage analysis model to fully explore possible factors affecting how Norwegian learners in particular acquire the three DMs. However, as my field of study is English and not Norwegian, and since this comparison is entirely based on my subjective intuition, I might not have captured all potential aspects of the use of the Norwegian equivalents. It would therefore be useful for future research to use empirical studies, and not a researcher's own introspection of the Norwegian language, as basis for comparing an English RLV with a Norwegian IL variety.

6.2 Pedagogical implications

In Chapter 2 we saw that discourse markers are defined as items creating coherence (Schiffrin, 1987; Fraser, 1990; Hansen, 1997) and “facilitating the hearer’s interpretation of the utterance” (Aijmer, 2002: 2). They are also found to contribute to the conversation in situations where it is not going very well, such as incidents of disagreements, or situations where the speaker needs to soften the force of the utterance in order to avoid face-threatening situations (Holmes, 1986). This study has confirmed these characteristics, discovering a widespread usage of *like*, *well* and *you know* serving pragmatic functions at the textual, interpersonal and qualifying level. Moreover, this study has revealed a statistically significant underrepresentation of all these DMs in the NLE data, indicating that the Norwegian learners do not use the markers as often as their reference language peers. Whether this underrepresentation leads to situations where the learners are perceived as being what Svartvik refers to as “dogmatic, impolite, boring, [and] awkward to talk to” (1980: 171) is difficult to say based on the corpus material at hand. Yet, the study suggests that a more frequent or different use of the target DM would in many situations clarify the messages intended by the speaker.

As briefly discussed in section 1.1, little attention is given to DMs in the two textbooks *Passage* and *Targets* briefly examined for this study. Although this quick examination cannot conclude that all textbooks used for the teaching of English in Norway or all classroom discussions comprise the same lack of focus, it serves to stress a challenge already addressed in previous research: “the language in textbooks often differ from language in use” (Lam, 2009: 261). This study’s results, revealing a general underrepresentation of the three DMs under investigations in the NLE data, further underline that more input and focus on such DMs is needed in the teaching of English in Norway. With this study I thus hope to have shed light on discourse markers as an important aspect of the English language, which needs to be included in textbooks and classroom instruction so that learners can receive more input and thereby improve their output of these items. Being a future teacher of English as a foreign language, I hope that the present study provides useful information for learners and teachers of English and for textbook compilers in the development of pedagogical methods and theories. Although having received an unmerited reputation of belonging to the informal and colloquial field of spoken English, discourse markers are still important items essential for learners in their attempt to achieve communicative competence.

References

Primary sources

The Louvain Corpus of Native English Conversation. Used by courtesy of Sylvie De Cock, Université catholique de Louvain. <https://www.uclouvain.be/en-cecl-lindsei.html>

The Louvain International Database of Spoken English Interaction – Norwegian. Accessed through CD-ROM at the University of Oslo by courtesy of Susan Nacey, Hedmark University of Applied Sciences. <http://www.uclouvain.be/en-307845.html>

Secondary sources

Abraham, W. (1991). "Discourse particles in German: How does their illocutive force come about? ". In: Abraham, W. (ed.) *Discourse Particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German*. Amsterdam & Philadelphia: John Benjamins.

Aijmer, K. (2002). *English Discourse Particles: Evidence from a corpus*. Amsterdam & Philadelphia: John Benjamins B. V. .

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-Vandenberg (2011). "Pragmatic Markers ". In: Zienkowski, J., Östman, J.-O. & Verschueren, J. (eds.), *Discursive Pragmatics*. Amsterdam and Philadelphia: John Benjamins

Andersen, G. (1997). "*They gave us these yeah, and they like wanna see like how we talk and all that*. The use of *like* and other pragmatic markers in London teenage speech. ". In: Kotsinas, U.-B., Stenstrom, A.-B. & Karlsson, A.-M. (eds.), *Ungdomsspråk i Norden: föredrag från et forskarsymposium*. Stockholm: Institutionen för nordiska språk, Stockholms universitet

Andersen, G. (2001). *Pragmatic Markers and Sociolinguistic Variations: A Relevance-Theoretic Approach to the Language of Adolescents*. Amsterdam: John Benjamins.

Ariel, M. 1993. Pragmatic operators. *The Encyclopedia of Language and Linguistics*. Oxford Pergamon Press.

Ariel, M. (1998). "Discourse markers and form-function correlations". In: Jucker, A. H. & Ziv, Y. (eds.), *Discourse Markers: Descriptions and theory*. Amsterdam & Philadelphia: John Benjamins.

- Barbieri, F. (2010). "Review of G. Gilquin, S. Papp & M. B. Díez-Bedmar (Eds.) Linking Up Contrastive and Learner Corpus Research. Amsterdam: Rodopi (2008)". *Studies in Second Language Acquisition*, 32 (1), 148-150.
- Barlow, M. (2005). "Computer-based analyses of learner language". In: Ellis, R. & Barkhuizen, G. (eds.), *Analysing Learner Language*. Oxford: Oxford University Press.
- Bazzanella, C. (1990). "Phatic connectives as intonational cues in contemporary spoken Italian". *Journal of Pragmatics*, 14 (4), 629-647.
- Bell, D. M. (2010). "Nevertheless, still and yet: Concessive cancellative discourse markers". *Journal of Pragmatics*, 42 1912-1927.
- Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (2006). "Discourse Markers". In: Horn, L. R. & Ward, G. (eds.), *The Handbook of Pragmatics*. Oxford Blackwell Publishing
- Bley-Vroman, R. (1983). "The comparative fallacy in interlanguage studies: the case of systematicity". *Language Learning*, 33 1-17.
- Blyth, C., Jr., S. Recktenwald & J. Wang. (1990). "'I'm like, Say What?!': A new quotative in American oral narrative". *American Speech*, 65 215-227.
- Brand, C. & S. Götz. (2011). "Fluency versus accuracy in advanced spoken learner language". *International Journal of Corpus Linguistics*, 16 (2), 255-275.
- Brinton, L. J. (1996). *Pragmatic Markers in English: Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter
- Brown, G. (1977). *Listening to Spoken English*. London: Longman.
- Brown, P. & S. C. Levinson. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Brutt-Griffler, J. & K. K. Samimy. (2001). "Transcending the nativeness paradigm". *World Englishes*, 20 (1), 99-106.
- Burgess, R. & T. B. Sørhus. (2009). *Passage: Engelsk VGI Studieforberevende Program*. Cappelen Damm.
- Buyse, L. (2012). "So as a multifunctional discourse marker in native and learner speech". *Journal of Pragmatics*, 44 1764-1782.
- Cadiot, A., O. Ducrot, B. Fradin & T. B. Nguyen. (1985). "Enfin, Marqueur Metalinguistique". *Journal of Pragmatics*, 9 199-239.
- Callies, M. & M. Paquot. (2015). "Learner Corpus Research: An interdisciplinary field on the move". *International Journal of Corpus Linguistics*, 1 (1), 1-6.

- Castele, A. V. & K. Collewaert. (2013). "The Use of Discourse Markers in Spanish Language Learners' Written Compositions". *Procedia - Social and Behavioral Sciences*, 95 550-556.
- Coed 1976. *The Concise Oxford Dictionary of Current English*. 6th ed. ed. Oxford: Clarendon Press.
- Cook, H. M. (1990). "An indexical account of the Japanese sentence-final particle no". *Discourse Processes*, 13 (4), 401-439.
- Corder, S. P. *The Study of Interlanguage*. Proceedings of the Fourth International Conference of Applied Linguistics, 1976 Munich, Hochschulverlag. Oxford University Press.
- Cotter, C. (1996). "Engaging the reader: the changing use of connectives in newspaper discourse ". In: Arnold, J., Blake, R., Davidson, B., Schwenter, S. & Solomon, J. (eds.), *Sociolinguistic Variation: Data, Theory and Analysis*. Stanford University: CSLI Publications.
- Cuenca, M.-J. (2008). "Pragmatic markers in contrast: The case of *well*". *Journal of Pragmatics*, 40 1373-1391.
- Davies, A. (1978). "Language testing: Survey article Part 1". *Language Teaching and Linguistics Abstracts*, 11 145-159.
- De Cock, S. (2004). "Preferred sequences of words in NS and NNS speech". *Belgian journal of English language and literatures*, 225-246.
- Ellis, R. & G. Barkhuizen. (2005). *Analyzing Learner Language*. Oxford and New York: Oxford University Press.
- Erman, B. (1987). *Pragmatic expressions in English: A study of 'you know', 'you see', and 'I mean' in face-to-face conversation*. . 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.
- Ferguson, G. (2009). "Issues in researching English as a lingua franca: a conceptual enquiry". *International Journal of Corpus Linguistics*, 19 (2), 117-135.
- Fox Tree, J. E. & J. C. Schrock. (2002). "Basic meanings of *you know* and *I mean*". *Journal of Pragmatics*, 34 727-747.
- Fraser, B. (1988). "Types of English discourse markers". *Acta Linguistica Hungarica*, 38 (1-4), 19-33.
- Fraser, B. (1990). "An Approach to Discourse Markers". *Journal of Pragmatics*, 14 383-395.
- Fraser, B. (1999). "What are discourse markers? ". *Journal of Pragmatics*, 31 931-952.

- Fuller, J. M. (2003). "The influence of speaker roles on discourse marker use". *Journal of Pragmatics*, 35 23-45.
- Fung, L. & R. Carter. (2007). "Discourse Markers and Spoken English: Native and Learner Use in Pedagogic Settings". *Applied Linguistics*, 28 (3), 410-439.
- Gilquin, G. (2000/2001). "The Integrated Contrastive Model: Spicing up your data". *Languages in Contrast*, 3 (1), 95-123.
- Gilquin, G. (2008). "Hesitation markers among EFL learners: Pragmatic deficiency or difference? ". In: Romero-Trillo, J. (ed.) *Corpus and Pragmatics. A Mutualistic Entente*. Berlin/New York: Mouton de Gruyter.
- Gilquin, G. (2012). *LINDSEI* [Online]. UCL: Université catholique de Louvain. Available: <https://www.uclouvain.be/en-cecl-lindsei.html> [Accessed 13.08 2015].
- Gilquin, G. (2015). *LINDSEI partners* [Online]. Louvain-la-Neuve. Available: <https://www.uclouvain.be/en-307845.html> [Accessed 15. December 2015].
- Gilquin, G., S. De Cock & S. Granger. (2010). *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain
- Gilquin, G. & M. Paquot. (2008). "Too chatty: Learner academic writing and register variation". *English Text Construction*, 1 41-61.
- Goffman, E. (1981a). "Replies and Responses". *Language in Society*, 5 (3), 257-313.
- Granger, S. (1996). "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora". In: Aijmer, K., Altenberg, B., and Johansson, M. (ed.) *Languages in Contrast: Text-based cross-linguistic studies*. Lund: Lund University Press.
- Granger, S. (1998b). "The Computer Learner Corpus: A versatile new source of data for SLA research ". In: Granger, S. (ed.) *Learner English on Computer* London: Longman.
- Granger, S. (2002). "A bird's eye view of learner corpus research". In: Granger, S., Hung, J. & Petch-Tyson, S. (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* Amsterdam and Philadelphia: Benjamins.
- Granger, S. (2008). "Learner Corpora". In: Lüdelung, A. a. K. M. (ed.) *Corpus Linguistics: An International Handbook*. Berlin and New York: Walter de Gruyter.
- Granger, S. (2009). "The contribution of learner corpora to second language acquisition and foreign lanugage teaching: A critical evaluation". In: Aijmer, K. (ed.) *Corpora and Language Teaching* Amsterdam: John Benjamins

- Granger, S. (2015). "Contrastive interlanguage analysis: A reappraisal". *International Journal of Corpus Linguistics*, 1 (1), 7-24.
- Greasley, P. (1994). "An investigation into the use of the particle *well*: Commentaries on a game of snooker". *Journal of Pragmatics*, 22 477-494.
- Grice, H. P. (1975). "Logic and Conversation". In: Cole, P. & Morgan, J. L. (eds.), *Syntax and Semantics, Volume 3: Speech Acts* New York: Academic Press.
- Gries, S. T. (2009). "Ch. 2: The Three central Corpus-linguistic Methods". *Quantitative Corpus Linguistics with R. A Practical Introduction* Routledge
- Gupta, A. F. (1992). "The pragmatic particles of Singapore colloquial English". *Journal of Pragmatics*, 18 (1), 31-57.
- Gülich, E. (1970). *Makrosyntax der Gliederungssignale im gesprochen Französisch*. Walter Fink
- Gülich, E. & T. Kotschi. (1983). "Les marqueurs de la réformulation paraphrastique. Connecteurs pragmatiques et structure du discours. Actes du 2éme Colloque de Pragmatique de Genève". *Cahiers de linguistique française*, 5 305-351.
- Halliday, M. a. K. (1985). *Introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. a. K. (1991). "Corpus studies and probabilistic grammar". In: Aijmer, K. & Altenberg, B. (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. Longman.
- Halliday, M. a. K. & R. Hasan. (1976). *Cohesion in English*. London: Longman.
- Hansen, M.-B. M. (1997). "*Alors* and *donc* in spoken French: A reanalysis". *Journal of Pragmatics*, 28 (2), 153-187.
- Hansen, M.-B. M. (1998). "The semantic status of discourse markers". *Lingua*, 104 235-260.
- Hasselgård, H. & S. Johansson (2011). "Learner corpora and contrastive interlanguage analysis". In: Meunier, F., De Cock, S., Gilquin, G. And Paquot, M. (ed.) *A Taste for Corpora: In honour of Sylviane Granger*. John Benjamins Publishing Company.
- Hasund, I. K. (2002). "Congratulations, *like!* - Gratulerer, *liksom!* Pragmatic particles in English and Norwegian". In: Breivik, L. E. & Hasselgren, A. (eds.), *From the COLT's mouth ... and others': Language Corpora Studies In honour of Anna-Brita Stenström* Amsterdam & New York: Rodopi B.V. .
- Haugen, H., E. Haugum, J. Kagge, C. A. Ljones, S. Natvig & A. Rugset. (2009). *Targets: Engelsk VGI*. Aschehoug.
- He, A. W. & R. Young (1998). "Language Proficiency Interviews: A Discourse Approach". In: Young, R. & He, A. W. (eds.), *Talking and Testing*. Amsterdam: John Benjamins.

- Hellermann, J. & A. Vergun. (2007). "Language which is not taught: The discourse marker use of beginning adult learners of English". *Journal of Pragmatics*, 39 157-179.
- Helt, M. E. & S. H. Foster-Cohen (1996). "A relevance theoretic approach to older children's use of discourse markers". In: Stringfellow, A., Cahana-Amitay, D., Hughes, E. & Zukowski, A. (eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development*. Sommerville, MA: Cascadilla Press
- Hirzalla, H. (2007). "The construction of discourse and its development in first and second language". *California Linguistic Notes*, 32 (1).
- Holmes, J. (1986). "Functions of *you know* in women's and men's speech". *Language in Society*, 15 1-21.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hymes, D. H. (1966). "Two types of linguistic relativity ". In: Bright, W. (ed.) *Sociolinguistics: Proceedings of the UCLA sociolinguistics conference* The Hague: Mouton
- Hymes, D. H. (1972). "On communicative competence". In: Bright, W. (ed.) *Sociolinguistics*. The Hague: Mouton.
- Hölker, K. (1991). "Französisch: Partikkelforschung". *Lexicon der Romanistischen Linguistic*, 5 (1), 77-88.
- Johansson, S. (1998). "On the Role of Corpora in Cross-Linguistic Research". In: Johansson, S. & Oksefjell, S. (eds.), *Corpora and Cross-Linguistic Research, Theory, Method and Case Studies*. Amsterdam: Rodopi.
- Johansson, S. (2008). "Some aspects of the development of corpus linguistics in the 1970s and 1980s". In: Lüdelung, A. & Kytö, M. (eds.), *Corpus linguistics. An international handbook*. Walter de Gruyter.
- Jucker, A. H. (1993). "The discourse marker *well*. A relevance-theoretical account". *Journal of Pragmatics*, 19 (5), 435-453.
- Jucker, A. H. & S. W. Smith (1998). "And people just you know like *wow*: Discourse markers as negotiating strategies ". In: Jucker, A. H. & Ziv, Y. (eds.), *Discourse Markers: Descriptions and Theory*. Amsterdam & Philadelphia: John Benjamins.
- Kachru, B. (1985). "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle ". In: Quirk, R. & Widdowson, H. G. (eds.), *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press.

- Kilgarriff, A. (1997). "Putting frequencies in the dictionary". *International Journal of Lexicography*, 10 (2), 135-155.
- Knott, A. A. & R. Dale. (1994). "Using linguistic phenomena to motivate a set of coherence relations ". *Discourse Processes* 18 (1), 35-62.
- Koike, D. A. (1996). "Functions of the adverbial *ya* in Spanish narrative discourse". *Journal of Pragmatics*, 25 267-279.
- Krashen, S. (1985). *The Input Hypothesis*. London: Longman.
- Krashen, S. (1991). "The Input Hypothesis: An Update". In: Alatis, J. E. (ed.) *Georgetown University Round Table on Languages and Linguistics 1991*. Washington D.C.: Georgetown University Press.
- Lado, R. (1957). *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.
- Lakoff, R. (1973). "Questionable Answers and Answerable Questions". In: Kachru, B., Lees, R. B., Malkiel, Y., Pietrangeli, A. & Saporta, S. (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane* Urbana: University of Illinois Press
- Lam, P. W. Y. (2009). "Discourse Particles in Corpus Data and Textbooks: The Case of *Well*". *Applied Linguistics*, 31 (2), 260-281.
- Larsen-Freeman, D. (2014). "Another step to be taken - Rethinking the end point of the interlanguage continuum". In: Han, Z. & Tarone, E. (eds.), *Interlanguage. Forty years later* Amsterdam and Philadelphia: John Benjamins.
- Lawson, A. (2001). "Collecting, aligning and analyzing parallel corpora ". In: Ghadessy, M., Henry, A. & Roseberry, R. L. (eds.), *Small Corpus Studies and ELT: Theory and practice* Amsterdam and Philadelphia: John Benjamins.
- Lee-Goldman, R. (2011). "*No* as a discourse marker". *Journal of Pragmatics*, 43 2627-2649.
- Leech, G. (1992). "Corpora and Theories of Linguistic Performance". In: Svartvik, J. (ed.) *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin and New York: Mouton de Gruyter.
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Liu, B. (2013). "Effect of first language on the use of English discourse markers by L1 Chinese speakers of English". *Journal of Pragmatics*, 45 149-172.
- Macaulay, R. (2002). "You know, it depends". *Journal of Pragmatics*, 34 749-767.
- Macmillan 2002. *Macmillan English Dictionary for Advanced Learners*. Macmillan. Oxford: Macmillan/Bloomsbury.

- Maschler, Y. (1997). "Discourse markers at frame shifts in Israeli Hebrew talk-in-interaction". *Pragmatics*, 7 (2), 183-211.
- 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.
- Mittwoch, A. (1976). "Grammar and illocutionary force". *Lingua*, 40 21-42.
- Montolío Durán, E. & V. Unamuno. (2001). "The discourse marker *a ver* (Catalan, *a veure*) in teacher-student interaction". *Journal of Pragmatics*, 33 193-208.
- Mukherjee, J. (2009). "The grammar of conversation in advanced spoken learner English. Learner corpus data and language-pedagogical implications". In: Aijmer, K. (ed.) *Corpora and Language Teaching* Amsterdam and Philadelphia: John Benjamins B. V. .
- Mukherjee, J. & J.-M. Rohrbach (2006). "Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research ". In: Kettemann, B. G. M. (ed.) *Planing, Guiding and Painting Corpora: Inside the Applied Corpus Linguist's Workshop*. Peter Lang
- Müller, S. (2004). "*Well you know that type of person*: functions of *well* in the speech of American and German students". *Journal of Pragmatics*, 36 1157-1182.
- Müller, S. (2005). *Discourse Markers in Native and Non-native English Discourse*. Amsterdam and Philadelphia John Benjamins Publishing Company
- Norrick, N. R. (2001). "Discourse markers in oral narrative". *Journal of Pragmatics*, 33 849-878.
- Oald 2005. Oxford Advanced Learner's Dictionary of Current English. In: Whemeier, S., Mcintosh, C., Turnbull, J. & Ashby, M. (eds.) *Oxford Advanced learner's Dictionary* 7th ed. Oxford and New York: Oxford University Press.
- Ortega, L. (2009). *Understanding Second Language Acquisition*. London & New York: Routledge.
- Polanyi, L. & R. Scha. (1983). "The syntax of discourse". *Text*, 3 261-270.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Rayson, P. (1993-2014). *Log-Likelihood and Effect Size Calculator* [Online]. Lancaster Lancaster University: University Centre for Computer Corpus Research on Language Available: <http://ucrel.lancs.ac.uk/llwizard.html> [Accessed 03 March 2016].
- Redeker, G. (1990). "Ideational and Pragmatic Markers of Discourse Structure". *Journal of Pragmatics*, 14 367-381.

- Redeker, G. (1991). "Linguistic markers of discourse structure". *Linguistics*, 29 1139-1172.
- Romaine, S. & D. Lange. (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.
- Romero Trillo, J. 2001. The pragmatic "dis-functions" of discourse markers in non-native speakers of English: a corpus-driven approach. *8th Biennial Congress of the International Association for Dialogue Analysis (IADA)*. Göteborg.
- Romero Trillo, J. (2002). "The pragmatic fossilization of discourse markers in non-native speakers of English". *Journal of Pragmatics*, 34 769-784.
- Schiffrin, D. (1986). "The functions of *and* in discourse ". *Journal of Pragmatics* 10 (1), 41-66.
- Schiffrin, D. (1987). *Discourse Markers* Cambridge and New York: Cambridge University Press
- Schiffrin, D., D. Tannen & H. E. Hamilton. (2001). *The Handbook of Discourse Analysis* Malden and Oxford Blackwell Publishers Ltd.
- Schourup, L. (1985). *Common discourse particles in English conversation*. New York: Garland.
- Schourup, L. (1999). "Discourse markers". *Lingua*, 107 227-265.
- Schourup, L. (2001). "Rethinking well". *Journal of Pragmatics*, 33 1025-1060.
- Schourup, L. (2011). "The discourse marker *now*: A relevance-theoretic approach". *Journal of Pragmatics*, 43 2110-2129.
- Scott, M. 2012. Wordsmith Tools 6.9. Stroud: Lexical Analysis Software.
- Selinker, L. (1972). "Interlanguage". *International Review of Applied Linguistics*, 10 209-231.
- Selinker, L. (2014). "Interlanguage 40 years on. Three themes from here". In: Han, Z. & Tarone, E. (eds.), *Interlanguage. Forty years later*. Amsterdam and Philadelphia: John Benjamins.
- Sinclair, J. (1996). "EAGLES. Preliminary Recommendations on Corpus Typology. ".
- Sommerseth, H. (2013). *English subject curriculum* [Online]. Ministry of Education and Research. Available: <http://www.udir.no/kl06/ENG1-03> [Accessed 12.04 2016].
- Stubbs, M. (1983). *Discourse Analysis*. Chicago, IL: The University of Chicago Press.
- Svartvik, J. (1980). "*Well* in conversation". In: S. Greenbaum, G. L., And J. Svartvik (ed.) *Studies in English Linguistics for Randolph Quirk*. London: Longman.

- Tan, M. (2005). "Authentic language or language errors? Lessons from a learner corpus". *ELT Journal*, 59 (2), 126-134.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam & Philadelphia: John Benjamins Publishing Company
- Traugott, E. C. (1995). "Subjectification in grammaticalization". In: Stein, D. & Wright, S. (eds.), *Subjectivity and subjectivisation in language* Cambridge: Cambridge University Press.
- Unaldi, I. (2013). "Overuse of Discourse Markers in Turkish English as a Foreign Language (EFL) Learners' Writings: The Case of *I think* and *in my opinion*". *Anthropologist*, 16 (3), 575-584.
- Underhill, R. (1988). "*Like* is, like, focus". *American Speech*, 63 234-246.
- Van Lier, L. (1989). "Reeling, writing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversation ". *TESOL Quarterly*, 23 489-508.
- Watts, R. J. (1988). "A relevance-theoretic approach to commentary pragmatic markers: the case of *actually*, *really* and *basically*". *Acta Linguistica Hungarica*, 38 (1-4), 235-260.
- Weydt, H. (1969). *Abtönungspartikel*. Bad Homburg: Gehlen.
- Wilson, A. (2013). "Embracing Bayes factors for key item analysis in corpus linguistics". In: Bieswanger, M. & Koll-Stobbe, A. (eds.), *New approaches to the study of linguistic variability*. Frankfurt: Peter Lang.
- Wouk, F. (2001). "Solidarity in Indonesian conversation: The discourse marker *ya*". *Journal of Pragmatics*, 33 171-191.
- Östman, J.-O. (1981). *'You Know': A discourse-functional approach*. Amsterdam: John Benjamins
- Östman, J.-O. (1982). "The symbiotic relationship between pragmatic particles and impromptu speech". In: Enkvist, N. E. (ed.) *Impromptu speech: A symposium*. Åbo: Publications of the Research Institute of the Åbo Akademi Foundation.
- Östman, J.-O. (1995). "Pragmatic particles twenty years later ". In: Warvik, B., Tanskanen, S.-K. & Hiltunen, R. (eds.), *Organization in Discourse. Proceedings from the Turku Conference*. Turku: University of Turku

Appendix 1: LINDSEI-NO metadata

LINDSEI-NO METADATA

LINDSEI ID	DM/like	DM/well	DM/you know	DM/I mean	Comments	Gender	Age	Current studies	Institution	Current year of study	Medium of instruction	Years Eng. school	Stay in Eng speaking country	Where	How long (months)	When	Native language	Father's 1L	Mother's 1L	Home language (with % if more than one)	Primary school Medium of instruction
NO001	3	22	10	0		Male	26	Teacher training education	Hedmark University College	3	English	10	TRUE	England	4	2003	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO002	2	1	0	0		Female	23	Teacher training education (Music and English)	Hedmark University College	3	English and Norwegian	8	FALSE		0		Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO003	1	22	5	0		Female	43	Teacher training education	Hedmark University College	3	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO004	16	0	4	0		Female	23	Teacher training education	Hedmark University College	3	English	8	FALSE				Norwegian	Norwegian	Norwegian	English (60%) Norwegian (40%)	Norwegian
NO005	1	1	0	1		Female	23	Teacher training education	Hedmark University College	3	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO006	1	0	0	0		Female	23	Teacher training education	Hedmark University College	3	English	12	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO007	5	9	2	0		Female	23	Teacher training education	Hedmark University College	4	English	11	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO008	0	1	2	0		Female	26	Teacher training education	Hedmark University College	4	English	11	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO009	2	0	0	1		Female	40	Teacher training education	Hedmark University College	3	English and Norwegian	10	TRUE	England	1	1985	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO010	1	10	17	0		Male	40	English Foundation Course (1-year)	Hedmark University College	1	English	8	TRUE	[1] Michigan, USA [2] Reading, England	48	[1] 1971-1974 [2] 1983-1984	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO011	1	3	2	0		Female	47	English Foundation Course (1-year)	Hedmark University College	1	English	9	TRUE	England	11	2006-2007	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO012	6	6	0	0		Female	19	English Foundation Course (1-year)	Hedmark University College	1	English	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO013	0	4	3	0		Female	42	English Foundation course (1-year) and Social Pedagogy	Hedmark University College	1	English and Norwegian	9	TRUE	Illinois, USA	18	1986-1987	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO014	22	22	0	0		Female	22	English Foundation course (1-year)	Hedmark University College	1	English	11	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO015	4	11	0	1		Female	23	Teacher Training Education	Hedmark University College	3	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO016	6	4	0	0		Female	22	Teacher Training Education	Hedmark University College	3	English	7	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO017	19	9	1	0		Female	21	English Foundation course (one year)	Hedmark University College	1	English	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO018	3	1	2	0		Female	20	English Foundation Course (one-year)	Hedmark University College	1	English	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO019	5	11	5	3		Male	23	English Foundation Course (one-year)	Hedmark University College	1	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO020	5	0	1	0		Male	20	English Foundation Course (one-year)	Hedmark University College	1	English	8	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO021	1	11	1	0		Female	21	Teacher Training Education	Hedmark University College	3	English	8	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO022	6	1	0	27		Male	29	Teacher Training Education	Hedmark University College	2	English		TRUE	New Jersey, USA	12	1995	Norwegian	Norwegian	English	Norwegian 70% English 30%	Norwegian
NO023	6	4	13	0		Female	39	English Foundation course (one year)	Hedmark University College	5	English	9	FALSE	Seattle, WA, USA	12	1979	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO024	3	0	1	0		Female	19	English Foundation course (one-year)	Hedmark University College	1	English	11	TRUE	Portsmouth, England	11	2006-2007	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO025	27	4	8	5		Female	22	English Foundation course (one-year), VUE	Hedmark University College	2	English	13	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian

LINSEI ID	DM/like	DM/well	DM/you	DM/I	Comments	Gender	Age	Current studies	Institution	Current year of study	Medium of instruction	Years Eng speaking	Stay in Eng speaking country	Where	How long (months)	When	Native language	Father's LL	Mother's LL	Home language (with % if more than one)	Primary school Medium of instruction
NO026	5	8	1	0		Male	32	Teacher Training Education	Hedmark University College	4	English	8	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO027	1	0	0	0		Female	22	Bachelor in Culture and language	Hedmark University College	2	English	12	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO028	1	15	1	0		Female	46	English foundation course (one-year)	Hedmark University College	1	English	9	TRUE	USA	4	1983 and 1984	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO029	6	7	9	2		Female	41	English foundation course (one-year)	Hedmark University College	1	English	10	TRUE	USA	12	1986-1987	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO030	19	1	1	2		Female	26	English foundation course (part of Bachelor's)	Hedmark University College	2	English	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO031	69	6	26	2	Mostly quote like (I was like, he was like)	Female	22	English foundation course	Hedmark University College	3	English	9	TRUE	Sydney, Australia	12	2005-2010	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO032	1	7	1	0		Male	20	English foundation course	Hedmark University College	1	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO033	25	3	0	1		Female	21	English foundation course	Hedmark University College	1	English	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian 50%, Finnish 30%, English 20%	Norwegian
NO034	8	11	12	0		Male	24	Teacher training education	Hedmark University College	4	English	11	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO035	4	4	0	0		Female	22	Teacher training education	Hedmark University College (foundation course)	4	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO036	21	1	2	0		Male	22	English foundation course	Hedmark University College	1	English	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO037	13	20	4	0		Male	19	Teacher training education (GUU 5-10), English 60 credits	Hedmark University College	1	English	9	FALSE	US, Mid-west 2	2000 and 2006		Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO038	0	0	0	0		Female	45	English foundation course	Hedmark University College	1	English	10	TRUE	Newbury, England	1	1981	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO039	2	8	2	0		Female	23	Teacher training education (GUU 5-10), 60 credits English total	Hedmark University College	1	English and Norwegian	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO040	10	1	0	0		Male	22	Teacher training education (GUU 5-10), 60 credits English total	Hedmark University College	1	English and Norwegian	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO041	7	1	0	0		Female	21	Teacher training education (GUU 5-10), 60 credits English total	Hedmark University College	1	English and Norwegian	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO042	7	0	5	0		Female	22	Teacher training education (GUU 5-10), 60 credits English total	Hedmark University College	1	English and Norwegian	10	TRUE	USA	12	1) 2008 (3 months), 2) 2010/11 (9 months)	Norwegian	Norwegian	Norwegian	Norwegian 98%, German 2%	Norwegian
NO043	25	2	1	0		Female	24	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	1	English and Norwegian	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO044	2	4	0	0		Female	21	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	2	English and Norwegian	11	FALSE				Norwegian	Norwegian	Norwegian	Norwegian 98%, Danish 2%	Norwegian
NO045	30	34	1	22		Male	23	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	2	English and Norwegian	9	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO046	2	1	3	0		Male	28	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	2	English and Norwegian	10	TRUE	USA	2	2005	Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO047	1	9	0	0		Male	28	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	2	English and Norwegian	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO048	0	3	0	0		Male	21	Teacher training education (GUU 5-10), 60 credits English	Hedmark University College	2	English and Norwegian	10	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO049	6	1	1	0		Female	23	English foundation course (1 year)	Hedmark University College	1	English	13	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
NO050	1	15	0	0		Female	19	English foundation course (1 year)	Hedmark University College	1	English	12	FALSE				Norwegian	Norwegian	Norwegian	Norwegian	Norwegian
Average per speaker	8	6.38		2.84																	

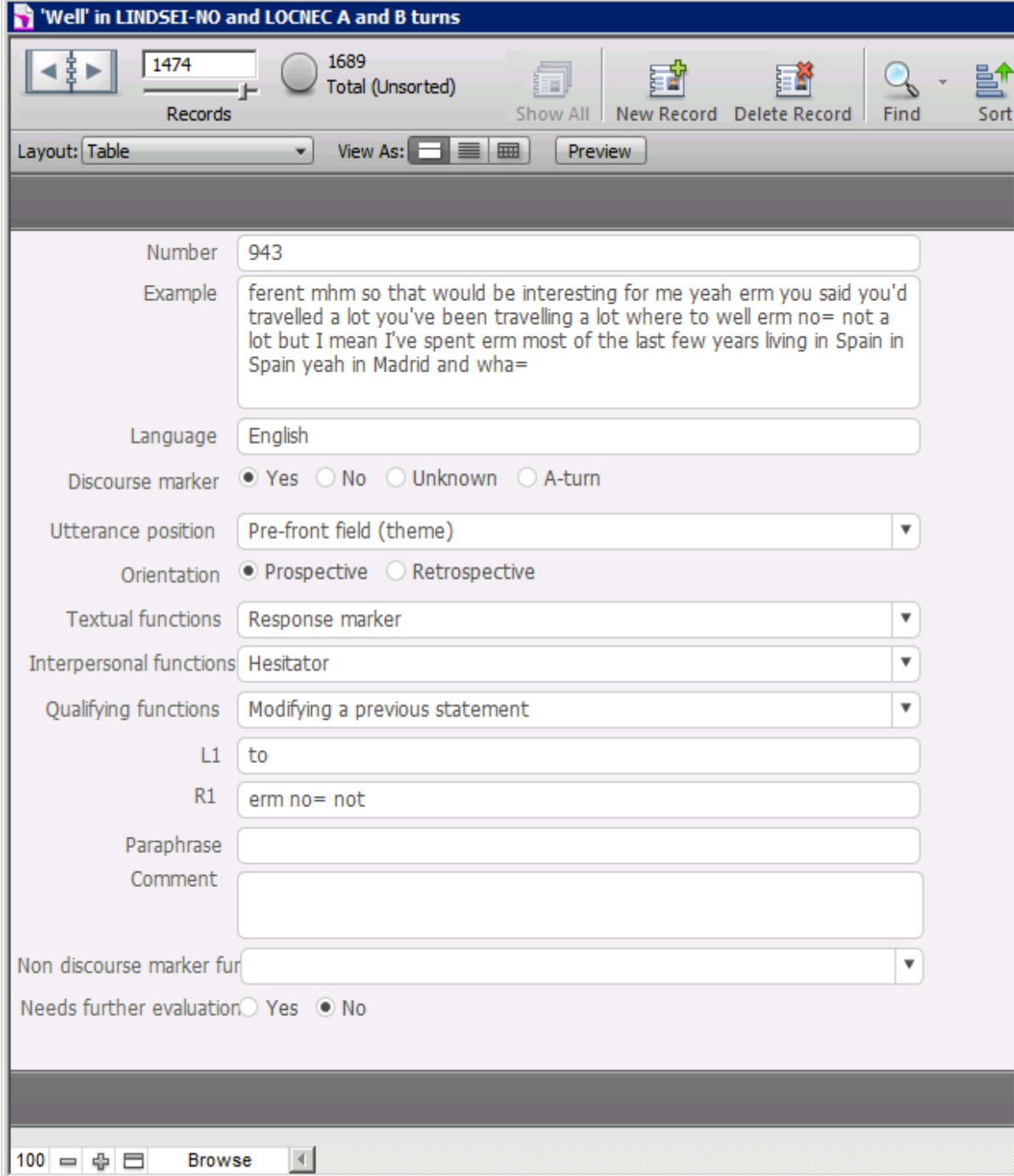
Appendix 2: LOCNEC metadata

LOCNEC METADATA

LOCNEC ID	Introduction topic	DM like	DM well known	duration	age	gender	country	language	home lang. 1	home lang. 2	insti	schooling	uniling	monthse	olang1	olang2	olang3	Interviewer's gender	Interviewer's language	Interviewer's lang1	Interviewer's lang2	status	date	comments
EN001	Film	2	11	7	13.51	20	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-11-27	Interviewer's olang3 = German
EN002	Film	41	24	4	16.2	19	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-11-27	Interviewer's lang1 = German
EN003	Film	25	8	11	16.02	19	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-11-28	Interviewer's olang3 = German
EN004	Experience	9	17	3	17.19	18	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-12-01	Interviewer's lang1 = German; nonparallel3 = Interviewer's
EN005	Experience	8	2	1	12.59	18	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-12-01	Interviewer's olang3 = German
EN006	Country	8	7	2	18.09	19	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-12-04	Interviewer's olang3 = German
EN007	Experience	1	10	9	15.53	25	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Vaguely familiar	1995-12-04	Interviewer's lang1 = German; Interviewer's olang3 =
EN008	Country	20	10	5	16.26	21	Female	Great Britain	English	English	Code699	0	0	0	0	French		Female	French	English	Dutch	Unfamiliar	1995-12-15	Interviewer's olang3 = German
EN009	Film	10	7	17	18.3	19	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-12-05	Interviewer's olang3 = German
EN010	Film	71	15	27	15.37	20	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-12-04	Interviewer's olang3 = German
EN011	Experience	3	10	39	18.44	24	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-11-30	Interviewer's lang1 = German; Interviewer's olang3 =
EN012	Country	12	8	9	15.25	24	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1995-11-29	Interviewer's lang1 = German; Interviewer's olang3 =
EN013	Country	19	19	0	17.07	22	Female	Great Britain	English	English	Code699	0	0	0	0	French	German	Female	French	English	Dutch	Unfamiliar	1995-12-12	Interviewer's olang3 = German
EN014	Country	10	11	17	14.4	20	Male	Great Britain	English	English	Code699	0	0	0	0	Italian		Female	French	English	Dutch	Unfamiliar	1996-01-24	Interviewer's lang1 = German; Interviewer's olang3 =
EN015	Film	6	0	26	20.46	22	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-24	Interviewer's olang3 = German
EN016	Country	4	10	3	15.46	23	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-24	Interviewer's olang3 = German; Interviewer's olang3 =
EN017	Country	5	7	10	15.24	20	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-26	Interviewer's lang1 = German; Interviewer's olang3 =
EN018	Country	10	14	14	16.53	19	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-25	Interviewer's olang3 = German
EN019	Experience	24	14	27	18.44	20	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-31	Interviewer's lang1 = German; Interviewer's olang3 =
EN020	Film	1	3	15	12.28	19	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-31	Interviewer's olang3 = German
EN021	Film	6	15	4	16.5	19	Male	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-31	Interviewer's olang3 = German
EN022	Film	41	20	33	18.57	25	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-31	Interviewer's lang1 = German; Interviewer's olang3 =
EN023	Country	0	8	4	11.37	20	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-25	Interviewer's olang3 = German
EN024	Country	18	9	2	15.56	19	Female	Great Britain	English	English	Code699	0	0	0	0			Female	French	English	Dutch	Unfamiliar	1996-01-24	Interviewer's olang3 = German
EN025	Country	23	11	13	15.26	20	Female	Great Britain	English	English	Code699	0	0	0	0	Chinese		Female	French	English	Dutch	Unfamiliar	1996-01-24	Interviewer's lang1 = German; Interviewer's olang3 =

LOGNEC ID	Introduction topic	DM like	DM well	DM you know	duration	age	gender	country	language	home lang 1	home lang 2	inst	schooling	uneng	monthling	olang1	olang2	olang3	Interviewer's gender	Interviewer's language	Interviewer's lang1	Interviewer's lang2	status	date	comments	
END26	Film	5	13	25	24.14		24	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-08	Interviewer's lang3 = Interviewer's lang3 =
END27	Film	20	18	1	14.58		21	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1996-02-02	Gerrman Interviewer's lang3 =
END28	Experience	3	9	12	13.19		18	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-11-05	Gerrman Interviewer's lang3 =
END29	Film	0	13	12	14.07		19	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-05	Gerrman Interviewer's lang3 =
END30	Country	22	6	0	15.29		22	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-04	Gerrman Interviewer's lang3 =
END31	Film	3	14	3	14.35		28	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-12-07	Gerrman Interviewer's lang3 =
END32	Country	18	18	1	14.17		21	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-07	Gerrman Interviewer's lang3 =
END33	Film	5	14	6	15.13		30	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-11-23	Gerrman Interviewer's lang3 =
END34	Experience	6	13	15	16.39		22	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Familiar	1995-11-22	Gerrman Interviewer's lang3 =
END35	Country	17	4	4	16.47		19	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-11	Gerrman Interviewer's lang3 =
END36	Country	6	5	10	15.51		25	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-12	Gerrman Interviewer's lang3 =
END37	Film	10	5	19	11.41		21	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-11-28	Gerrman Interviewer's lang3 =
END38	Experience	52	21	5	18.17		26	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-12-01	Gerrman Interviewer's lang3 =
END39	Country	10	13	12	18.19		23	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-11-22	Gerrman Interviewer's lang3 =
END40	Country	1	5	1	15.11		22	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-12-14	Gerrman Interviewer's lang3 =
END41	Country	37	13	16	16.27		23	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1996-02-02	Gerrman Interviewer's lang3 =
END42	Film	5	15	10	14.17		27	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-11-23	Gerrman Interviewer's lang3 =
END43	Film	4	16	42	18.26		28	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1995-11-22	Gerrman Interviewer's lang3 =
END44	Country	61	15	60	15.08		19	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-11-28	Gerrman Interviewer's lang3 =
END45	Country	6	6	12	17.58		30	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1995-11-27	Gerrman Interviewer's lang3 =
END46	Experience	5	6	38	16.36		20	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1996-03-03	Gerrman Interviewer's lang3 =
END47	Country	1	1	1	12.06		18	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1997-03-03	Gerrman Interviewer's lang3 =
END48	Country	36	4	5	12.17		22	Female	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1997-03-05	Gerrman Interviewer's lang3 =
END49	Film	30	18	6	19.35		19	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Vaguely familiar	1997-02-26	Gerrman Interviewer's lang3 =
END50	Country	28	14	2	16.29		19	Male	Great Britain	English	English			0	0	0				Female	French	English	Dutch	Unfamiliar	1997-02-25	Gerrman Interviewer's lang3 =
Average per Speaker		15.88	13.62	12.64																						

Appendix 3: Screenshot of FilemakerPro file



'Well' in LINDSEI-NO and LOCNEC A and B turns

Records: 1474 / 1689 Total (Unsorted)

Layout: Table View As: [Table View] [List View] [Grid View] Preview

Number: 943

Example: ferent mhm so that would be interesting for me yeah erm you said you'd travelled a lot you've been travelling a lot where to well erm no= not a lot but I mean I've spent erm most of the last few years living in Spain in Spain yeah in Madrid and wha=

Language: English

Discourse marker: Yes No Unknown A-turn

Utterance position: Pre-front field (theme)

Orientation: Prospective Retrospective

Textual functions: Response marker

Interpersonal functions: Hesitator

Qualifying functions: Modifying a previous statement

L1: to

R1: erm no= not

Paraphrase:

Comment:

Non discourse marker fur:

Needs further evaluation: Yes No

100 Browse

Appendix 4: Transcription guidelines

Transcription guidelines³⁸

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 `` or ``.

e.g. `<A>` okay so which topic have you chosen ``

`` the film or play that I thought was particularly good or bad really ``

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. `` yeah I went on a bus to London once and I'll never `<overlap />` do it again ``

`<A>` `<overlap />` that's even worse ``

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. `` (erm) .. it's a British film there aren't many of those these days ``

³⁸ Copied from University of Louvain, *Centre for English Corpus Linguistics* [URL], <https://www.uclouvain.be/en-307849.html> (Accessed April 28th, 2016)

6. Filled pauses and backchannelling

Filled pauses and backchannelling are put between brackets and marked as (eh) [brief], (er), (em), (erm), (mm), (uhu) and (mhm). No other fillers should be used.

e.g. yeah . well Namur was warmer (er) it was (eh) a really little town

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. <X> they're just begging <XX> there's there's honestly he did a course .. for a few weeks

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. I went to see a<?> friend at university there and stayed

Unclear names of towns or titles of films for example may be indicated as <name of city> or <title of film>.

e.g. where else did we go (er) <name of city> it's in Bolivia

8. Anonymisation

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

9. Truncated words

Truncated words are immediately followed by an equals sign.

e.g. it still resem= resembled the theatre

10. Spelling and capitalisation

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. yes not really I did sort of basic G C S E French and German

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

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. an awful lot of people complain and say well the grants were two thousand two hundred

15. Foreign words and pronunciation

Foreign words are indicated by <foreign> (before the word) and </foreign> (after the word).

e.g. we couldn't go with (er) knives and so on <foreign> enfin </foreign> we were (er)

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. I didn't have the (erm) . <foreign> distinction </foreign>

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. that's something I'll I'll plan to: to learn

(b) Articles

- when pronounced as [ei], the article *a* is transcribed as a[ei];

e.g. and it's about (erm) . life in a[ei] (eh) public school in America I think

- when pronounced as [i:], the article *the* is transcribed as the[i:].

e.g. and the[i:] villa we were staying in was in one of the valleys

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. <starts laughing> I don't have to assess it I only have to write it <stops laughing>

18. Nonverbal vocal sounds

Nonverbal vocal sounds are enclosed between angle brackets.

e.g. I hope so I've I've got some <coughs> friends out there

e.g. so I went back into Breda .. and sat down again <imitates the sound of a guitar>

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
<somebody enters the room>
 hi

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

Appendix IV: LINDSEI Picture description task

LINDSEI picture description task³⁹



³⁹ Copied from Gilquin, G., S. De Cock & S. Granger. (2010). *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain page 109.