“I just need to like see what I can do” –
A contrastive study of hedging strategies in English and Norwegian informal spoken conversations

Stine Hulleberg Johansen

Dissertation presented for the degree of Philosophiae Doctor (PhD)
Department of Literature, Area Studies and European Languages
Faculty of Humanities
UNIVERSITY OF OSLO
2020
Title: “I just need to like see what I can do” – A contrastive study of hedging strategies in English and Norwegian informal spoken conversations
Acknowledgements

Many people have contributed to the completion of this thesis, and I would like to express my heartfelt gratitude to all of them here.

First, I would like to express my very great appreciation and gratitude to my supervisor, Hilde Hasselgård. Her expertise and sound advice have been instrumental in enhancing the quality of this thesis. Without her knowledge of the field, dedication and patience, this thesis would not have been possible. I am particularly grateful for our sessions deconstructing feedback from reviewers and for her restoring my faith in myself.

Second, I would like to extend my warmest thanks to my co-supervisor, Hildegunn Dirdal. Without her sharp eye and insightful perspective and comments, the completion of this thesis would not have been possible. I am also greatly indebted to my midway evaluator Ingrid Kristine Hasund whose feedback helped me shape the project into what it is.

Next, I would like to thank the academic staff of the Department of English Language for their general support throughout the PhD process, especially Signe Oksefjell Ebeling and the members of the research group English Language and Corpus Linguistics Research (ELC). I would also like to thank my office cohabitants and partners in crime, Øyvind Thormodsæter and Sabina Nedelius. I am grateful for all our conversations and for sharing the ups and downs of being a PhD candidate with you.

I am particularly grateful for Anne Schad Bergsaker’s support on statistical calculations in R. I appreciate your patience, availability and pedagogical approach. I also wish to acknowledge the help Joanna Nykiel provided with R and the technical support Kristin Hagen offered on the Norwegian corpus data.

I would like to offer my special thanks to Tina Skouen for caring about the wellbeing of the PhD candidates at ILOS and to Sarah Salameh for taking care of all the practicalities and for always being available for questions.

I would also like to thank my colleagues at OsloMet for all the support and for checking up on me. You have played an important role in making this possible. I am particularly grateful for all the help and support from Siri Aanstad. Thank you for taking time to read parts of this thesis and for all our walks and discussions. Thank you also to Florissa Abreu for your feedback on parts of this thesis.

Finally, this thesis would not have been possible without the support I received from my family and friends. Thank you for always encouraging me and for being my safe haven at the end of a long day.
Det er den draumen

Det er den draumen me ber på
at noko vedunderleg skal skje,
at det må skje –
at tidi skal opna seg
at hjarta skal opna seg
at dører skal opna seg
at berget skal opna seg
at kjeldor skal springa –
at draumen skal opna seg
at me ei morgenstund skal glida inn
på ein våg me ikkje har visst um.

Olav H. Hauge
Table of Contents

Acknowledgements ........................................................................................................ iii

Chapter 1 Introduction .................................................................................................... 1
  1.1 Topic and context ........................................................................................................ 1
  1.2 Overall aim .................................................................................................................. 2
  1.3 Relevance and importance ......................................................................................... 2
  1.4 Overall research design ............................................................................................. 5
  1.5 Thesis outline ............................................................................................................. 7

Chapter 2 Previous research ............................................................................................ 9
  2.1 Hedging ....................................................................................................................... 9
    2.1.1 Previous research on hedging .............................................................................. 11
    2.1.2 Hedging in related areas of research ................................................................. 21
    2.1.3 Summary of previous research on hedging and related concepts ....................... 28
  2.2 Corpus pragmatics ..................................................................................................... 31
    2.2.1 Corpus linguistics ............................................................................................... 32
    2.2.2 Learner corpus research ...................................................................................... 37
    2.2.3 Pragmatics ........................................................................................................... 39
    2.2.4 Contrastive, cross-cultural, intercultural and interlanguage pragmatics ............... 42
    2.2.5 The influence of corpus linguistics and pragmatics on corpus pragmatics .......... 43
    2.2.6 Learner corpus pragmatics ................................................................................. 48

Chapter 3 Overview of the three phases of the project ................................................ 51
  3.1 Phase 1: Developing a method and identifying a probe (article 1) ....................... 51
  3.2 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3) 53
  3.3 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4) 54

Chapter 4 Presentation of the data ................................................................................ 57
  4.1 Why study spoken corpora? ....................................................................................... 57
  4.2 The issue of comparability and the choice of corpora ............................................ 58
  4.3 Phase 1: Developing a method and identifying a probe (article 1) ....................... 59
  4.4 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3) 61
  4.5 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4) 67

Chapter 5 Methodological considerations .................................................................... 71
  5.1 Phase 1: Developing a method and identifying a probe (article 1) ....................... 71
    5.1.1 Developing a method ......................................................................................... 71
    5.1.2 Identifying a probe ......................................................................................... 74
5.1.3 No/nei .................................................................................................................... 76
5.1.4 But/men .................................................................................................................. 83
5.1.5 Comparing no/nei and but/men ........................................................................... 88

5.2 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3) .......... 92
  5.2.1 Procedure .............................................................................................................. 92
  5.2.2 Analysis of the data ............................................................................................ 97

5.3 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4) .... 100
  5.3.1 Procedure ............................................................................................................ 101
  5.3.2 Analysis of the data ............................................................................................ 104

Chapter 6 Summaries of articles 1–4 ................................................................................. 109
  Article 1: Exploring the use of probes in a corpus pragmatic study of hedging strategies ................. 109
  Article 2: Classifying hedging strategies in Norwegian and English informal spoken conversations .... 109
  Article 3: “I guess anyone would do that wouldn't they?” How do native speakers of Norwegian and English hedge in informal conversations? ............................................................. 110
  Article 4: Hedging in spoken conversations by Norwegian Learners of English .............................. 110

Chapter 7 Summary of results and conclusions ........................................................................ 111
  7.1 Summary of main findings ...................................................................................... 111
  7.2 Contributions .......................................................................................................... 114
  7.3 Limitations and potential for further research ............................................................. 116

References .......................................................................................................................... 119

Appendix 1 A working guide for analysis - English .............................................................. 131
Appendix 2 Working guide for Analysis – Norwegian .......................................................... 147
Appendices references ......................................................................................................... 163
Article 1 ............................................................................................................................ 167
Article 2 ............................................................................................................................ 193
Article 3 ............................................................................................................................ 219
Article 4 ............................................................................................................................ 245
List of tables

Table 2-1 Variations in the definitions of hedging ................................................................. 10
Table 4-1 Number of words in the direct speech samples ......................................................... 61
Table 4-2 Key figures in the Norwegian and English corpora ................................................... 62
Table 4-3 Overview of relevant transcription symbols ............................................................... 67
Table 4-4 Total number of words in LINDSEI-no and LOCNEC ............................................... 69

Table 5-1 Filters applied in the ENPC and the number of instances of no/nei ............................ 80
Table 5-2 Filters applied in the NDC, NoTa, BB and BNC2014 (pilot study) ............................. 81
Table 5-3 Numbers of no/nei in its refusing sense and co-occurring with hedging strategies .......... 83
Table 5-4 Filters applied in the ENPC and number of instances of but/men ............................... 85
Table 5-5 Filters applied in the NDC, NoTa, BB and BNC2014 (main contrastive study) .......... 86
Table 5-6 Number of uses of but and men in their contrastive sense ........................................ 87
Table 5-7 Summary of the selection criteria for the choice of probe ......................................... 89
Table 5-8 Number of contrastive uses of the probe and number of the probe co-occurring with hedging strategies .......................................................... 95
Table 5-9 An overview of the categories applied in article 2 ..................................................... 98
Table 5-10 Overview of the categories of forms ...................................................................... 100
Table 5-11 Overview of the 10 most frequent forms retrieved from the BNC2014 in phase 2 .... 101
Table 5-12 Overview of the hedging functions of the forms with corpus examples ................. 101
Table 5-13 Overview of the number of forms in both corpora .................................................. 104
Table 5-14 Variation coefficient for the expressions investigated in article 4 ........................... 106

List of figures

Figure 1-1 Overview of the project phases ............................................................................... 5
Figure 2-1 A timeline summarising previous research on hedging in English (below the line) and Norwegian (above the line) ........................................................................ 29
Figure 2-2 Hedging and overlapping concepts ........................................................................ 29
Figure 2-3 Corpus-informed–corpus-driven continuum ............................................................. 37
Figure 2-4 Overview of some of the characteristic features of corpus linguistics, pragmatics and corpus pragmatics ............................................................. 44
Figure 2-5 Form-to-function–function-to-form continuum ....................................................... 47

Figure 4-1 The structure of the English-Norwegian Parallel Corpus (Johansson et al., 2002) .............. 60

Figure 5-1 Placing the approach within the corpus-based–corpus-driven/form-to-function–function-to-form spectrum ................................................................. 74
Figure 5-2 Illustration of the probe in context ........................................................................ 96
Chapter 1 Introduction

1.1 Topic and context

The topic of this thesis is hedging in Norwegian and English informal spoken conversations. Hedging is an essential part of language and an important tool in the linguistic toolbox enabling speakers to communicate something beyond the propositional content, and to guide the hearer in the interpretation of what is said. By hedging, speakers may signal caution, tentativeness or possibility, or tone down, soften or disarm their utterances to lessen the impact on the hearer. In example (1.1), just serves this downtoning function, and seems indicates tentativeness. Hedging may also be used to reduce speakers’ commitment to the message to shield them from potential criticism, as exemplified in (1.2). The use of generally makes the opinion less categorical by implying that the speaker likes the actor Julia Stiles for the most part, leaving the speaker with a way out if the hearer disagrees.

(1.1)
[...] it just seems that Mary\(^1\) [...] hasn't had a lot of friends over lately
BNC2014 SZFG no. 375

(1.2)
but I generally like Julia Stiles
BNC2014 SZ7Y no. 335

Just, seems, and generally in (1.1) and (1.2) are only three examples of realisations of the pragmatic function of hedging. Such realisations are referred to as hedging strategies. Hedging strategies may take nearly any linguistic or para-linguistic form, i.e. ranging from words or phrases to gestures or facial expressions, and are ubiquitous in both informal and formal spoken and written discourse. The ability to hedge forms an integral part of speakers’ pragmatic competence, i.e. the ability to communicate the intended message with all its nuances in any socio-cultural context (Fraser, 2010, p. 15).

Hedging is also a complex phenomenon which requires a high degree of sophistication and skill to master. The hedging quality is not an inherent property of individual linguistic devices, and often it can

\(^1\) Fictive name.
be challenging to identify what exactly in an utterance causes the hedging effect (Stenström, 1994). Moreover, although hedging is a phenomenon which occurs in most, if not all, languages, the type of strategies used to express hedging varies across language cultures. The situations in which it is appropriate to hedge may also vary between language cultures. This makes it particularly challenging for language learners who may speak grammatically correctly, but fail to communicate their intended meaning, because they do not understand the underlying social conventions of the target language (Nikula, 1997; Romero-Trillo, 2018).

1.2 Overall aim

The overall aim of this thesis is to provide more insight into hedging as an aspect of pragmatic usage by studying linguistic hedging strategies in English and Norwegian informal spoken conversations. By using a methodological approach associated with the emerging field of corpus pragmatics, the thesis aims to discover which strategies native speakers of Norwegian and English use to express hedging linguistically in their respective languages, including the similarities and differences between them. A further aim is to provide insight into the use of hedging strategies by Norwegian learners of English in comparison with native speakers. It is believed that insight into the learners’ appropriate (or inappropriate) use of hedging strategies can shed a light on their pragmatic competence. An understanding of how hedging is performed across languages and cultures may reduce the risk of communicative failure.

1.3 Relevance and importance

The field of hedging research comprises nearly 50 years of research, and the understanding of hedging has developed from a logico-semantic concept into a fully-fledged pragmatic concept. Hedging has been studied across genres of spoken and written discourse, and from the perspective of other related concepts within linguistics such as vague language, pragmatic markers, modality, and more. Nevertheless, there are still aspects of hedging that remain underexplored. First, there is a need to study hedging in the light of recent methodological developments, i.e. the use of corpora to study pragmatic phenomena (Romero-Trillo, 2014). The empirical nature of corpus studies can benefit studies of pragmatic phenomena methodologically (Rühlemann & Clancy, 2018, p. 241). Until recently, corpus linguistics and pragmatics were regarded as mutually exclusive (Romero-Trillo, 2008), mainly because core features of pragmatics are challenging to catch with corpus methodology (Taavitsainen & Jucker, 2015, p. 12). However, the desire to combine quantitative aspects characteristic of corpus linguistics
and qualitative aspects characteristic of pragmatics has resulted in the emergence of corpus pragmatics as a new subfield within linguistics. Studying hedging using corpus pragmatics approaches may shed new light on how hedging is used in authentic communication situations.

Within corpus pragmatics, the majority of studies have taken a form-to-function perspective, i.e. starting from pre-defined lexical items and investigating the functions of those items in corpora. A consequence of this may be that forms remain undiscovered. A limited number of studies have taken the function-to-form perspective, where a pragmatic function is the starting point, and forms realising the function are retrieved from a corpus. This is mainly because corpora are constructed to access forms rather than functions. Due to the limited number of corpora tagged for pragmatic functions (Weisser, 2018), there is a need to find “robust workarounds” to access pragmatic functions in corpora (O'Keeffe, Clancy, & Adolphs, 2020, p. 9). Therefore, more studies are needed to explore and test methodological approaches for accessing pragmatic phenomena in corpora. Furthermore, in corpus-based research, studies on written corpora greatly outnumber those on spoken corpora, mainly due to the greater availability of written corpora. The compilation of spoken corpora, by contrast, has been less widespread because of the costs involved. However, in recent years, due to new methods of accessing participants, such as crowdsourcing, as well as better tools for transcription, more corpora of spoken informal conversations have seen the light of day, for example the spoken BNC2014 with over 11 million words launched in 2017. The availability of such corpora opens new trajectories of language research which may lead to new insight into language use in general, and the use of hedging strategies in particular.

Second, there is also a need for studies on hedging across languages as such studies can “shine light on similarities and differences between languages and cultures” (Kaltenböck, Mihatsch, & Schneider, 2010, p. 2). By comparing two languages, new insights are given into both languages compared which most probably would have been missed in a monolingual study (Aijmer & Altenberg, 1996). Contrastive studies are also beneficial in discovering language-specific or culture-specific features as well as features that are shared across languages and cultures. Previous research on hedging has mainly been focused on how hedging is expressed in English, and as a result, the majority of classificatory frameworks and descriptions of types of hedging are based on English data. Comparing English with other languages may shed new light on existing presuppositions about hedging and challenge the general applicability of existing classificatory frameworks and descriptions.

---

2 Pre-publication manuscript.
Furthermore, the use of hedging strategies in languages other than English remains relatively underexplored (Vold, 2006, p. 63). This is particularly true of Norwegian where the research on hedging has been intertwined with the research on the properties of pragmatic/modal particles, and where there are few studies focusing exclusively on hedging. Thus, there is a need for studies that looks at how hedging is expressed in Norwegian beyond the use of pragmatic/modal particles.

Potential differences between languages also make hedging interesting to study from the perspective of second or foreign language learners, and studying pragmatic phenomena such as hedging in learner language has been highlighted as an area in need of further scholarly attention (Paquot & Plonsky, 2017, p. 87). Acquiring pragmatic competence in one’s first language is generally “an unconscious process which co-occurs with socialisation and acculturation into the community into which one is born” (Holmes, Vine, & Marra, 2020, p. 1). For learners in a non-target language environment, however, language skills are often acquired through formal instruction, and the process might not be as smooth. Studies of the use of pragmatic expressions by second or foreign language learners have shown that learners often reach a stage where they have “a near-to-native grammatical and lexical competence but with a limited range of pragmatic resources” (Romero-Trillo, 2018, p. 113).

Norwegian learners of English are in general perceived to have a high level of grammatical competence and tend to do well on national and international tests measuring elements of grammatical competence, such as Assessment of English (2004) and EF English Proficiency Index (2019). However, little is known about their pragmatic competence in English. From the Norwegian popular press, one could get the impression that the level of pragmatic competence in English of Norwegians in general is rather low, as the press regularly publishes articles on how Norwegians are perceived as impolite when they speak English, and how Norwegians lose international business contracts because of their lack of English skills. Moreover, previous studies of hedging by Norwegian learners, such as Hasselgreen (2004) and Thomson (2016), have shown somewhat conflicting results concerning the degree to which Norwegian learners master the art of hedging. Furthermore, Hasselgreen (2004) and Thomson (2016) focus on lower secondary school learners, but limited research has been done on advanced learners of English, particularly spoken English of advanced learners. As the oral skills of learners are becoming

---

more and more important in an increasingly global society (Friginal, Lee, Polat, & Roberson, 2017, p. 3), the need to study spoken learner language becomes even more urgent.

This thesis aims to address all of these needs. By exploring a methodological approach that circumvents the need to use pre-decided forms as a point of departure for a corpus study of hedging strategies, the thesis hopes to add to the empirical base of English and Norwegian hedging research and contribute to a better understanding of how hedging is expressed in Norwegian. Furthermore, the results of the contrastive study will form the basis for an investigation into the use of hedging strategies in spoken English by Norwegian advanced learners.

1.4 Overall research design

To achieve the overall aim, this thesis has been organised as a research project with three phases, each with its own objective, corresponding research question(s) and outputs. Figure 1-1 presents an overview of the three phases. The studies in the three phases are mainly qualitative in nature, and their contributions range from methodological to empirical.

![Figure 1-1 Overview of the project phases.](image)

The objective of phase 1 is to develop a methodological approach, i.e. a “workaround” (O'Keeffe et al., 2020, p. 9), to access the pragmatic function of hedging in Norwegian and English corpora of spoken conversations. Article 1 is the output of phase 1, and it aims to strengthen the body of research on
methodological approaches from the function-to-form perspective. More precisely, article 1 explores the use of probes (defined in 5.1.1) to access speech situations in corpora in which hedging strategies are likely to occur.

The objective of phase 2 is to discover similarities and differences in the realisation of hedging strategies in Norwegian and English spoken conversations. The empirical contributions of the studies conducted in phase 2 aim to increase our understanding of how hedging is expressed in Norwegian and English informal conversations. The studies apply the methodological approach tested in phase 1 on data from corpora of spoken informal Norwegian and English conversations (described further in chapter 4), and analyse the results from two different perspectives. Whereas article 2 focuses on the types of hedging strategies used, article 3 compares the actual linguistic forms used to express hedging. Specifically, article 2 studies the differences and similarities in the types of hedging strategies used in Norwegian and English, i.e. whether the hedging takes place within the proposition, on the relationship between the speaker and the proposition, or on the relationship between the speaker and the hearer. The study reported in article 2 also proposes a framework for a classification of hedging strategies in spoken conversations in order to ensure a sound basis for the cross-linguistic comparison. Article 3 compares the forms, i.e. particles, adverbs, tag questions, etc., used to express hedging in Norwegian and English.

In phase 3, the use of hedging strategies by Norwegian advanced learners of English is compared to that of native speakers of English. Previous studies of hedging in learner language have shown that overall, learners tend to hedge less frequently than their native speaker counterparts and that they restrict their hedging repertoire to a limited number of strategies (e.g. Hyland and Milton, 1997 and Friginal et al., 2017). The objective of phase 3 is to discover whether there are any differences between Norwegian learners and native speakers of English in the use of a selection of hedging strategies. Building on the results of article 3, the 10 proportionally most frequent forms from the English material are used as a point of departure. These forms are studied in comparable corpora of spoken English by Norwegian learners and native speakers of English (described further in chapter 4), and the results are reported in article 4. As opposed to many previous studies of over- and underuse of different expressions by learners, the study in article 4 looks at differences both at group level and individual level. By looking at both inter- and intraspeaker differences, the degree of over- and underuse becomes more nuanced. The contribution of article 4 is mainly empirical and adds to the research base on learner language in general and English by Norwegian learners in particular.
The four articles are linked together by an extended introduction ("kappe") which outlines the theoretical background and previous research within the fields of hedging and corpus pragmatics in chapter 2 and describes how the present thesis builds on this foundation and where it seeks to add new knowledge in chapter 3. It presents the corpora used in the studies in chapter 4, and reflects on some of the methodological considerations done throughout the project in chapter 5. Summaries of the articles are provided in chapter 6, and a summary of the results, limitations and suggestions for further research is given in chapter 7.

This thesis comprises the following papers, which are referred to in the text by their Arabic numerals.


### 1.5 Thesis outline

Chapter 1: Introduction
Chapter 2: Previous research
Chapter 3: Overview of the phases
Chapter 4: Presentation of the data
Chapter 5: Methodological considerations
Chapter 6: Summary of articles
Chapter 7: Summary of the results and conclusions
Appendices:

1. A working guide for analysis (English)
2. A working guide for analysis (Norwegian)

Articles 1–4
Chapter 2 Previous research

This chapter presents the theoretical backdrop for the thesis, cutting across two subfields of linguistics: hedging research and corpus pragmatics. The aim is to present an overview of the state-of-the-art of the research within both of these subfields with relevance to the studies presented in articles 1–4. The chapter also discusses how hedging research and corpus pragmatics overlap with other areas of pragmatic research; however, these related fields are described only to the extent that they contribute to shine a light on hedging and corpus pragmatics. The chapter consists of two main parts. The first part outlines some important points in the development of the concept of hedging, with an emphasis on work done on English and Norwegian, given the focus of the study. In section 2.1, various definitions of hedging are presented. Section 2.1.1 outlines previous research on hedging and discusses how the understanding of the concept has developed. Section 2.1.2 presents some overlapping fields and how they have contributed to the current understanding of hedging. Section 2.1.3 summarises the discussion.

The second part of the chapter outlines some important aspects of corpus pragmatics. In section 2.2, corpus pragmatics is defined. Section 2.2.1 outlines some of the central elements of corpus linguistics which have influenced corpus pragmatics. Learner corpus research is described in section 2.2.2. Section 2.2.3 outlines some central elements of pragmatics which have influenced corpus pragmatics. Four overlapping subfields of pragmatics, contrastive pragmatics, cross-cultural pragmatics, intercultural pragmatics and interlanguage pragmatics, are described in section 2.2.4. Section 2.2.5 summarises the main characteristics of corpus linguistics and pragmatics and discusses how they have influenced corpus pragmatics. Section 2.2.6 gives a brief summary of the characteristics of learner corpus research and interlanguage pragmatics relevant for corpus pragmatics.

2.1 Hedging

Hedging may be defined as a type of discourse strategy that reduces the force or truth of an utterance (Kaltenböck et al., 2010, p. 1). The term discourse strategy may be understood as any means or method of bringing about a result (Sanders, 2015), thus the definition opens up for a variety of realisations, i.e. words and phrases, syntactic configurations such as agentless passive constructions, tag questions, if-clauses, etc., as well as paralinguistic elements of communication such as prosody, gestures, facial
expressions. For practical reasons, I limit my investigation to linguistic elements such as words, phrases, clauses, and various syntactic constructions used to perform the pragmatic function of hedging.

Most of the current studies of hedging view hedging as a strategy reducing the force or truth of an utterance, as proposed by Kaltenböck et al. (2010). However, although researchers seem to agree that the concept of hedging includes components of non-prototypicality, uncertainty, doubt, and caution towards the proposition and the audience (Holmes, 1982, p. 18), the exact wordings of existing definitions vary. The definitions typically refer to hedging as either a (type of) strategy, a behaviour, or the linguistic devices used to express hedging. Table 2.1 presents some examples.

Table 2.1 Variations in the definitions of hedging.

<table>
<thead>
<tr>
<th>Hedging is …</th>
<th>… of tentativeness and possibility</th>
<th>Hyland (1996, p. 433), Knight, Adolphs, and Carter (2013, p. 135)</th>
</tr>
</thead>
<tbody>
<tr>
<td>… one or more lexico-syntactic elements or a strategy that modify a proposition in a certain context for specific communicative purposes</td>
<td></td>
<td>Schröder and Zimmer (1997, p. 249)</td>
</tr>
<tr>
<td>… an interactional strategy that speakers and writers use in communication in a variety of ways and for a variety of purposes</td>
<td></td>
<td>Farr and O'Keefe (2002, p. 25), Gries and David (2007, p. 1)</td>
</tr>
<tr>
<td>… a realisation of an interactional/communicative strategy that mitigates the harshness or hostility, or softens the force of utterances and makes them more acceptable to the hearer</td>
<td></td>
<td>Tchizmarova (2005, p. 1146)</td>
</tr>
<tr>
<td>… a rhetorical strategy which attenuates either the full semantic value of a particular expression or the force of a speech act</td>
<td></td>
<td>Fraser (2010, p. 22)</td>
</tr>
<tr>
<td>… a discourse strategy that reduces the truth or force of the whole or parts of an utterance</td>
<td></td>
<td>Kaltenböck et al. (2010, p. 1)</td>
</tr>
<tr>
<td>… a linguistic strategy / linguistic marker that allow the speaker to weaken the force of a proposition</td>
<td></td>
<td>Kranich (2011, p. 80)</td>
</tr>
<tr>
<td>… a behaviour wherein speakers or writers attempt to distance themselves from the proposition they are communicating</td>
<td></td>
<td>Prokofieva and Hirschberg (2014, p. 1)</td>
</tr>
</tbody>
</table>

The variation illustrated in table 2.1 suggests that hedging is a complex and multi-faceted phenomenon, which could potentially “pose a number of tightly interwoven definitional, terminological, and typological difficulties” (Adamczyk, 2015, p. 321). This variation in definitions is a result of the broadening of the concept of hedging since it became a topic of interest in the late 1960s, the overlap with related fields such as mitigation, vagueness, modality and pragmatic markers, and the heterogeneity of research perspectives from which hedging has been studied. All of which will be discussed in more detail below.
2.1.1 Previous research on hedging

Although research on hedging is a fairly recent field within linguistics, it comprises nearly 50 years of research and the number of publications on the subject has warranted several review articles, e.g. Clemen (1997), Schröder and Zimmer (1997) and Fraser (2010). The term hedge was introduced by G. Lakoff (1972) in the early 1970s as a term for a class of words implicitly indicating fuzziness, i.e. words which made things more or less vague, but the concept of hedging was already discussed in the mid-1960s in the work of Lotfi Aliasker Zadeh (1965) and Uriel Weinreich (1966).

The initial research on hedging was rooted in logic, formal semantics, and the truth-conditional meaning of sentences. Relevant to the background of hedging research is set theory, which is central to formal semantics. Sets can be understood as an abstract collection of objects. Zadeh (1965) introduced the idea of fuzzy sets as “a class of objects with a continuum of grades of membership” (1965, pp. 338-339), thus fuzzy set theory can be understood as an extension of the classical notion of sets. In set theory, elements either belong to a set or they do not, whereas fuzzy set theory opens for degrees of membership of a set. The idea behind fuzzy sets was that fuzziness plays an important role in human cognition and that many of the classes of objects in the real world do not have clear membership criteria. For example, the class of animals would include objects such as dogs and horses and exclude objects such as rocks and plants, but there are other objects for which it seems more difficult to determine whether they belong to this class or not, e.g. bacteria (Zadeh, 1965, p. 338).

Semantic categories were also studied by Rosch (1973) who argued that properties defining a category are not necessarily shared by all speakers equally and that category membership is a matter of degree. An object can be a central, i.e. a prototypical, member of a category or a more peripheral member.

The idea of real-life concepts having fuzzy boundaries influenced the work of Lakoff (1972), who challenged the traditional distinction between true, false or non-sensical declarative sentences characteristic of formal semantics. Lakoff (1972, p. 458) argued that real-life concepts have fuzzy boundaries, which cause natural sentences to often neither be true, false or non-sensical, but rather that there are degrees of truth. The classical example proposed by Lakoff is that of the bird: the category of birds has vague or fuzzy boundaries or membership criteria. At least to a non-ornithologist, a creature can be a bird to a greater or lesser degree so whether or not a creature is a bird is not a yes/no question. Their properties are measured against the ideal of a bird, i.e. a robin is more of a bird than a duck or a penguin is (Lakoff, 1972, p. 459). Lakoff (1972) expressed a particular interest in words whose meaning “implicitly involves fuzzyness” (Lakoff, 1972, p. 471). The effects of such words, i.e. hedges,
could not be described in a two-value system where sentences are either true or false. The effect of hedges was that of altering the truth value of the sentence and signalling degree of category membership, for example, *A robin is sort of a bird* is false because a robin is a bird, whereas *A penguin is sort of a bird* is true or close to true because a penguin is less of a bird than a robin (Lakoff, 1972, p. 471). Thus in Lakoff’s view, hedges either have an attenuating effect on the meaning of an expression, e.g. *sort of, pretty much*, or reinforce its characteristics, e.g. *extremely, par excellence*.

The development of the concept of hedging is also related to the work of Weinreich (1966) on the semantic structure of language. According to Weinreich (1966, p. 4), language is a repertoire of signs, which can be divided into designators and formators. A designator consists of a sign and a designatum, i.e. the object referred to by a sign, whereas a formator consists of a sign and an implicit instruction for an operation, e.g. negation. In his view, hedges are closely linked to metalanguage. Metalinguistic operators, as Weinreich labels them, function as “instructions for the loose or strict interpretation of designata” (Weinreich, 1966, p. 130).

Although Lakoff (1972) was mostly concerned with expressions that modify the category boundaries of a concept (Apróné, 2011), he also mentions that the meaning of hedges may be context dependent and that hedges may be used to soften requests or modify claims (Lakoff, 1972, p. 484). However, he does not pursue this kind of hedging. Still, gradually the focus of hedging research shifted from logico-semantic aspects to pragmatic aspects, i.e. from *langue* (semantics) to *parole* (pragmatics) (Taavitsainen & Jucker, 2015, p. 5), and how hedges affect speech acts became a topic of great interest. In contrast to hedging as a measure signalling degrees of category membership, speech act hedging has a different purpose. Fraser (1975) introduced the term *hedged performatives* for certain performative verbs preceded by a modal verb causing an attenuated illocutionary force of the speech act expressed, e.g. *I must advise you to leave* (Fraser, 1975, p. 187). He distinguished between strong and weak performatives, the former being acts denoted by the performative verb and the latter cases where the performative use is dubious. The shift in focus from propositional hedging to speech act hedging anticipated the binary distinction which is still found in most descriptions of hedging today. The reinforcement aspect of hedging, originally included by Lakoff (1972) was more or less abandoned (Fraser, 2010, p. 22). However, it is still found in studies of modality and stance. Forms used to reinforce utterances are often referred to as *boosters* (see further section 2.1.2).

This binary distinction between propositional hedging and speech act hedging is also found in Norwegian hedging research. The beginning of Norwegian hedging research is often associated with
Thorstein Fretheim’s (1979) paper on *dempere* (‘attenuators’, ‘hedges’) from 1979. Fretheim discusses the pragmatic functions of the Norwegian word *nesten* (‘almost’) and points out that it has two main uses; one where it affects the truth value of the proposition and one where it modifies the speech act. When *nesten* is used to modify a speech act, it is referred to as a demper. Fretheim’s work was inspired by the work of Lakoff on hedges but also by Quirk, Greenbaum, Leech, and Svartvik’s (1972) concept of downtoners. Quirk et al. (1972: 439ff) explain downtoners as adverbs that decrease the force of an utterance.

The speech act aspect of hedging was pursued by Brown and Levinson (1987) in their work on politeness. Whereas Lakoff (1972) only studied propositional hedging and Fraser (1975) only considered hedging with certain performative verbs, Brown and Levinson (1987) investigated hedging on the illocutionary force in greater detail. A central concept to their theory of politeness was the concept of *face* and how speakers threaten their own or the interlocutor’s face when performing speech acts. The anthropological concept of face was originally introduced by the Chinese anthropologist Hu in the 1940s; however, the expression *to lose face* has existed in the English language since the mid-1800s. The expression is a translation of the Chinese *tu lien* and can be paraphrased as *to lose prestige* (Harper, 2000). The concept of face was further explored by Goffman (1955) and has become a central term in sociology and other scientific fields. Goffman (1955, p. 213) defines face as “the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact,” thus face is not something that exists in or on a person but something which is negotiated when interacting. In this respect, a person is also responsible for the face of others and it would be seen as “heartless” to see another person lose his/her face and “shameless” to not care about the loss of one’s own face (Goffman, 1955, p. 215). It is, therefore, a temporary working acceptance between the participants of spoken interaction to uphold each other’s face (Goffman, 1955, p. 216).

The notion of face as it is described in Brown and Levinson (1987, p. 61) derives from Goffman’s work. Face is explained as something “that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction” (Brown & Levinson, 1987, p. 61). Brown and Levinson (1987, p. 61) distinguish between positive and negative aspects of face. Positive face is the positive consistent self-image of a person, and negative face is every person’s right to freedom of action and not to be imposed on. In a conversation, a speaker will be concerned with both, i.e. to save his own face and to protect the face of the person he/she is speaking to (Goffman, 1955, p. 217). Since it is every speaker’s wish that his/her face is maintained or enhanced, Brown and Levinson (1987, p. 62) reformulated the positive and negative aspects of face as basic wants, in which
negative face is regarded as the want of a speaker that his/her actions are unimpeded by others, and positive face is regarded as the want of a speaker that his/her wants are desirable to at least some others. Acts of communication, whether verbal or non-verbal, that run contrary to these wants, threaten the speaker or hearer’s face (Brown & Levinson, 1987, p. 65).

Nevertheless, a speaker may choose to perform a speech act despite the risk of threatening an interlocutor’s face. If the speaker chooses to perform a face-threatening act, there are strategies which can be employed to minimise the threat, i.e. redressive action (Brown & Levinson, 1987, p. 68). Such redressive action could be oriented towards either aspect of face or both simultaneously. The use of hedging strategies is regarded as one type of redressive action oriented towards negative face (Brown & Levinson, 1987, p. 145). Brown and Levinson (1987, p. 145) define hedges as “a particle, word, or phrase that modifies the degree of membership of a predicate or noun phrase in a set; it says of that membership that it is partial, or true only in certain respects, or that it is more true and complete than perhaps might be expected” (emphasis in original), but state that there are the performative hedges that are the most important when it comes to satisfying the speaker’s want (Brown & Levinson, 1987, p. 146). The reinforcement aspect of hedges mentioned is not investigated in detail. One of the presuppositions of Brown and Levinson is that “ordinary communicative intentions are often potential threats to cooperative interaction” (Brown & Levinson, 1987, p. 145) and that these intentions are often encoded in speech acts. For example by promising to do something, the speaker simultaneously admits to not having done it already. To hedge these assumptions is to avoid commitment to them and thus disarming potential threats. Unlike Lakoff, Brown and Levinson (1987, p. 146) argue that hedging can be achieved through an indefinite number of surface forms.

Although highly influential, Brown and Levinson’s theory has been subjected to criticism. The developments within politeness research will not be addressed in detail here, but a few of the criticisms will be mentioned. One of the major criticisms is that Brown and Levinson’s theory is too Anglo-centric (Beeching, 2016, p. 17; Yeung, 1997, p. 510), particularly in relation to how the concept of face is understood and described. Researchers have argued that Brown and Levinson disregard the interpersonal and social aspects of face and that in many cultures, face is more of a collective concern than an individual concern, thus their understanding of face is too narrow (Spencer-Oatey, 2002). Furthermore, the theory has been criticised for being too focused on the needs of the speaker and less on the hearer (Eelen, 2001). With respect to hedging, the broadening of the concept has led to it no longer being seen as involving merely redressive action towards the negative aspect of face, but also strategies towards the positive aspect of face.
Brown and Levinson’s (1987) account of hedging was not the only mark of the transition in hedging research from semantics into pragmatics. Grice’s (1975) theory of communication also played an important role in this shift. Grice distinguished between what is said, i.e. the truth-conditional aspects of meaning, and what is meant, i.e. the non-truth-conditional aspects of meaning. A central point to this theory is that there is often a difference between what is said and what is meant or implied in natural language. Grice observed that some inference on behalf of the hearer often was needed to make sense of what was meant by an utterance, the explanation of which stretched beyond truth-conditional theory. For example, by saying, *I sleep all the time, Doctor*, a speaker is usually understood to mean, *I sleep more than I think is normal* (Archer, Aijmer, & Wichmann, 2012, p. 47). Thus there is a difference between sentence meaning i.e. natural meaning and speaker meaning i.e. non-natural meaning (Jaszczolt, 2012, p. 5). Grice coined the term *implicature* (1975 p. 44) for implied meaning generated intentionally by the speaker and potentially inferred by the hearer. An implicature is created when speakers on the surface do not seem to adhere to the cooperative principle, “make your conversational contribution as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.” In addition to the cooperative principle, Grice (1975) proposed four maxims regulating speakers’ contributions, i.e. the maxims of quantity, quality, relation, and manner (1975, p. 45). It is because of the assumption that speakers do adhere to the cooperative principle and the maxims even if the surface forms of their utterances do not look that way that we work out the implicatures, i.e. hearers will construe the speaker’s meaning of an utterance so that it conforms to the maxims. However, a speaker may choose not to adhere to these maxims. This may be done in three different ways, i.e. opting out, flouting, or violating a maxim. When flouting a maxim, the speaker blatantly fails to observe it in order to draw the hearer’s attention to the implicature being created. When a speaker opts out of a maxim, the speaker indicates that he/she is unwilling to cooperate, for example due to legal or ethical reasons. When a speaker violates a maxim, the speaker deliberately misleads the hearer. From this perspective, hedging is seen as a way of tactfully opting out of a maxim. For example, in the constructed sentence *As far as I know, there are three cookies left in the jar.*. *As far as I know* illustrates how a speaker opts out of the maxim of quantity.

The concept of hedging evolved as parts of these important theories of communication, and by the end of the 1970s, it was firmly established within the field of pragmatics (Kaltenböck et al., 2010, p. 5). The growing influence of pragmatic research on hedging also resulted in a broadening of the concept. In the 1980s, a considerable effort was devoted to attempting to sub-classify hedges into different types. Many of these attempts were based on empirical studies of hedging in various types of discourse. In
recent years, many of these classification systems have been met with criticism (Kaltenböck et al., 2010, p. 6); however, they are still referred to and applied in studies of hedging today. Kaltenböck et al. (2010, p. 7) suggest that the reason for upholding criticised classificatory frameworks is that they are “necessary heuristic devices,” i.e. they contribute to our understanding of hedging strategies. Nevertheless, there is a need to develop new classificatory frameworks or revise existing frameworks in order to account for hedging strategies in all genres of spoken and written discourse. (See also article 2.)

The seminal paper by Prince, Frader, and Bosk (1982) is one example of a study which outlines a classification framework for hedging strategies. Prince et al. (1982) compiled a corpus of recorded and transcribed morning rounds at a hospital ward in the period 1978–1979. Based on the definition of hedging by Lakoff (1972), they identified between 150 and 450 hedges per hour of speech. Prince et al. (1982, p. 4) observed that hedges made things fuzzy in at least two distinct ways, i.e. within the propositional content and between the speaker and the proposition. Hedges that affect the truth value of propositional content are referred to as approximators, whereas hedges that affect the commitment of the speaker to the truth of the proposition are referred to as shields (Prince et al., 1982, p. 5). Referring to the Gricean division between truth-conditional semantics and non-truth-conditional pragmatics, Prince et al. (1982, p. 6) argue that approximators and shields have little to do with each other. Approximators say something about the semantics and affect class membership, whereas shields affect the pragmatics by saying something about speaker commitment.

In Prince et al.’s framework, approximators are further sub-divided into adaptors and rounders. Adaptors are related to class-membership and the lexicon of words speakers have at their disposal. When for example describing symptoms, as would frequently be the case in Prince et al.’s (1982) corpus, there are certain prototypical terms to choose from. If, however, the patient’s symptoms are not prototypical, the speaker may have to choose a non-prototypical term to describe the symptoms. The speaker may use an irrelevant term (i.e. a term which exists in the lexicon but has no relevance to the description of symptoms), coin a new term, choose the closest relevant term, or choose the closest relevant term and indicate by hedging that this is not the most accurate one (Prince et al., 1982, p. 9). Hedges marking adaption in this way are called adaptors. The second category, rounders, comprises hedges indicating that the precise term is not relevant. Fraser (2010, p. 19) illustrates this with the use of something in the example The baby’s blood pressure was something between 40 and 50. Both types of approximators indicate that “the actual situation is close to but not identical to the prototypical situation” (Prince et al., 1982, p. 11).
Shields, on the other hand, do not affect the truth conditions of the proposition, but rather indicate that the speaker is not fully committed to it (Prince et al., 1982, p. 11). The class of shields is also further divided into two sub-classes: plausibility shields and attribution shields. Plausibility shields are hedges which enable the speaker to express various degrees of doubt regarding the belief expressed, whereas attribution shields attribute the belief to someone else (Prince et al., 1982, p. 11). In the utterance *I think his feet were blue*, the construction *I think* is an example of a plausibility shield indicating doubt on behalf of the speaker. In the utterance *according to her estimates, she got the baby’s heart rate back within two minutes*, the -ing -clause *according to her estimates* is an example of an attribution shield.

Although the classificatory framework of Prince et al. (1982) is based on physician–physician discourse, it has had a general appeal and is still used in its original form, e.g. Prokofieva and Hirschberg (2014), or serve as a basis for other classificatory systems in studies of other types of discourse e.g. Salager-Meyer (1994). Its appeal may be a result of the fact that Prince et al.’s (1982) study is one of the few that base a generic frameworks for hedging strategies on authentic spoken data. The majority of other frameworks are based on written texts. Despite its appeal, the system has also received criticism. Skelton (1988) argues that the subdivisions found in Prince et al. (1982) only works on an abstract level and not in practice.

Another important contribution to the classification of hedging strategies was proposed by Hübler (1983), who wrote the first monograph on hedges. The point of departure of the study was the idea that assertive sentences represent a hypothesis about a state of affairs in the world, which can be negated (Hübler, 1983, p. 155). The speakers’ aim is to avoid rejection of their proposition by the hearer, and therefore he/she will make use of a variety of linguistic devices to reduce the risk of negation. The use of such devices “lessens the liability for acceptability by reducing the number of acceptability conditions” (Hübler, 1983, p. 19). Hübler (1983) suggests a division between *understatements* and *hedges*. Understatements concern the phrastic, i.e. the propositional content and hedges concern the neustic, i.e. the relationship between the speaker and the proposition (Hübler, 1983, p. 20). Hübler (1983, p. 1) uses the term *understatements* as a metalinguistic term “collecting certain verbal expressions into one class and is thus concerned with a linguistic pattern of behaviour.” Although he distinguishes between hedges and understatements, he also uses understatements as an overarching term for both. Hübler (1983, p. 23) underlines that understatements and hedges are not stylistic strategies but strategies pertaining to content since they operate within and on the sentence proposition. They aim to make sentences more acceptable and “to increase their chance of ratification by the hearer” (Hübler, 1983, p. 23). Thus Hübler (1983) mentions the hearer in his account of hedging, but maintains
the binary distinction between propositional hedging and hedging between the speaker and the proposition which is found in Prince et al. (1982) and most other existing classificatory frameworks.

Attempts to classify hedging strategies in various text types continued well into the 1990s and several of the resulting taxonomies are still frequently used. Salager-Meyer (1994) studied hedging strategies in medical written discourse, research papers and case reports, and built a classificatory framework based on that of Prince et al. (1982). Salager-Meyer (1994) observed that previous research on hedging in scientific discourse states that hedges are often used to signal distance to the claims being made and that this is done with the purpose of protecting one’s own reputation as a scientist, avoiding absolute statements that may damage the researcher or the researcher’s institution and indicating that one is open to more interpretations (1994, p. 150). She presents an additional view of hedges, in which hedges can be seen as ways of reporting results more precisely. By representing the right state of knowledge, “greater preciseness in scientific claims” is achieved (Salager-Meyer, 1994, p. 151). Salager-Meyer (1994) understands hedging in scientific writing as a three-dimensional concept, and regards it as a threat-minimising strategy, a reflection of the author’s modesty concerning their achievements, and the impossibility or unwillingness to accurately quantify the phenomena studied (1994, p. 153). Although she uses more or less the same terminology as Prince et al. (1982), she uses it in a slightly different way. The category of shields corresponds to Prince et al.’s probability shields. Her approximators comprise both adaptors and rounders in the same way as in Prince et al.’s taxonomy, but Salager-Meyer does not sub-divide hedges into the two sub-categories. The third category, author’s personal doubt and direct involvement, is not found in Prince et al. (1982). This category includes expressions such as I believe and to our knowledge, which would be included in the shields categories in Prince et al. (1982). The fourth category, emotionally charged intensifiers, includes expressions which project the author’s reaction, e.g. surprisingly. The final category, compound hedges, includes instances of multiple hedges occurring together (Salager-Meyer, 1994, p. 154).

Another classificatory framework based on written scientific texts is that by Hyland (1996, 1998). Hyland developed a framework for classifying hedging strategies based on a 75,000-word corpus of 26 articles from journals in the field of cell and molecular biology (Hyland, 1996, p. 434) and distinguishes between content-oriented and reader-oriented hedges. Content-oriented hedges concern the relationship between the proposition and a representation of reality, i.e. what a writer says about the world and what the world is actually like (Hyland, 1996, pp. 436, 439), and reader-oriented hedges concern the interactional aspect of a statement and function as an invitation for the reader to take part in a dialogue (Hyland, 1996, p. 446). The class of content-oriented hedges is further divided into
accuracy-oriented and writer-oriented hedges. Accuracy-oriented hedges enable the writer to make uncertain claims with the appropriate caution (Hyland, 1996, p. 440). The accuracy-oriented class is subdivided into attribute and reliability hedges. Attribute hedges indicate a discrepancy between actual results and what was expected or between the results and the terms used to explain them (Hyland, 1996, p. 441). Reliability hedges say something about the writer’s confidence in the truth of the proposition (Hyland, 1996, p. 441). Writer-oriented hedges are writer-focused and are often used to diminish the author’s presence in the text to shield him/her from potential opposition (Hyland, 1996, p. 443). Although frequently applied, Hyland’s framework has also received criticism, e.g. for not making clear distinctions between the sub-categories (Diewald, 2006, pp. 303-305) and for “including interpersonal communicative effects,” which are hard to measure (Kaltenböck et al., 2010, p. 5).

What characterises most of the classificatory frameworks of hedging strategies is the binary distinction between propositional hedging and hedging on the relationship between the speaker and the proposition. A third perspective, the relationship between the speaker and the hearer is mentioned in both Hübler (1983) and in Holmes’ (1982) account of epistemic modality (see section 2.1.2). However, it is only fully integrated into Hyland’s framework (1996, 1998). A challenge with many of the proposed classificatory frameworks is that they are based on different definitions of hedging. Furthermore, the classificatory frameworks are typically based on only one type of discourse in only one language, mainly English. However, the way the function of hedging is realised often varies between languages (Holmes, 1982, p. 10). Therefore, there is still a need for a classificatory framework which can be applied to studies of hedging across languages and genres.

Norwegian hedging research has not focused on classifying types of hedging strategies, but some references to different types of hedging strategies can be found in the literature. Fretheim (1981) distinguishes between ego- and alter-hedges, the former expressing uncertainty on the part of the speaker towards the content of the proposition and the latter expressing uncertainty on the part of the speaker as well as seeking confirmation from the addressee (Fretheim, 1981, p. 87). Although Fretheim only describes the properties of two particular words, the modal particles vel (‘well’, ‘right’) and nok (‘probably’), the perspectives introduced are highly relevant. Fretheim argues that vel appeals to the hearer, and that the speaker, through the use of vel, seeks confirmation from the hearer. Alter-hedges such as vel share some of the interactional properties of reader-oriented hedging strategies in Hyland’s framework, described above. In contrast, nok does not signal the same appeal, but merely signals uncertainty on the part of the speaker.
In addition to establishing classificatory frameworks for hedging strategies, another trend in hedging research from the 1980s and onwards is the study of properties of individual hedges, e.g. Aijmer (1984), Holmes (1986), Farr and O’Keeffe (2002), and Fetzer (2010). Such studies often look at the distribution and different functions of particular expressions in one or more types of discourse. This trend overlaps to a large extent with research on pragmatic markers. (See further Section 2.1.2). Studying hedging as one of several functions of pragmatic markers has been characteristic of Norwegian hedging research as well. The research on hedging in Norway has to a large degree become intertwined with research on pragmatic markers and modal particles in particular (see 2.1.2 for definitions), e.g. *jo* (‘after all’, ‘you know’, ‘of course’) (Andvik, 1992; Berthelin & Borthen, 2019; Lind, 1994; Solberg, 1990) and *bare* (‘just’) (Hasund, Opsahl, & Svennevig, 2012; Opsahl, 2002; Opsahl & Svennevig, 2007). These studies are typically form-to-function (see section 2.2.5 for a definition and examples), taking a particular linguistic expression as the point of departure, and investigating its various functions are investigated in different contexts. Borthen (2018, pp. 232-233) presents an extensive overview of research on Norwegian particles from the late 1970s to the present. Many of the studies are rooted in a cognitive-pragmatic tradition and use relevance-theoretical frameworks to account for the various uses. The purpose of many of these studies has been to establish a proper semantic description of the expression in question and show how they are used pragmatically (Unger, 2018).

From the end of the 1980s, two further trends of hedging studies have emerged: genre-specific studies and applied studies. Within the category of genre-specific studies, hedging in scientific discourse developed as a large subfield of its own. The already mentioned studies by Salager-Meyer (1994) and Hyland (1996, 1998) are examples of research on hedging in scientific discourse. Other examples are Varttala (1998, 2001), who studied hedging in scientific discourse, and Vass (2004, 2017), who studied hedging in legal discourse. Vold (2006) studied certain hedging expressions in Norwegian scientific discourse. Applied studies comprise a variety of perspectives on hedging, e.g. contrastive studies, translation studies and intervarietal studies (Markkanen & Schröder, 1997, p. 12). Vold’s (2006) study is also an example of a contrastive study, as she compares hedging strategies across Norwegian, English and, French scientific texts. Another contrastive study is Kärkkäinen (1990) comparing hedging in spoken Finnish and English. An example of a translation study is Kranich (2011), investigating hedging in English texts, English-German translations and German original texts. An example of an intervarietal

---

4 Relevance theory is a cognitive-pragmatic framework which sprung out of Grice’s maxim of relevance. A premise for the theory is that human beings are relevance-oriented and that they seek to find the most relevant interpretation of an utterance. See Sperber and Wilson (1986, 1995) for a full account of relevance theory.
study is Uysal (2014), comparing hedging strategies in conference proposals by non-native and native speakers of English.

2.1.2 Hedging in related areas of research

Hedging research does not exist in a vacuum and has been studied in connection with other areas of research, particularly pragmatic markers, modality, vague language and mitigation. Hedging in the light of each of these areas will be discussed below. First, however, the reinforcement aspect of hedging will be addressed in a bit more detail.

Boosters

Boosters have developed from the reinforcement aspect of hedging introduced by Lakoff (1972) and were originally seen as a part of hedging, i.e. hedging strategies made concepts more or less fuzzy. Today, the term hedge is used for attenuating and downtoning expressions, whereas booster is used for expressions indicating speaker confidence, certainty and precision. Although hedging and boosting seem to be clearly distinguishable in theory, hedging and boosting expressions can be difficult to distinguish in practice. One of the reasons for this problem is that the same type of expressions may be used to express both hedging and boosting. For example, the use of vague language may from a hedging perspective be interpreted as a way of indicating shared knowledge between the speaker and the hearer. From a boosting perspective, it may be interpreted as indicating precision. Jucker, Smith, and Lüdge (2003, pp. 1751-1752) exemplify this by referring to an example originally introduced by Moxey and Sanford (1997, p. 211). They argue that it would be more informative and precise to report that Most of our students passed an advanced exam. than Twenty-two of our students passed an advanced exam. when the hearer does not know how many students were eligible to take it. From a hedging perspective, this example could be interpreted as the speaker using an approximate estimate because he/she does not know the actual number or does not think it is important (cf. the concept of rounders discussed in 2.1.1).

Pragmatic markers

The term pragmatic marker represents a broad and fuzzy concept, and this fuzziness is clearly reflected in the use of terminology in the extensive body of research on the topic. Jucker and Ziv (1998, p. 2) argue that the multiplicity in the terminology not only reflects different elements in different languages with comparable functions, but also distinct theoretical perspectives from which these elements have been studied. This view is also supported by Borthen (2018, p. 232). In this thesis, the term pragmatic
marker is used as in Borthen (2018), Andersen and Fretheim (2000) and Fraser (1996, 1999), and is understood as a linguistic sign (lexeme, inflectional suffix, word order, intonation pattern, etc.) which communicates something beyond the proposition. The term marker is preferred because it encompasses a broad range of linguistic signs, e.g. single word items, phrases, etc. (Brinton, 1996, pp. 29-30). The term pragmatic is preferred here over discourse, which is also frequently used about such elements, because it is a broad term covering the range of functions such elements may fill (Brinton, 1996, p. 30). This is in contrast to Jucker and Ziv (1998), who argue that discourse marker is the broader term. In this thesis, however, discourse marker is understood as a type of pragmatic marker having a primarily discourse-structuring function, in line with Aijmer and Simon-Vandenbergen (2006, p. 2), who state that “discourse marker is the term which we use when we want to describe how a particular marker signals coherence relations,” and Beeching (2016, p. 132), who states the discourse markers bracket units of talk. Thus, pragmatic markers may be defined as “signals in the communication situation guiding the addressee’s interpretation” (Aijmer & Simon-Vandenbergen, 2006, p. 2), or “elements of language […] that weave the net of discourse between the addresser, the addressee, and the context of a given message” (Romero-Trillo, 2012, p. 4222).

Another sub-type of pragmatic markers is pragmatic particles. Pragmatic particles may be used to refer to lexical pragmatic markers which say something about the interlocutors’ commitment or attitude towards the content of the proposition or the relationship between the content of the proposition and the interlocutors’ presuppositions about the world (Borthen, 2018, p. 227). Here, pragmatic particles will mainly be used to refer to “the uninflectable semantically and phonologically reduced lexical items, which occur postverbally, cannot take stress, and have no inherent lexical-semantic meaning, but rather function as expressions of emotive modality, similar to intonation and gestures in English” (Andvik, 1992, p. vi) characteristic of Scandinavian and Germanic languages. In a study of Norwegian pragmatic particles, Lind (1994) introduce a subtype of pragmatic particles called modal particles. In an effort to distinguish modal particles as a separate sub-category, Lind (1994, p. 21) states that modal particles primarily say something about the speaker’s attitude towards the message and that they simultaneously are hearer-oriented and guide the hearer in the interpretation of the message. However, it is difficult to make a clear distinction between pragmatic particles and the subcategory of modal particles. Therefore, I have chosen to use the label pragmatic particles and avoid further subdivisions.

Pragmatic markers occur frequently in spoken conversations and may serve a variety of functions. Traditionally, researchers have focused on their text-structuring function connecting an utterance to a preceding utterance or its context (Brinton, 1996, p. 30), i.e. the features I have associated with
discourse markers. They may either signal relevance to preceding utterances (Blakemore, 1987, p. 125) or sequential ordering of utterances (Fraser, 1990). Pragmatic markers have also been regarded as conventionalised response signals, ways of maintaining continuity of discourse, and interactive vehicles (Brinton, 1996, p. 31). Many of the linguistic expressions used as pragmatic markers may have a hedging function in certain contexts. One example of a pragmatic marker with a variety of functions is like. Like has been a topic of research interest for over 30 years and there have been numerous studies on its uses and functions (see e.g. Beeching, 2016, p. 126; and Hasund, 2003, p. 1) for a list of previous studies). Beeching (2016, p. 128) distinguishes between five different functions of non-canonical like. These are exemplifying (We like could go to Nepal), approximative (It will cost me like a couple of grand), quotative (an employer will be like oh you’ve actually worked), focusing (there’re loads of stories in the newspaper about a couple who went abroad and like she’s found dead in the back of a car), and hedging (I wondered like, if your sister could find out for us).\(^5\) Hedging like in Beeching (2016, p. 132) is understood as a means to modalise the text, downplay a potentially dogmatic delivery, and play a face-saving role. However, depending on how the concept of hedging is defined, most of the other subtypes could be understood as hedging as well. In a broad understanding of hedging, like used in its exemplifying, approximative, quotative and hedging sense could all be seen as hedging. Exemplifying and approximative like could be used to express non-prototypicality. Quotative like could indicate that the quote given is an approximate rendering of what was actually said.

Although a lot of effort has gone into studying the functions of pragmatic markers, it is challenging to determine exactly what type of function(s) a pragmatic marker has in every situation. Furthermore, it is challenging to distinguish between functions in natural language, and sometimes it seems as if a pragmatic marker may serve more than one function simultaneously. Like, for example, often occurs at the beginning or end of clauses and has been described as having a linking function between units of talk, i.e. functioning as a discourse marker structuring the text (Beeching, 2016, p. 132). However, in this position, it also often functions as a focuser highlighting an upcoming piece of text.

Another challenge when studying hedging and pragmatic markers is the issue of what the overarching concept is. As illustrated with like above, hedging is often seen as a function of certain pragmatic markers in certain contexts. In this view, hedging is defined in a narrow sense, often either as a means of downplaying the message or as a means of indicating vagueness. However, when hedging is defined in a broad sense, e.g. in line with Kaltenböck et al. (2010) in section 2.1, a pragmatic marker can be

\(^5\) The examples are abridged from Beeching (2016, p. 128).
seen as one potential realisation of the hedging function, with several distinct types of meaning. Thus the perspective varies depending on what is studied.

Modality

Modality has been said to be the most important concept that cuts across the field of hedging (Markkanen & Schröder, 1997, p. 6). Modality can be defined as a philosophical concept, a logical concept, or a grammatical concept. From a grammatical perspective, modality is concerned with the status of a proposition with respect to the event it modifies (Palmer, 2001, p. 1). In other words, modality concerns the speaker or writer’s attitude towards the factual status of the proposition. Since hedging concerns the speaker or writer’s attitude towards the content of the proposition, there is an overlap between the two concepts.

A distinction can be made between propositional modality and event modality. Propositional modality is concerned with “the speaker’s attitude to the truth-value or factual status of the proposition”, e.g. *Kate must be at home now*, while event modality is concerned with “the speaker’s attitude towards a potential future event,” e.g. *Kate must come in now* (Palmer, 2001, pp. 7-8). Propositional modality is further sub-divided into epistemic and evidential modality. Epistemic modality concerns the speaker’s judgement about the factual status of the proposition or, more specifically, the speaker’s assumptions, assessment of possibilities, and confidence or lack of such in the truth of the proposition (Coates, 1987, p. 112), whereas evidential modality concerns the speaker’s indication of the evidence he/she has for the factual status of the proposition (Palmer, 2001, p. 8). From the perspective of hedging, epistemic modality is the most relevant concept. Hyland (1998, p. 2) states that a “writer or speaker’s judgements about statements and their possible effects on interlocutors is the essence of hedging” and places epistemic modality at the centre of this field. However, epistemic modality may also be regarded as a broader concept than hedging. In a study of epistemic modality, Holmes (1982) argues that epistemic modality comprises all degrees of certainty and conviction. Nevertheless, epistemic modal markers, such as modal verbs (*might, could*) and adverbs (*perhaps, possibly*) are often used to indicate uncertainty on behalf of the speaker towards the truth of the proposition, and are acknowledged as a frequent and important type of hedging devices (Kranich, 2011, p. 80; Vold, 2006, p. 61). Similarly, expressions such as *I think, sort of* and tag questions are linguistic forms which are used to express epistemic modality and thus also hedging (Coates, 1987, pp. 114-120).

Although there is a clear relationship between hedging and modality, the nature of the relationship is less clear. The degree to which the two concepts overlap, depends on their respective definitions
(Markkanen & Schröder, 1997, p. 7). In the same way as pragmatic markers and hedging are placed in different relations to each other, modality, particularly epistemic modality, can also be seen as either a way of realising the function of hedging or vice versa. Hyland (1998), Holmes (1982) and Wong (2009) all see hedging as a part of modality: “[…] hedging is one part of epistemic modality; it indicates unwillingness to make an explicit and complete commitment to the truth of propositions” (Hyland, 1998, p. 3). In Wong’s view, hedging indicates a lack of commitment or confidence in the proposition, i.e. uncertainty, whereas epistemic modality indicates the degree of commitment in the proposition, ranging from “uncertain possibility […] to confident necessity” (Wong, 2009, p. 40). Kranich (2011, p. 80), on the other hand, observes that expressions of epistemic modality are an important subtype of the linguistic strategies used for hedging. Epistemic modality, when used to weaken the force of a proposition may be considered a type of hedging.

Vague language

Vague language is another concept which overlaps with hedging and has been studied alongside it. In their bibliographical guide to the study of hedging, Schröder and Zimmer (1997, p. 255) list some 30 studies on hedging within the related concept of vagueness between the late 1970s and mid-1990s. Vague language is a prevalent feature of spoken language in general and informal conversations in particular (O’Keeffe, McCarthy, & Carter, 2007, pp. 74, 177). Like hedging, vague language is defined in a variety of ways and has been studied within a variety of fields, e.g. psychology, philosophy, literary criticism and linguistics (Channell, 1994, p. 5).

From a grammatical point of view, vague language has been explained as the use of generic nouns such as thing and stuff, i.e. noun phrases which refer to “a whole class rather than to an individual person or thing” (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 265), general extenders such as or whatever, and and so on, and words and phrases such as sort of, kind of, like, etc. (Carter & McCarthy, 2006, p. 202). Vague language can be realised in several ways. Stenström, Andersen, and Hasund (2002, pp. 87-88) refer to expressions that indicate total vagueness, e.g. thingy, summarising phrases, e.g. and things, vague generic terms, more specifically, vague quantifiers e.g. bags of, adverbs of frequency e.g. seldom and approximation in the form of approximate numbers and quantities, e.g. around 10 people, and the suffix -ish. Approximating expressions, which are typically used with quantities to indicate an approximate rather than precise amount, are however sometimes seen as separate from but related to vague language. Fraser (2010, p. 26) sees vagueness as a perlocutionary effect, rather than an intrinsic property of certain words or expressions, which occurs “when the information you receive from a speaker lacks the expected precision.”
The many intentions behind using vague expressions overlap to a large degree with those of hedging. Both hedging and vague language have been associated with fuzziness (O’Keeffe et al., 2007, p. 177), and propositional vagueness may create propositional hedging (Fraser, 2010, p. 26). In the same way as hedging, vague language may be used to soften or tone down statements that could otherwise be perceived as too direct, assertive, or pedantic. Vague language can also be used to “indicate assumed or shared knowledge” (O’Keeffe et al., 2007, p. 177). Vague expressions signalling shared knowledge may either be endocentric, i.e. oriented towards previously agreed common ground between the speaker and the hearer, or exocentric, i.e. oriented towards an alternative meaning outside the interlocutors’ realm (Romero-Trillo, 2015, p. 55). Furthermore, vague language may be seen as a part of a cooperative strategy “to avoid providing more information than is required” (Andersen, 2010, p. 35). Thus, like hedging, vague language may have both approximative and relational functions.

Other reasons for using vague language could be memory loss, lack of a suitable precise word, lack of knowledge, or to signal in-group membership, as exemplified by Stenström et al. (2002, p. 88), who found that it is regarded as cool among teenagers to demonstrate that one cannot be bothered to be precise. In writing, the use of vague expressions may be a way for writers to “hedge the claims for their theories and research findings” (Carter & McCarthy, 2006, p. 202). Vague language is also often seen in the level of formality of the situation (Stenström et al., 2002, p. 86). The less formal the situation, the more prevalent vagueness is.

Looking at vague language in relation to hedging poses some challenges. First, “the literature on vagueness does not always include the hedging phenomenon despite the fact that the lexical and or syntactic features of vague utterances are closely related to hedging” (Clemen, 1997, p. 240). Second, when explaining the two concepts, one could potentially get lost in circular argumentation. For example, vague language is often seen as one potential realisation of hedging, i.e. a hedging effect may be achieved through the use of vague language. For example, generic expressions may be used to signal approximation, the purpose of which may be to cause a hedging effect (Kaltenböck et al., 2010, p. 6). In Prince et al.’s (1982) classification framework for hedging strategies, words and phrases indicating approximation are classified as a sub-type of hedging strategies called approximators. At the same time, in many descriptions and classifications of vague language, hedges are seen as a type of vague language alongside approximators. Thus, form a vague language perspective, hedging and approximators are seen as types of vague language. Furthermore, not all types of approximating expressions are necessarily hedging strategies, e.g. exaggerations, as in the example provided by
Channell (1994, p. 89), *A million students came to see me today*, where the approximation constitutes a booster rather than hedging.

**Mitigation**

Mitigation is another concept which overlaps with hedging to a large extent. Mitigation has been the object of theoretical and empirical studies for more than three decades. (See e.g. Thaler, 2012 for an overview of studies.) Mitigation can be defined as the modification of a speech act to reduce unwelcome effects on the hearer (Fraser, 1980, p. 341). Fraser (1980, pp. 341–344) explains mitigation by addressing four aspects: 1), Mitigation is not in itself a speech act but rather an attempt to reduce any unwelcome effects of a speech act on the hearer; 2), mitigation only modifies those effects that are unwelcome to the hearer; 3), mitigation is not the same as politeness, but mitigation entails politeness; and 4), mitigation is not the same thing as hedging. At the time, hedging was seen as a way of making something more or less fuzzy, whereas mitigation was seen as a way of softening one’s statement (Fraser, 1980, p. 344). As the concept of hedging has broadened, the overlap between the two concepts has become more extensive, and now hedging is often regarded as the mitigation of claims to influence the interlocutor (Vold, 2006, p. 62). Furthermore, both hedging and mitigation have been referred to as relational concepts or relational language, which are used to create and maintain a good relationship between interlocutors (Caffi, 1999, p. 885; O’Keeffe et al., 2007, p. 159).

As with hedging, the need to explain and classify mitigation strategies created a need for classificatory frameworks. Fraser (1980, p. 344) distinguishes between *self-serving* and *altruistic* mitigation strategies. Self-serving mitigation strategies are strategies used by the speaker to reduce unwelcome effects on the hearer while simultaneously relieving him/herself of some of the responsibility for causing these effects in the first place (e.g. *It is my duty to tell you that you are fired*) or from any implications these effects may have (e.g. *Difficult as it is to believe, given your lifestyle, the analysis shows you have a social disease*) (Fraser, 1980, p. 345). Altruistic mitigation strategies are strategies only concerned with reducing the negative effects on the hearer. Both types of strategies may be applied at the same time.

Mitigation strategies may also be classified according to their different scopes in a particular speech act in the same way as hedging strategies in the framework of Prince et al. (1982) for example. Caffi (1999, p. 888) distinguishes between propositional strategies and strategies which influence the relationship between the speaker and the proposition. Mitigation may take place on or within the proposition, on the illocution, i.e. the speaker’s commitment towards the proposition, and on the deictic
origin of the utterance (Caffi, 1999, p. 888). Caffi (1999, p. 890) calls these three different types of mitigation strategies *bushes*, also referred to as *propositional hedges, hedges*, and *shields* respectively. Bushes mitigate the propositional content and often function in the same way as approximators, i.e. some element of the propositional content is typically made less precise. Hedges mitigate the illocution and say something about the speaker’s epistemic commitment to the propositional content (Caffi, 1999, p. 892). Shields mitigate the source of the utterance; however, it is not always mitigated by explicit linguistic means. Shields are related to the category of shields in Prince et al. (1982). Both descriptions involve some kind of “defocalisation of the speaker,” i.e. where the speaker shifts the responsibility to another source (Caffi, 1999, p. 896). In Caffi’s description of shields, this source is not always shared or accessible to the hearer. Shields function as a way of weakening the speaker’s commitment to what he/she is saying, but “the weakening operation takes place at a deeper, more abstract level: for instance, it affects syntax, as in passive transformations, or morphology, as in the shift from first-person singular pronouns to other person pronouns” (Caffi, 1999, p. 889). These types of strategies may be applied at the same time (Caffi, 1999, p. 888). Today, hedging in a mitigation perspective is regarded as a subtype of mitigation in line with the classificatory framework for mitigation strategies proposed by Caffi (1999) (Thaler, 2012). However, from a hedging perspective, particularly when applying a broad understanding of hedging, mitigation and hedging are more or less the same thing.

2.1.3 Summary of previous research on hedging and related concepts

Sections 2.1.1 and 2.1.2 have illustrated some central aspects of the development of the concept of hedging and how hedging research cuts across research on other related pragmatic phenomena. Figure 2-1 summarises this development chronologically. This is, however, by no means a complete overview of hedging research, but an overview of some of the important contributions which serve as the theoretical backdrop for the studies performed as a part of this thesis. Figure 2-2 illustrates how hedging overlaps with other related pragmatic concepts.
Building on the research presented in sections 2.1.1 and 2.1.2, I will use this section to explain how hedging is defined and understood in this thesis. Here, hedging is understood as a pragmatic function belonging to the interpersonal domain of language. It may be realised through an indefinite number of linguistic forms and strategies, which can be employed singly or in combination with each other. Hedging may be described as a type of metalanguage guiding the hearer in the understanding of an utterance. However, hedging does not entail all linguistic devices expressing attitudes or feelings on
behalf of a speaker. For example, in an utterance such as To my surprise, I was allowed to cross the border the phrase to my surprise, although expressing a feeling, does not constitute a hedging strategy. Hedging strategies, by contrast, are expressions of uncertainty, caution or reduced commitment on the part of the speaker, such as in If I’m not mistaken, the earth consists of 70% water, where the speaker is reducing his/her commitment to the proposition.

The definition of hedging applied in this thesis is based on the definition by Kaltenböck et al. (2010) presented in section 2.1; however, an additional perspective is included. In this thesis, hedging is defined as discourse strategies that reduce the force, truth or perceived negative effect on the hearer of an utterance. Thus, hedging signals the speaker’s commitment to the truth of a proposition, the speaker’s commitment to the force of an utterance, and the speaker’s wish to invite the hearer to engage in a dialogue to seek ratification for the content of the utterance. This latter aspect of hedging has been discussed in studies of epistemic modality. Coates (1987, p. 120), for example, states that “in conversation, speakers communicate not just propositions and attitudes to propositions, but also attitudes to addressees.” However, this has not been given similar attention in hedging research and thus is often not included in definitions of hedging although there are exceptions. Hedging on the relationship between the interlocutors is described in Hyland’s (1996, 1998) account of hedging in scientific discourse. In scientific texts, claims are often hedged to invite the reader to take part in the ratification of the knowledge they contain. In my opinion, this idea is transferable to the use of hedging in informal conversations, where speakers often invite their interlocutors to co-construct meaning or support their claims and beliefs. For example, the use of the first you know in (2.1) may serve as a way of inviting the interlocutor to co-construct the rest of the utterance, which speaker B does. The second use of you know in final position with rising intonation is an example of speaker A seeking confirmation for his/her views.

(2.1)⁶
A: (...) she raised me (. ) albeit in a you know
A: in a bad way but
B: >> a cack handed way yeah
A: she still did it and and there was a lot of love there for me you know?

(BNC2014 SEPP no. 613)

⁶ Examples included from the corpora are included as they are transcribed in the corpora including transcription symbols. See table 4-3 in section 4.4 for an explanation of relevant transcription symbols.
2.2 Corpus pragmatics

Corpus pragmatics has been defined as “the science that describes language use in real contexts through corpora” (Romero-Trillo, 2017, p. 1). Corpus pragmatics is a relatively new sub-field of linguistics, which stands at the intersection between the established fields of corpus linguistics and pragmatics. Corpus linguistics and pragmatics are not fields on the same level, so to speak. Corpus linguistics is defined by a way of doing language research (see section 2.2.1 for a discussion of the status of corpus linguistics), whereas pragmatics is defined by the aspect of language that it deals with. Corpus pragmatics has therefore also been regarded as a specialised sub-field of corpus linguistics (Weisser, 2018, p. 9). However, researchers have also argued that corpus pragmatics forms a part of empirical pragmatics (Huang, 2017a). Corpus linguistics and pragmatics were for a long time regarded as two distinct and incompatible disciplines characterised by two distinct ways of thought (Romero-Trillo, 2008, p. 2). In general, corpus linguistics has been strongly associated with quantitative methods, whereas pragmatics has been associated with qualitative methods and strong dependence on context (Rühlemann & Aijmer, 2015, p. 2). In recent years, however, the two disciplines have found some common ground resulting in the new discipline of corpus pragmatics.

In simple terms, corpus pragmatics can be defined as a discipline which combines key methodologies from corpus linguistics and pragmatics, with the purpose of researching pragmatic phenomena with the aid of corpus linguistic methodology (Rühlemann & Aijmer, 2015, p. 9). Yet, studying pragmatic phenomena through corpora is challenging as pragmatics is concerned with how language is used as a tool in communication, including what is not being said explicitly, but is situationally, culturally or socially given, which may be difficult to discover from corpus data. In general, corpora have been compiled to study linguistic forms in large databases (Flöck & Geluykens, 2015, p. 9), and as a result, core features of pragmatics, such as hedging, are harder to catch with corpus methodology (Taavitsainen & Jucker, 2015, p. 12). This is also why corpus linguistics came into pragmatics later than into other sub-fields of linguistics, such as lexicography. However, being a young field, corpus pragmatics is constantly under development, and there is still a need for studies uniting elements from the two fields (Romero-Trillo, 2008). Furthermore, technological developments, e.g. more advanced pragmatic tagging tools, may impact the field and change the research agenda in the future.

The following sections will outline how corpus pragmatics developed from corpus linguistics (section 2.2.1) and pragmatics (section 2.2.3) and give an overview of some central approaches to corpus pragmatics today. Relevant sub-fields of corpus linguistics and pragmatics respectively will also be
addressed. Learner corpus research will be presented in section 2.2.2 and contrastive, cross-cultural, intercultural and interlanguage pragmatics in section 2.2.4. A summary of the characteristics of corpus linguistics and pragmatics, and how they have been carried over into corpus pragmatics will be given in section 2.2.5. In section 2.2.6, the overlap between corpus pragmatics and learner corpus research is addressed.

2.2.1 Corpus linguistics

Whereas corpus pragmatics can be described as a new tradition, corpus linguistics and pragmatics are well-established traditions within linguistics. The term *corpus linguistics* first appeared in the early 1980s, but depending on how a corpus is defined, its roots go back further than that. The word *corpus*, comes from Latin and means a body, mass, collection, or matter of any kind, and the tradition of using collections of texts as a foundation for studying various aspects of language can be traced back to the 13th century and the attempts of word indexing and concordancing of the Christian Bible (Yao, 2015). Examples of corpus work are also found in the development of dictionaries, e.g. Samuel Johnson’s *Dictionary of the English Language* published in 1755, which was based on natural language in the form of approximately 150,000 sentences recorded on slips of paper (Biber & Reppen, 2015, p. 2). At the beginning of the 20th century, collections of naturally occurring language were also starting to be used in language teaching and the creation of grammars. A famous example is Otto Jespersen’s *A modern English grammar: on historical principles*. Gradually, corpora were applied in a wide range of linguistic disciplines: lexicography, translation, sociolinguistics, forensic linguistics, and more. (See O’Keeffe and McCarthy (2010) for more examples.)

The emergence of corpus linguistics as an academic discipline, however, is closely connected to the increasing ability to collect and store large quantities of data. The compilation of the first electronic corpus in the 1960s marked the beginning of a new era of corpus linguistics. In 1967, the 1-million-word Brown corpus of written American English texts was published (Kučera & Francis, 1967). A few years later, the parallel Lancaster-Oslo/Bergen (LOB) corpus was compiled. These first-generation corpora are relatively small by today’s standards. In the 1980s, studies based on larger electronic corpora began to appear, mainly due to their widespread availability. The 1980s and early 1990s are referred to as the era of the second-generation corpora and are characterised by the introduction of corpora with more than 100 million words, such as the British National Corpus, completed in 1994, and the Bank of English, with 211 million words per 1995 (Sinclair, 2004). Towards the end of the 1990s, researchers were using quantitative methods to a larger degree to describe how different
structures were used in actual text and explored the degree to which their results were generalisable (Ädel & Reppen, 2008, p. 1). Since then, many more corpora have been compiled, and the easier availability of corpora has led to corpus linguistics having an impact on many fields. Furthermore, researchers are applying increasingly sophisticated statistical methods to their corpus data. The great influence of corpora on language research in general has by several been referred to as a corpus revolution or a paradigm shift (Rühlemann & Aijmer, 2015, p. 4). The introduction of corpora into linguistics has made it possible to study patterns which previously have been impossible to discover through hand and eye analysis, and today, corpus linguistics is characterised by quantitative investigations and vertical reading of the data, often using a key word in context (KWIC) function.

Today, many different types of corpora are available, some of which will be described here. A monolingual corpus consists of texts in one language only and is the most frequent type of corpus. In a multilingual corpus, more languages are included. A parallel (translation) corpus have been said to contain “the same text samples in each of two languages, in the sense that the samples are translations of one another” (Oakes & McEnery, 2000, p. 1). The text samples are usually aligned so that pieces of text, sentences, paragraphs, etc., correspond across the languages. An advantage of a parallel corpus is that it allows the researcher “to make specific and fine-grained comparisons on the basis of texts which are interlingually comparable” (Aijmer, 2020, p. 3). A parallel corpus can be either unidirectional, i.e. the original texts are in one language, and the translations are in another, or bidirectional, i.e. original texts in all the concerned languages, and translations into the other languages (Hasselgård, Forthcoming, pp. 4-5). Comparable corpora are two or more monolingual corpora with texts relating to the same topic and “whose components are chosen to be similar samples of their respective languages in terms of external criteria such as spoken vs. written language, register, etc.” (Tognini-Bonelli, 2001, p. 7). These corpora often also have corresponding metadata, i.e. sociolinguistic information about the contributors. However, the terms parallel and comparable have been used in different ways (Johansson & Oksefjell, 1998, pp. 5-6). The term parallel has been used to refer to both corpora including translated texts and comparable corpora. The term comparable has been used for both monolingual corpora of comparable texts (e.g. Brown and LOB) and corpora of comparable texts in different languages. Comparable corpora are different from translation corpora in that no direct alignment is possible between pieces of text, but such corpora enable comparisons of phenomena across languages (Tognini-Bonelli, 2001, p. 7). A learner corpus consists of text produced by learners of a second/foreign
language and is an example of specialised corpora which contribute to the description of a language variety rather than a specific language norm (Gilquin, 2015, p. 9). A learner corpus can be defined as a large computerised systematic collection of authentic, continuous, and contextualised spoken or written language produced by learners of a foreign or second language (Callies & Paquot, 2015, p. 1; Gilquin, Papp, & Díez-Bedmar, 2008, p. vii). Learner corpora can be used to identify characteristic patterns in non-native speech or writing and they are often annotated for errors, making it possible to study types of errors made by language learners. There are also comparable learner corpora which enable the study of language produced by learners with different first languages. The study of learner corpora has developed into a separate subfield of corpus linguistics, Learner Corpus Research (Meunier, Gilquin, & Granger, 2015) which will be discussed in more detail below.

Corpora may also be described according to the types of texts they include, e.g. spoken or written texts, or whether they contain raw texts or annotated texts. When a corpus is annotated, interpretative linguistic information is added to the corpus (Leech, 2005). Part of speech (POS) tagging is frequently used on various types of corpora and involves each word in a corpus being marked according to its part of speech. As for spoken and written corpora, written corpora greatly outnumber spoken corpora mainly due to the cost-intensive and time-consuming process of transcribing spoken data. This has contributed to what is referred to as “the written language bias” in linguistics (Linell, 2005, 2019). This bias is related to the status of spoken language. For many years, the notion prevailed that speaking was a corruption of writing, and grammars and research on language were based solely on written language (Carter & McCarthy, 2015, pp. 3-4). Although the status of spoken language has improved, the written language bias remains. In corpus linguistics, the main reason for the written language bias is the imbalance in the availability of spoken and written corpora. Although several spoken corpora have been compiled and made available in recent years (e.g. BNC2014 in 2017, London–Lund Corpus 2 in 2019, and corpora which are not publicly available, such as the Cambridge and Nottingham Corpus of Discourse in English CANCODE (McCarthy, 1998), these are greatly outnumbered by corpora of written language (Adolphs & Carter, 2013, p. 1). As pointed out, the greater availability of written corpora is mainly due to the fact that a written corpus is easier and less costly to compile. A lot of research in corpus linguistics, e.g. in the area of lexicography, has required a high degree of recurrence of the word or expression investigated, and therefore corpora with multi-million entries have been necessary (Adolphs, 2008, pp. 2-3). This has only been possible to achieve in writing, mainly for financial reasons. Whereas a written corpus can be collected relatively quickly, especially if the texts needed are electronically available on the internet, compiling a spoken corpus involves the time- and
cost-intensive process of transcription. Burnard (2002) estimates that the cost of collecting and transcribing 1,000,000 spoken words is 10 times higher than adding 1,000,000 words from newspapers to a corpus. A representative corpus of spoken texts is even harder to compile due to requirements of e.g. size and demographic spread (Love, Dembry, Hardie, Brezina, & McEnery, 2017, p. 325). Another reason for the lower number of studies using spoken corpora is the “scepticism towards the applicability of corpus-based techniques to issues beyond the clause boundary” (Conrad, 2002, p. 86). Spoken language, particularly informal unplanned spoken language, is often fragmented and challenging to study within a framework primarily based on written language.

A major drawback of the focus on written discourse is lack of studies on phenomena which are rare or non-existent in writing, e.g. turn transitions, prosodic information, etc. Carter and McCarthy (2015, p. 6) mention a range of items which are rare in writing but frequent in spoken interaction. Many of these items occur outside the conventional clause structure and are thus often not included in traditional grammars that focus on clauses and their structures. Such items, however, may be extremely important in a conversational setting, where they convey politeness, emotions, and other conversational cues guiding the hearer in the interpretation of the utterance. The study of such items has challenged existing metalanguage. For example, expressions such as fine, great, absolutely, definitely and indeed are labelled response tokens by Carter and McCarthy because the traditional word-class labels, adjective, and adverb, “fail to do justice to their ability to comment on whole previous turns and prior stretches of discourse” (2015, p. 8). This is also relevant to hedging strategies, which may take on nearly any linguistic form and belong to nearly any word class (see Article 3), and whose function may vary depending on the context.

There is an ongoing debate about the status of corpus linguistics; whether it is a theory or a method. When corpus linguistics is regarded as a theoretical approach, it is seen as a paradigm leading to new research questions and new perspectives on language. Thus it is not regarded as merely an analytical tool, but an important concept or a paradigm in its own right. According to Sinclair (2004, p. 9), one of the most important advocates of this view, language should be inspected with as little attention as possible to existing theory. Language needs to be reinvestigated taking the available large quantity of empirical evidence, i.e. corpora of authentic text, into consideration. The main purpose of such studies is to discover things that have not yet been discovered (Huang, 2017a). From the corpus-as-theory perspective, raw text is the source of patterns of language use, and annotation of data, e.g. part of speech tagging, etc., should not be given too much focus, because it may hide language patterns (Stubbs, 2009, p. 116). Instead, one should “trust the text” (Sinclair, 2004). However, only a minority regard corpus
linguistics as a theory. Researchers mainly focus on the methodological aspects, although most corpus linguists agree that it is more than merely a methodological approach because of its significant impact on linguistics (Reppen & Biber, 2015, p. 2).

When corpus linguistics is considered a methodology, it is not regarded as a paradigm alongside sociolinguistics, or cognitive linguistics but as “a way of doing linguistics” (Meyer, 2002, p. xi). Corpus linguistics is also often regarded as a part of the applied linguistics domain. Whereas sociolinguistics and cognitive linguistics study language in relation to social factors and cognitive processes respectively, there is no such relationship for corpus linguistics (Reppen & Biber, 2015, p. 1). Corpus linguistics, on the other hand, is characterised by its ability to represent a specific language domain and is “a methodological basis for pursuing linguistic research” (Leech, 1992, p. 105) and an approach to the study of language which does not adhere to any particular linguistic theory (McArthur, Lam-McArthur, & Fontaine, 2018). However, even within the corpus as a method strand, there are disagreements concerning how corpus linguistics should be defined. If corpus linguistics is regarded as a set of procedures and methods for studying language, there is no consensus on the set of methods and procedures involved (McEnery & Hardie, 2012).

Despite this ongoing discussion, some researchers have pointed out that it might not be necessary to characterise corpus linguistics as either a theory or a method. In an interview published post-humously, Stig Johansson, an important contributor to corpus linguistics, stated that it would be a pity to restrict the term to a particular use or application, although for most people who use corpora, the emphasis is probably on methodology (Johansson, 2011, p. 117). The overarching belief is that language analysis should be based on real data, i.e. actual instances of spoken or written text, rather than made-up examples or pure intuition. Corpus linguistics can be seen as an empirical approach to language with authentic data as a starting point (Tognini-Bonelli, 2001, p. 2).

A distinction is generally made between three types of corpus studies: corpus-informed, corpus-based, and corpus-driven (Johansson, 2011, pp. 116-117). In corpus-informed studies, corpora are mainly used to give examples of a phenomenon or to support the theory presented. In corpus-based studies, the corpus is an essential element, but not necessarily the only source of data (Johansson, 2011, p. 117). Pre-existing theories and descriptions of language are used as a point of departure, and the corpus functions as a way to test, exemplify, adjust or challenge them, i.e. the corpus is approached with a set of pre-existing assumptions and expectations (Huang, 2017a). Such studies with existing theories as a point of departure are also referred to as *top-down* studies. Corpus-driven studies have been described
as the perhaps most radical use of corpora. “The hallmark of corpus-driven studies is respect for the data and the attempt to deal as fully as possible with the evidence the corpus provides” (Johansson, 2011, p. 117). In corpus-driven studies, theory has no independent existence from the evidence (Tognini-Bonelli, 2001, p. 84). The issue of representativeness is very important when the corpus is used to create new theories. If it is unrepresentative, the risk of errors is great. Corpus-driven studies are often also referred to as bottom-up studies, i.e. studies that do not impose any restrictions on the data but let the data decide. (See section 5.1.1 for an explanation of how the term bottom-up is used here.) However, it may be difficult to categorise a study of either corpus-based or corpus-driven, as studies may include both top-down and bottom-up elements and the distinction between the two is not as rigid in practice. Some authors have also chosen to use the term corpus-based in a broader fashion encompassing both corpus-based and corpus-driven approaches (McEnery & Hardie, 2012, p. 151), while others adhere to the stricter understanding of corpus-driven and do not accept any terminology or models developed prior to the investigation. It might therefore be more useful to regard them as points along a continuum illustrated in figure 2-3.

![Corpus-informed–corpus-driven continuum.
](image)

**2.2.2 Learner corpus research**

The field of corpus linguistics has developed several sub-fields, one of which is particularly relevant to this study, namely that of learner corpus research (LCR). LCR sprung out of corpus linguistics in the 1980s as a reaction to certain analytical approaches within the field of second language acquisition (SLA) research (Romero-Trillo, 2018, p. 114), but it did not begin to flourish until the 1990s (Paquot & Plonsky, 2017, p. 62). The purpose of LCR was to contribute to SLA research by understanding second-language (L2) production in comparison with first-language (L1) data, particularly by identifying linguistic elements that were overused or underused by learners (Granger, 1993, 1994; Romero-Trillo, 2018, pp. 114, 116), and to develop new pedagogical tools and methods in second- or
foreign-language teaching (Granger, 2002, 2008). Learner language was often compared to native speaker language, and quantitative analyses of learner language and native speaker language were performed (Romero-Trillo, 2018, p. 115).

LCR has been heavily influenced by the work of Sylviane Granger and her team (Romero-Trillo, 2018, p. 115). The compilation of several corpora of written and spoken learner language, e.g. the International Corpus of Learner English, ICLE (Granger, Dagneaux, & Meunier, 2002), and the Louvain International Database of Spoken English Interlanguage, LINDSEI (Gilquin, De Cock, & Granger, 2010), has made it easier to study differences between learner language and native language, as well as similarities and differences across learner varieties. Since the release of these corpora, several other learner corpora have also been released, such as the longitudinal corpus of Spanish learners of English, the Universidad Autónoma de Madrid-Learner English Spoken Corpus (UAMLESC) (Romero-Trillo & Llinares-García, 2008). This improved access to learner data “has profoundly increased the interest in learner language” (Hasselgård & Johansson, 2011, p. 37). However, LCR and the practice of comparing non-native varieties to native varieties have been criticised for not being informative enough for studying learner pragmatic development (Ishihara & Cohen, 2010) and for being guilty of the “comparative fallacy” (Bley-Vroman, 1983), i.e. failing to analyse interlanguage as a linguistic system in its own right. A central argument in this criticism has been that learners do not necessarily strive to reach a native-like level of proficiency, and that interlanguage should be analysed as a language variety in its own right (Hunston, 2002, p. 211). Romero-Trillo (2018, p. 114), however, argues that an object of such studies is to “bridge the gap between the language L2 learners listen to and the language they produce.” Similarly, researchers have argued that the comparison of learners to native speakers can be instrumental in improving learners’ language in terms of lexicogrammaticality, acceptability, and idiomaticity (Mukherjee, 2005, p. 16).

In a summary of the state of the art of learner corpus research, Callies (2015, p. 38) states that the majority of LCR studies to date are “corpus-based, quantitative, cross-sectional and comparative.” Another characteristic of most LCR research is that it is mostly done on written text produced by learners, i.e. mostly argumentative essays produced in university settings by young adult advanced learners of English (Callies, 2018). Friginal et al. (2017, p. 3) specify this further and state that school essays, standardised tests and proficiency tests, and laboratory or research reports have been studied extensively in various publications. However, learner corpus research is a field in constant development and is undergoing a methodological reform introducing more complex statistical methods (Paquot & Plonsky, 2017, p. 61). In this endeavour, LCR has been criticised for over-relying on tests of statistical
significance (Paquot & Plonsky, 2017, p. 61). The statistical measures typically applied have been measures of frequency and that other statistical measures, such as dispersion is overlooked (Paquot & Plonsky, 2017, p. 63). To be able to make sound comparisons across languages, researchers have argued in favour of combining measures of frequency and dispersion (Gablasova, Brezina, & McEnery, 2017; Gries, 2006). By studying both the frequency and dispersion of a linguistic feature, the heterogeneity of groups of language learners and native speakers is acknowledged. (See also article 4.)

Despite a variety of contributions to LCR, there is still a gap in the knowledge of spoken English L2 registers. This is also true of research on hedging in spoken English by Norwegian learners. A study of Norwegian learners which has received some attention is that of Hasselgreen (2004). Hasselgreen studied the use of “smallwords” by Norwegian learners and the role of these words with respect to the learners’ fluency in English. She compared the use of expressions such as I think, a bit, and things, etc., in two groups of Norwegian learners, a ‘more fluent’ group and a ‘less fluent’ group, the division into which was based on the pupils’ grades on a speaking proficiency test (Hasselgreen, 2004, p. 160), and one group of native English speakers. The results showed that overall, the native English speakers used the investigated expressions more frequently than the Norwegian learners. Furthermore, there are two unpublished MA theses which focused on the use of certain expressions related to hedging among Norwegian learners of English. Sandal (2016) reported a general under-representation of the discourse marker use of like, well and you know among Norwegian advanced learners of English in the LINDSEI-no corpus. (See chapter 4 for more information about the corpora.) Thomson (2016) studied hedging strategies in English texts written by Norwegian lower secondary school students and reported that the students had a good understanding of hedging strategies, but that the variation among realisations could be improved. Nevertheless, there is still a need for more research on spoken learner language in general and on Norwegian learners in particular.

2.2.3 Pragmatics

Pragmatics, the other subfield of linguistics on which corpus pragmatics is based, also has a long history. Pragmatics has its roots in the philosophy of language, but the origin of modern pragmatics is often credited to the work of Morris (1938). Morris (1938) distinguished between syntax, which addressed the relation of signs to one another, semantics, which addressed the relation of signs to what they denote, and pragmatics, which addresses the relation of signs to their users. Pragmatics can be said to have come into being as a separate field in the 1970s with Grice’s theory of communication (see section 2.1.1 for a presentation of the main traits of Grice’s theory) which is regarded as an important
landmark in the development of a systematic framework for pragmatics. However, the term *pragmatics* was introduced in the 1980s (Archer et al., 2012, p. 3). In broad terms, pragmatics has been defined as the study of language in context (Huang, 2017a, p. 1). For a long time, it was regarded as the “wastebasket” of the study of meaning (Bar-Hillel, 1971), encompassing everything related to utterance meaning which could not be accounted for by formal theories of syntax and semantics. Now, however, pragmatics is regarded as an important and rapidly growing subfield of linguistics (Huang, 2017a, p. 1) and has been said to concern “the interesting stuff about language” (R. T. Lakoff, 1993, p. 367).

Pragmatics is a large field which comprises a variety of different sub-fields, e.g. computational pragmatics, cognitive pragmatics, cross-cultural or intercultural pragmatics, interlanguage pragmatics, etc. The latter two are relevant for this project and will be described in more detail in section 2.2.3 below. Pragmatics also interfaces with other linguistic fields such as second language acquisition and semantics, as well as other disciplines such as artificial intelligence, neuroscience, sociology, and more. Despite its versatility, the perspectives from which pragmatics has been defined and studied have mainly been divided into two traditions: the Anglo-American (cognitive philosophical) tradition, and the European continental (sociocultural-interactional) tradition (Huang, 2017a, p. 2). In the Anglo-American view, pragmatics is regarded as a component of linguistic theory alongside other components such as phonetics, syntax, and semantics. Thus a clear distinction is made between semantics and pragmatics. Pragmatics is regarded as the systematic study of meaning in use and is characterised by theories such as Neo-Gricean pragmatic theory and relevance theory (see footnote 3 in section 2.1.1). The European continental tradition employs a broader and more sociological view of pragmatics (Horn & Ward, 2006, p. 1) in which it is seen as a perspective on linguistic phenomena. This view builds on the work of Morris (1938) and his view of pragmatics as the study of any relations between signs and their users (Verschueren, 2017). However, aspects of the distinction between these two traditions have also been questioned. Verschueren (2017) agrees with the view that these perspectives are two contrastive ways of looking at pragmatics, but also argues that the geographical anchoring reflected in the names is problematic. Not all linguists belonging to the European continental tradition subscribe to the characterisation of pragmatics as a perspective rather than a component.

Different traditions and perspectives have influenced the way pragmatics has been defined, and today there is still “little consensus about, how pragmatics should be delimited from semantics and grammar, the scope of a discipline of linguistic pragmatics and the terminology we need to describe pragmatic phenomena” (Archer et al., 2012, p. 3). For example, Mey (1993, p. 135) defines pragmatics as “the societally necessary and consciously interactive dimension of the study of language”; Crystal (1997, p. 40
301) defines pragmatics as “the study of language from the point of view of users, […] the choices they make, the constraints they encounter in using language in social interaction, and the effects their use of language has on other participants in the act of communication.” Huang (2007, p. 4), on the other hand, defines pragmatics as “the systematic study of meaning by virtue of, or dependent on, the use of language.” No attempt to synthesise the various definitions and approaches to pragmatics will be made here. (See e.g. Huang, 2017b or Horn and Ward, 2004 for a more detailed coverage of the field.) However, one characteristic of the field will be highlighted, i.e. the dependence on context.

Meaning, as it is understood in pragmatics, is characterised by a strong dependence on context. “What is said is always in response to what has been said before, and it creates conditions for what comes afterwards” (Rühlemann & Aijmer, 2015, p. 2). However, context is not restricted to the textual surroundings of a particular linguistic phenomenon, it may also include the physical surroundings of an utterance (Falkum & Kjøll, 2014, p. 182) or even the “cultural contexts of shared meanings and world views” as well as “social contexts through which definitions of self and situation are constructed, to cognitive contexts of past experience and knowledge” (Schiffrin, 1987a, p. 3). Aijmer (2018, p. 566) summarises the aspects of context which affect the interpretation of an utterance: previous discourse, immediate co-text, immediate physical surroundings, social context, and the relationship between the speakers and stored and shared knowledge. The dependence on context is also evident in the methodological approaches associated with pragmatic research.

Pragmatic research is in essence qualitative due to the focus on individual texts and the need to know the contextual parameters to interpret the pragmatic meaning (Rühlemann & Aijmer, 2015, p. 3). Because of its dependence on context, pragmatics is characterised by horizontal reading (as opposed to the vertical KWIC concordance perspective of corpus linguistics) and consequently, pragmatic analyses have typically relied on a small number of texts of which horizontal reading is possible. Methodologically, pragmatics has also been associated with methods where the situational factors can be controlled to some extent, for example Discourse Completion Tasks (DCTs), role plays, and other elicitation methods. Such methods have been used to study a variety of pragmatic phenomena, such as various speech acts and expressions of politeness. (See for example Blum-Kulka & House, 1989; Eisenstein & Bodman, 1986; and Hartford & Bardovi-Harlig, 1992.) Pragmatics takes as its starting point the notion that meaning is not a stable counterpart of linguistic form, but is dynamically generated in the process of using language (Verschueren, 1999, p. 10). Thus, pragmatic studies tend to go from function to form. An overview of the most frequently used methods in pragmatics and the level of researcher inference and control can be found in O'Keeffe et al. (2020).
In DCTs, a speaker is provided certain situational prompts including a variety of contextual and situational clues to elicit certain language functions. An advantage this method is thus that the researcher has more control over the variables which in turn makes the study easier to replicate (Golato, 2003). However, recent comparative studies of speech acts have indicated that speech acts elicited by DCTs differ from those observed in authentic conditions and that the type of methodology used can greatly influence the obtained results (Flöck & Geluykens, 2015). DCTs have also been criticised for not eliciting what speakers would actually say in a situation, but what they think they would say (Kasper, 2000) and for giving data that are less complex and more direct than naturally occurring conversations (Beebe & Cummings, 1996). The lack of interaction markers has been explained by the absence of an interlocutor in questionnaire settings (Flöck & Geluykens, 2015). Nevertheless, the breadth of forms available in a corpus may come at the expense of the contextual depth of DCTs, for example (O'Keeffe et al., 2020, p. 6). Thus, various methodological approaches give insight into different aspects and are all important to the field.

2.2.4 Contrastive, cross-cultural, intercultural and interlanguage pragmatics

Contrastive pragmatics “encourages a focus on variation in the ways that different groups enact their distinctive cultural values and norms” (Holmes et al., 2020, p. 1). Thus, contrastive studies of pragmatic phenomena overlap with other disciplines dealing with language comparison such as cross-cultural pragmatics and intercultural pragmatics (Kecskes, 2017), as well as learner corpora and pragmatics (Romero-Trillo, 2018) (Aijmer, 2020). Both cross-cultural pragmatics and intercultural pragmatics are socially or culturally-oriented studies within pragmatics. Cross-cultural pragmatics “compares different cultures, based on the investigation of certain aspects of language use” (Kecskes, 2017, p. 400). The goal of the discipline can in many ways be summarised as highlighting “aspects of language behaviour in which speakers coming from various cultures have differences and similarities” (Kecskes, 2017, p. 400). Intercultural pragmatics focuses on the interaction between people from different cultures. “It investigates the speech production and comprehension of interlocutors who represent different cultures and languages and use a common language (lingua franca) for communication” (Kecskes, 2017, p. 400).

Interlanguage pragmatics focuses on the use of a target language by non-native speakers and has been defined as the investigation of “how L2 learners develop the ability to understand and perform [a speech] action in a target language” (Kasper & Rose, 2002, p. 5). A central element within this sub-field is the focus on the learners’ pragmatic competence. Pragmatic competence is often defined as “the
ability to use language effectively in order to achieve a specific purpose and to understand language in context” (Thomas, 1983, p. 92) or more specifically “the linguistic ability to communicate your intended message with all its nuances in any socio-cultural context and to interpret the message of your interlocutor as it was intended” (Fraser, 2010, p. 15). Within interlanguage pragmatics, a distinction between two types of pragmatic competence is usually made, i.e. pragmalinguistic and sociopragmatic competence (Leech, 1983; Thomas, 1983). Pragmalinguistic competence refers to knowledge about the resources available for conveying communicative acts in a language and the use of these resources to perform such communicative acts, e.g. knowing the various linguistic resources used to express an apology: *I’m so sorry; I apologise, please forgive me,* etc. (Félix-Brasdefer, 2017). Sociopragmatic competence refers to knowledge about the social perceptions underlying a speaker’s communicative actions, i.e. knowledge of social norms and how to act according to those norms. Research into the pragmatic competence of learners has shown that the development of pragmatic competence in a second or foreign language usually does not progress at the same pace as the development of grammatical competence (Romero-Trillo, 2018, p. 117). This may lead to pragmatic fossilisation, where learners have “a near-to-native grammatical and lexical competence but with a limited range of pragmatic resources” (Romero-Trillo, 2002; Romero-Trillo, 2018, p. 113).

### 2.2.5 The influence of corpus linguistics and pragmatics on corpus pragmatics

Given the dependence on context, studies of pragmatic phenomena have traditionally been qualitative in nature. This is in contrast to corpus linguistic studies which are generally quantitative. As noted above, pragmatics has often relied on a close horizontal reading of a small number of texts describing the use of language based on conversational data (Aijmer, 2018, p. 579), corpus linguistics has been characterised by a vertical reading of large quantities of text, often using quantitative statistical methods to study patterns of language use. Researchers have argued that “strictly speaking at least, the only thing corpora can provide is information on frequencies” (Gries, 2009, p. 11), and being a quantitative discipline, statistical operations are typically necessary to establish whether the frequencies are representative for a language as a whole. Within corpus pragmatics, key perspectives and methodologies from both fields are combined and vertical and horizontal analyses are integrated (Rühlemann & Aijmer, 2015, p. 12). Figure 2-4 summarises some of the characteristic features of corpus linguistics and pragmatics and how they are carried over into corpus pragmatics.
Corpus pragmatics studies have two different approaches to the study of pragmatic phenomena: the form-to-function approach, where you start from a linguistic form and investigate its function(s) in a corpus; and the function-to-form approach, where you start from a function and look for potential realisations of that function in a corpus. Most studies have opted for the form-to-function approach mainly because “core features of pragmatics studies [...] are harder to catch with corpus methodology than lexical or morpho-syntactic features” (Taavitsainen & Jucker, 2015, p. 12). Such studies are typically corpus-based. They take as their starting point lexical words or constructions which previous pragmatic research have shown to have pragmatic functions and first search for the word or phrase in question (vertical reading). Then they examine the findings, weeding out unwanted hits, and study the pragmatic functions of the word or phrase (horizontal reading).

Form-to-function approaches generally map words or constructions onto a range of functions (Aijmer, 2018, p. 555). An advantage of using corpora to explore forms is that they can be studied with great precision concerning frequency distribution, collocation, function etc., and corpora make it possible to base the description on authentic examples (Aijmer, 2018, pp. 556-557). Examples of elements which have been studied in this way are pragmatic markers and vocatives, whose function is often dependent on the communicative context. Many forms are multifunctional and function on several levels of
discourse, e.g. textual, subjective and interpersonal. For example, in her study of *like* and one of its Norwegian counterparts *liksom*, Hasund (2003) illustrates how these items functioned on all three levels. This multifunctionality has resulted in several form-to-function studies being performed to add to the semantic descriptions of the forms studied (Unger, 2018), e.g. Berthelin and Borthen (2019) who study the meanings and functions of the Norwegian particle *jo* (‘after all’, ‘of course’). In such studies, corpora often provide raw material for the analysis of functions, and the researcher interprets the functions from the examples (Aijmer, 2018, p. 564).

A distinction is typically made between four main types of form-to-function techniques (Ädel & Reppen, 2008, pp. 2-3): *one-to-one searching* is one approach which involves “investigating a linguistic form through a search term that only yields relevant hits.” The approach referred to as *sampling* involves the use of one or more search terms that are good examples of the linguistic phenomenon in question. This approach is related to the one suggested by Adolphs (2008), who starts with a definition of the speech act she wants to study and subsequently identifies lexico-grammatical strings realising this function (Adolphs, 2008, p. 52). The drawback of this approach is that not all occurrences of the linguistic phenomenon investigated will be captured. *Sifting*, on the other hand, involves searching for one or more forms, which perhaps have multiple meanings and functions, and then manually going through the hits retrieved to discard those that are not relevant. The fourth approach is referred to as *frequency-based listing* and involves using a frequency list generated by a corpus as a point of departure for the study. The researcher chooses one or more terms from the list and uses that as the starting point.

A general weakness of the form-to-function approach is that it often does not identify all instances of a particular function (Aijmer & Rühlemann, 2015). “While CL [corpus linguistics] aligns well with the core principle of pragmatics that meaning is not a stable counterpart of linguistic form, this is also its weakness when using a form-to-function approach” (O’Keeffe, 2018, p. 588). Function-to-form approaches, on the other hand, encompass techniques that start from a (pragmatic) function and identify realisations of that function in corpus data. Functional studies can be corpus-based, but it is difficult to retrieve realisations of functions in a corpus which is not annotated for pragmatic functions. To my knowledge, there are no such corpora in Norwegian and only a small number available in English, e.g. SPICE-Ireland (Kirk, 2013), which is annotated for speech act function. Most of the studies which have been done from a function-to-form perspective have been somewhere on the scale between corpus-based and corpus-driven (see figure 2-5 below). The most recent overview of such studies can be found in O’Keeffe (2018, p. 599ff) and is summarised here.
In her overview, O'Keeffe (2018) describes three techniques which are suitable for small corpora. The first technique is to search in a pragmatically annotated corpus, which enables the researcher to search for a function in a corpus and retrieve all its realisations. This is referred to as the “holy grail” of corpus pragmatics (O'Keeffe, 2018, p. 599). This may become increasingly possible as more corpora will be annotated for pragmatic functions, although the challenge of the mismatch between form and function for most pragmatic phenomena will probably remain. One step in the direction of creating an automatic tagging system for pragmatic functions is the Dialogue Annotation and Research Tool (DART) system (Weisser, 2016). The DART project aims to be able to investigate interaction on a higher level than individual forms (Weisser, 2016, p. 355). However, some level of manual pre- or post-processing will always be required as no automatic tagging system can be completely correct (Weisser, 2016, p. 357).

The DART annotation tool offers annotation on multiple levels, i.e. syntactic, semantic and pragmatic. Pragmatic annotation offers a solution for function-to-form research, but it is very time-consuming (even with the use of annotation tools) and therefore expensive and can be influenced by the subjective interpretations made by the annotator. O’Keeffe suggests that “because of these constraints, it is best applied in small scale studies, where the researcher is conducting the annotation and has an in-depth understanding of the contextual variables and conditions” (2018, p. 602).

The second technique described by O'Keeffe (2018, p. 602) is the sampling, searching, and sifting technique, where the researcher studies a random sample from a corpus and reads it horizontally. This makes the dataset more manageable for the researcher and enables him/her to sift through it and identify pragmatic phenomena. The third technique involves the use of existing research findings as “seeds” (O'Keeffe, 2018, p. 605), i.e. points of departure for further corpus studies. One example of application of this approach is Schauer and Adolphs (2006), who use results from a DCT as the start of their corpus investigation. One advantage is that it generates comparable datasets, and that the DCT enabled them to control the variables of the context for the scenarios.

O'Keeffe (2018, p. 607) also presents techniques, which are suitable for larger corpora. These techniques involve the use of Illocutionary Force Indicating Devices (IFIDs), i.e. linguistic elements that indicate or delimit the illocutionary force of an utterance; genre-specific search inventories from smaller samples; searches of typical lexical or grammatical features associated with a speech act; or searches for metacommunicative expressions (O'Keeffe, 2018, pp. 607-611). Deutschmann (2003) is an example of this technique, where expressions such as forgive, pardon, regret and sorry were used to locate apologies in the BNC. Deutschmann sifted through the data removing all non-apologies and analysed the remaining apologies in the contexts in which they were originally uttered. Kohnen (2008)
is an example of a study using genre-specific search inventories from smaller samples. Kohnen manually sifted through a pilot dataset of church sermons to identify all types of directives, creating a search inventory, which was used to extract directives from a large diachronic corpus. Taavitsainen and Jucker (2008) similarly studied compliments by searching for various adjectives and adjective strings that have a positive evaluative description, i.e. searching for typical lexical features associated with the speech act of evaluation. An example of the use of a metacommunicative expression is found in Jucker and Taavitsainen (2014). Building on their study from 2008, they used the search term *compliment* to find performative, descriptive, and discursive instances of the act. A downside to this method was that it mostly retrieved “accounts of a particular speech act rather than the actual speech act” (O'Keeffe, 2018, p. 612).

Although form-to-function and function-to-form approaches are described as two different sets of techniques, it is not always easy to make a clear distinction between the two, and many approaches overlap. Like corpus-based and corpus-driven studies, form-to-function and function-to-form may be regarded as ends on a scale, as illustrated in figure 2-5, where the end points reflect form-to-function and function-to-form in their purest sense.

![Form-to-function – function-to-form continuum.](image)

Although the use of corpora provides a range of possibilities for studying pragmatic phenomena, researchers have drawn attention to some caveats. The main issue raised is the discrepancy between form and function and the fact that meaning is often dynamically generated in interaction rather than being the property of one form O'Keeffe (2018). O'Keeffe (2018, p. 590) exemplifies this by referring to the form *sorry*, which is commonly used as an apology but can also be used as a request for clarification. This, therefore, requires the researcher to manually sift through the forms retrieved from the corpora, or in another way ensure that they are, in fact, the forms he/she was looking for. Another challenge with using corpora is that the corpus is often gathered by someone else for a specific purpose, which creates a distance between the researcher and the data. O'Keeffe (2018, p. 591) contrasts the use
of corpora with the use of DCTs and argues that one of the strengths of using DCTs is that the researcher can control the context and conditions and have access to the whole situational context.

The use of corpora has also had an influence on contrastive pragmatics to the extent that *contrastive corpus pragmatics* is emerging as a new field of research. Aijmer describes this new field as characterised by “the joint approaches of pragmatics, corpus linguistics and contrastive analysis for describing the similarities and differences between languages” (2020, p. 1). The research on pragmatic phenomena from a contrastive perspective comprises more than 30 years of work and covers a wide range of topics, such as the realisations of speech acts across languages (Holmes et al., 2020, p. 2). However, the introduction of electronic corpora has influenced contrastive studies and has invited (large-scale) empirical studies of similarities and differences across languages. Contrastive corpus pragmatics is often associated with the use of parallel (translation) corpora, but comparable corpora have also proven to be a valuable source in studies of genres where parallel corpora do not exist (Aijmer, 2020, p. 12).

### 2.2.6 Learner corpus pragmatics

*Learner corpus pragmatics* is not an established term in linguistics, but it could be used to describe work combining aspects of interlanguage pragmatics with aspects of LCR and corpus pragmatics. Learner corpora are beneficial to researchers studying interlanguage pragmatics because they contain large quantities of naturally occurring data, which facilitate empirical studies of learners’ performance, and because they enable studies of patterns of language (Vyatkina & Cunningham, 2015, p. 282). The access to learner corpora has led to researchers arguing that previous pragmatic research based on introspective methods and invented examples should be re-investigated using authentic data (Romero-Trillo, 2017, pp. 1-2).

Vyatkina and Cunningham (2015, p. 283) report that the first type of studies emerging from the intersection of interlanguage pragmatics and LCR was word-based studies, i.e. form-to-function, adding to the pragmalinguistic branch of interlanguage pragmatics. Studies of spoken interlanguage have typically focused on the use of pragmatic markers. Examples of word-based studies of both written and spoken text can be found in Vyatkina and Cunningham (2015, pp. 284-286). However, as with corpus pragmatics in general, there is a range of theoretical and methodological challenges to learner corpus pragmatics (Vyatkina & Cunningham, 2015, p. 283). First, access to the necessary level of context and metadata may be limited compared to established methods such as the use of DCTs, where the researcher can pre-define the variables of the study and carefully control the context and
conditions (O'Keeffe et al., 2020, p. 6). Second, the lack of pragmatically annotated corpora may limit the study of pragmatic functions from a function-to-form perspective. Theoretically, a comprehensive framework which integrates both the pragmalinguistic and sociopragmatic aspects is needed to explain the intricate process of acquiring pragmatic competence. A significant attempt to create such a model has been made by Romero-Trillo (2018): The polyhedric pragmatic model for L2 learners explains how interlocutors communicate successfully by understanding the speaker’s communicative intention and adapting existing pragmatic competence to new pragmatic competence in the target language. Nevertheless, learner corpus pragmatics is still in its infancy, and more research is needed to fully explore the potential of corpora in the study of learners’ pragmatic competence.
Chapter 3 Overview of the three phases of the project

This chapter presents the three phases of the research project (outlined in section 1.4) and the overarching theoretical and methodological aspects pertaining to each phase. The chapter builds on the overview of previous research outlined in chapter 2, and the aim of the chapter is to place the contributions of this project in the existing theoretical landscape. The chapter is structured according to the three phases of the project. Section 3.1 describes the theoretical considerations behind the development of a corpus pragmatic approach to the study of hedging strategies in spoken corpora (article 1). Section 3.2 outlines the theoretical background for the main part of the project, i.e. the contrastive study of hedging strategies in Norwegian and English (articles 2 and 3). Section 3.3 describes the theoretical backdrop of the intervarietal study of hedging in English by Norwegian learners (article 4). Since some of the aims of the study are of a methodological nature, it is inevitable that the theoretical considerations presented below blend in with methodological ones, as they emerge in previous studies as well as the present one. The methodological considerations made throughout the project are addressed in more detail in chapter 5 and articles 1–4.

As stated in section 2.1.3, in this thesis, hedging is defined as discourse strategies that reduce the force, truth or perceived negative effects on the hearer of an utterance. The term discourse strategy is understood as any means or method of bringing about a result, but is limited to linguistic elements such as words, phrases, clauses and various syntactic constructions used to perform the pragmatic function of hedging. Other hedging strategies, such as multi-modal and para-linguistic strategies, are not within the scope of this project. Hedging strategies may be realised through nearly any form. Here, form is used synonymously with realisation.

3.1 Phase 1: Developing a method and identifying a probe (article 1)

The first part of the project proposes a method for the study of hedging strategies across Norwegian and English corpora. Thus, the first phase of the project is somewhat experimental. The aim of this phase is to develop a methodological approach that could be placed towards the corpus-driven, bottom-up end of the continuum (described in section 2.2.1) and that would belong to the category of function-to-form approaches (described in section 2.2.5). More specifically, the aim is to find a “robust workaround” (O'Keeffe et al., 2020, p. 9) to access pragmatic functions in corpora which are not tagged for such functions (described in section 2.2.5). The approach involves the use of a probe (see section
5.1.1 for a definition) to access various forms of hedging strategies in the corpora and is described in more detail in chapter 5.

O'Keeffe (2018) outlines various previous approaches which have been classified as function-to-form, but concludes that there is still a need to study how pragmatic phenomena can best be investigated using corpora. The approach suggested for this study builds on elements from several of the function-to-form approaches described in section 2.2.5, for example, the manual sifting procedure applied by Deutschmann (2003) in his study of apologies. In order to arrive at a suitable probe, I have had to sift through the uses of the probe to make sure it adheres to the criteria outlined for the probe. Similarly, I have had to sift through all expressions in the context of the probe to retrieve hedging strategies. (See further chapter 5 for a description of these processes.) My study also shares some features with the approach of Taavitsainen and Jucker (2008) (described in section 2.2.5). Taavitsainen and Jucker (2008) retrieved compliments in corpora by searching for positive evaluative adjectives, examining both the forms they searched for and their contexts. The adjectives used as a starting point in Taavitsainen and Jucker’s (2008) study did not function solely as probes, as they were the subject of study as well, but the examination of the context in which they occurred is parallel to the approach which will be applied here. Inspired by these approaches and the need to explore more ways of studying pragmatic phenomena using corpora (O'Keeffe, 2018, p. 588), I propose a methodology using probes to retrieve a particular type of context in the corpora where hedging strategies are likely to occur. The approach (explained in more detail in section 5.1) is an attempt to bring together central concepts of corpus linguistics and pragmatics, i.e. combining horizontal and vertical reading, qualitative and quantitative approaches, etc. (see section 2.2.5). The process of identifying a probe is extensively described in section 5.1 and article 1.

The choice of probe was originally based on the notion presented by Brown and Levinson (1987) (described in section 2.1.1) that hedging strategies may be used as a remedial tool in face-threatening situations. Thus, the probe was chosen based on its ability to introduce face-threatening situations and the idea was that hedging strategies could be retrieved from the context following the probe. However, as discussed in section 5.1, choosing a probe based solely on criteria related to politeness was not successful, and other motivational factors for hedging in informal spoken conversations had to be considered as well. The discussion of the criteria for the choice of probe and the results from the comparison of different probes are presented in section 5.1.5.
3.2 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3)

The second phase of the project involves the analysis of data from four monolingual corpora, i.e. three corpora of informal spoken Norwegian and one corpus of informal spoken English. The corpora are described in more detail in chapter 4. The aim of this part of the project is to compare the use of hedging strategies across the two languages with respect to type of hedging strategies used, while simultaneously reviewing existing classificatory frameworks for hedging strategies described in section 2.1.1, and with respect to the forms used in each language, i.e. whether speakers use pragmatic markers, vague language, etc. The studies in the second phase of the project (see article 2 and 3) may be categorised as cross-cultural as defined in section 2.2.4.

The theoretical background for the study of hedging strategies involves a timeline for the concept of hedging which leads up to its present definition, as well as the classificatory frameworks for hedging strategies described in section 2.1.1 and the frameworks for mitigation presented in section 2.1.2. Classificatory frameworks have been described as “necessary heuristic devices” (Kaltenböck et al., 2010, p. 7) and are indeed useful in the description and understanding of hedging strategies. However, existing frameworks have also received criticism, particularly for the difficulties of distinguishing between the different sub-categories in practice (see section 2.1.1). Another challenge is that the majority of existing classificatory frameworks are either based on a narrow understanding of hedging, such as that proposed by Lakoff (1972), or are developed based on either spoken or written discourse from one genre in one language. This in turn challenges their general applicability as hedging often is realised differently in different languages (Holmes, 1982) and in different genres. In article 2, the present study seeks to investigate to what extent the existing classificatory frameworks can be used to describe hedging strategies identified according to the wide definition proposed in section 2.1 in both Norwegian and English. The hedging strategies retrieved through the methodological approach developed in phase 1 are described using a new combination of existing sub-categories from the frameworks presented in section 2.1.1 and the frameworks for mitigation strategies described in section 2.1.2. The results of the study are reported in article 2.

The comparison of realisations of hedging strategies across Norwegian and English in article 3 builds on research described in sections 2.1.1 and 2.1.2. As presented in these sections, previous studies of hedging in Norwegian generally overlap with studies of pragmatic markers (defined in section 2.1.2), often with the purpose of adding to the semantic descriptions of a particular marker (Unger, 2018), but
to this date, to my knowledge, there are no bottom-up studies of hedging strategies in spoken Norwegian. By applying the methodological approach described in section 5.1, a great number of different forms realising the hedging function were retrieved from the corpora. In order to determine whether the strategies were used to express hedging, additional secondary literature was consulted. This secondary literature is outlined in appendices 1 and 2. Literature on pragmatic markers, vague language, tag questions, general extenders was used to create a guide for analysis (appendices 1 and 2). The guide was used to identify and describe hedging strategies in the Norwegian and English data. For example, if an expression seemed to express hedging in the corpus data, the guide of secondary literature was consulted to find support for this claim. The analytical procedure is outlined in section 5.2, and the results of the study are reported in article 3.

3.3 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4)

The third phase of the project involves a study of hedging strategies by Norwegian learners of English using data from the free discussion part of the LINDSEI-no corpus and the comparable LOCNEC corpus (described in section 4.5). This part of the study builds on the results of the contrastive study in phase 2 identifying various forms of hedging strategies in the corpora. The 10 most frequent hedging expressions from the English material were used as the point of departure for the study of hedging strategies in the English of Norwegian learners. The findings from the learner corpus were compared to data from the LOCNEC corpus. Methodologically, the study applied an approach which could be placed towards the form-to-function end of the scale (see section 2.2.5 and section 5.3) As described in section 2.2.5, form-to-function approaches generally start with one or more forms and investigate their functions in various texts. The approach applied in phase 3 is in line with the approach Ädel and Reppen (2008) refer to as sifting, by which the data is sifted through manually and irrelevant cases are discarded. However, as the forms used in this study derive from a bottom-up study of hedging strategies, the approach shares some features with the approach applied by Schauer and Adolphs (2006) (described in section 2.2.5), who used results from a DCT as a point of departure for a corpus study, which according to O’Keeffe (2018) belongs to the function-to-form category of approaches.

Theoretically, the study also builds on previous research within the fields of learner corpus research and interlanguage pragmatics (described in sections 2.2.2 and 2.2.4). As stated in section 2.2.2, there is limited research on hedging in spoken English by Norwegian learners, therefore there is a need to explore this topic further. Furthermore, previous contrastive studies within the field of LCR have been
criticised for being too focused on frequency as the only statistical measure worth comparing and the other measures, such as dispersion, should also be applied (Paquot & Plonsky, 2017). Therefore, this study compares both frequencies of the individual expressions and inter- and intraspeaker variation. The procedure is outlined in more detail in section 5.3, and the results of the study are reported in article 4.
This chapter presents the corpora used in the studies in each phase of the project and discusses some aspects of using spoken corpora in the study of hedging strategies. Like chapter 3, this chapter is structured according to the different phases of the project (see figure 1-1 in section 1.4) and describes the corpora in more detail with respect to their compilation, principles of transcription and the samples used in the respective studies. First, however, some general remarks on the study of spoken data are made in section 4.1 and the choice of corpora reflected on in section 4.2. In section 4.3, the English-Norwegian Parallel Corpus (ENPC) is presented. The ENPC was used to find a suitable probe for the contrastive part of the study. The contrastive English-Norwegian study builds on data from four corpora of informal spoken conversations (three Norwegian and one British English corpus) presented in section 4.4. The Norwegian corpora are the Norwegian Speech Corpus (NoTa), the BigBrother Corpus (BB) and the Norwegian part of the Nordic Dialect Corpus (NDC). The English corpus is the Spoken British National Corpus 2014 (BNC2014). The study of learner language builds on data from a corpus of English spoken by Norwegian advanced learners (LINDSEI-no), a sub-corpus of the Louvain International Database of Spoken English Interlanguage (LINDSEI) and a comparable corpus of English spoken by native speakers, the Louvain Corpus of Native English Conversation (LOCNEC) presented in section 4.5.

4.1 Why study spoken corpora?

Hedging is a frequent phenomenon in both spoken and written discourse and consequently there have been several studies of the use of hedging strategies in both modes. The studies by Prince et al. (1982) and Holmes (1990) are two examples of studies of hedging in spoken discourse. The studies by Salager-Meyer (1994), Hyland (1996, 1998) and Vold (2006) are examples of studies of hedging in written discourse. The studies which are a part of this thesis investigate hedging exclusively in spoken discourse, mainly for three reasons. First, I was interested in how hedging is used as a tool to maintain a good relationship between interlocutors in informal conversations. I would argue that in many situations, maintaining a good relationship with the interlocutor is more important, or at least more urgent, in speech than in writing, when the addressee is actually present. Second, I wanted to investigate hedging as an aspect of pragmatic competence in spoken conversations and whether Norwegian learners of English are able to adjust their hedging behaviour to accommodate the practices of native
speakers of English, or if they translate their Norwegian strategies, favour certain hedging expressions or simply refrain from hedging altogether. Third, in linguistics in general there is a “written language bias” (Linell, 2005, 2019) (see section 2.2.1). Therefore, research based on informal speech is also much more in demand in the research community (Love et al., 2017, p. 325).

4.2 The issue of comparability and the choice of corpora

In corpus-based cross-linguistic studies, addressing the issue of comparability is crucial, both with respect to the phenomena compared and the data from which the phenomena are retrieved. Data comparability is a form of tertium comparationis. (See section 5.1.2 for other forms of tertium comparationis.) Tertium comparationis can be explained as a common reference against which similarities and differences can be identified (Lefer & Cartoni, 2011, p. 96). In parallel (translation) corpora, there is an in-built tertium comparationis provided by a translation relation (Ebeling & Ebeling, 2013, p. 17; Hasselgård, Forthcoming, p. 7). In comparable corpora, on the other hand, the tertium comparationis is related to text comparability, i.e. the tertium comparationis supposes that the corpora are similar in terms of genre, mode and time of compilation (Dupont, 2019, p. 14).

Another aspect of comparability is the degree to which the annotation schemes in the corpora compared is similar (Johansson, 2007). Johansson (2007, p. 306) problematises this and asks rhetorically “if corpora are annotated independently for each language, to what extent is the analysis comparable?” Text comparability is often supplemented by a perceived similarity or equivalence of the linguistic items compared (Hasselgård, Forthcoming, p. 7). The issue of data comparability in this study is particularly relevant to phase 2 of the project, the contrastive study of hedging strategies in Norwegian and English informal conversations. Whereas the corpora used in phase 3 of the project, LOCNEC and LINDSEI-no, are compiled according to the same principles for the same purposes and thus are largely comparable, the corpora used in the study of hedging strategies in Norwegian and English are compiled according to somewhat different principles and differ with respect to e.g. size and period and method of compilation. The use of natural language corpora for contrastive research has been discouraged because of the challenges of finding suitable corpora which can be used for language comparison (Aijmer, 2020, p. 4). However, such corpora are also particularly suitable for analysing pragmatic phenomena in spoken discourse where it is difficult to find translations. Although, the corpora used in the contrastive study here differ in some respects, they are comparable pertaining to genre in the sense that they all contain informal conversations between friends, family members, acquaintances and strangers (see further section 4.4).
The choice of corpora in phase 2 is a result of a compromise between the ideal and the constraints of what is realistically possible. The ideal material for a cross-linguistic study of hedging strategies in Norwegian and English spoken conversations would have been two large and completely comparable corpora of spoken conversations which would have been tagged for pragmatic functions such as hedging, making it possible to search for the function of hedging and compare the realisations in both languages. Unfortunately, no such pair of corpora is available, and it would have been too time-consuming and costly to compile such corpora within the scope of this project. Thus the study is based on existing corpora which are comparable to a certain degree.

In the following sections, the corpora are presented (with respect to manner of compilation, transcription procedures and sociolinguistic characteristics of the speakers), and their degree of comparability is discussed.

**4.3 Phase 1: Developing a method and identifying a probe (article 1)**

The first phase of the project was devoted to developing a procedure which would make it possible to study hedging strategies from a bottom-up, function-to-form perspective (see section 5.1.1 for an explanation of how bottom-up is understood here). The first step in this process was to find a probe which would function in the same manner in Norwegian and English to retrieve comparable hedging strategies from the two languages (see further section 3.1 and 5.1). The English-Norwegian Parallel Corpus (ENPC) was used to test whether the probe was used in the same way in English and Norwegian. This was a prerequisite for the use of the probe in the contrastive study in order to make the hedging strategies that would be compared as comparable as possible. The ENPC comprises written texts only, but has been tagged for speech in fictional texts, i.e. dialogues in novels, etc.

*The English-Norwegian Parallel Corpus (ENPC)*

The ENPC is a bidirectional parallel corpus (see section 2.2.1) which consists of written English and Norwegian original texts and their translations. The corpus has a fiction and a non-fiction part of which the fiction part consists of approx. 30 text extracts and the non-fiction part of approx. 20 text extracts.

The corpus was compiled in 1997 as a part of a research project at the Department of British and American Studies at the University of Oslo, and is one of several corpora in the Oslo Multilingual Corpus (OMC) (Johansson, 2007; Johansson, Ebeling, & Oksefjell, 2002). The OMC is a collection of several parallel corpora in various languages compiled in a similar way as the ENPC. Figure 4-1
presents the structure of the ENPC. For the purpose of this study, only the fictional texts and their translations were used.

![Diagram of the English-Norwegian Parallel Corpus](image)

*Figure 4-1 The structure of the English-Norwegian Parallel Corpus (Johansson et al., 2002).*

Of the 30 Norwegian original texts in the fictional part of the corpus, seven are children’s fiction, four detective fiction and 19 general fiction. Of the 30 English original texts, three are children’s fiction, eight detective fiction and 19 general fiction (Johansson et al., 2002, pp. 5-6). The text extracts are between 10,000 and 15,000 words long, and each s-unit (roughly comparable to a sentence) in the original texts is aligned with the corresponding s-unit in the translated texts, which makes it possible to compare parallel texts across more languages. To access the most spoken-like parts of the fictional texts, the filter “direct speech” was applied. This filter restricts the search to dialogues only, thus making it possible to see how the probe was used in spoken-like texts. Table 4-1 presents the number of words in the sample. The search procedure and results of this study are reported in section 5.1. See also article 1.
Table 4-1 Number of words in the direct speech samples.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>No. of words in sample(^8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENPC – English originals</td>
<td>91,995</td>
</tr>
<tr>
<td>ENPC – Norwegian originals</td>
<td>76,954</td>
</tr>
</tbody>
</table>

4.4 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3)

For the purpose of exploring hedging strategies in spoken Norwegian conversations, the conversational parts of the Norwegian Speech Corpus (NoTa) and the Nordic Dialect Corpus (NDC) were used, in addition to the whole of the BigBrother Corpus (BB). The conversations in these corpora were recorded between 2001 and 2012, thus they only partly overlap temporally with the recordings of the conversations in the BNC2014. The sizes of the corpora are also different. The Norwegian corpora (when only the conversational parts are included) comprise a total of approximately 2.3 million words, whereas the BNC2014 comprises a total of 11.4 million words. All corpora have an approximately equal distribution between the genders of the speakers. The age ranges of the speakers in the corpora are also rather comparable, with the exception of the BB corpus, which has a more limited age span (22–36). However, the majority of speakers in the BNC2014 (250 of 671) were between the ages of 19 and 29, so this may partly strengthen the representation of this age group in the Norwegian data and make the data more comparable with respect to age of the speakers. Table 4-2 summarises the key figures of the corpora. Each corpus is described in more detail below.

\(^8\) Signe O. Ebeling, personal communication.
Table 4-2 Key figures in the Norwegian and English corpora.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Year of recording</th>
<th>No. of words in sample</th>
<th>Age of speakers</th>
<th>Gender distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoTa</td>
<td>2005-2006</td>
<td>674,596</td>
<td>15-85</td>
<td>50% Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50% Male</td>
</tr>
<tr>
<td>NDC</td>
<td>2006-2012</td>
<td>1,199,651</td>
<td>15-88</td>
<td>51% Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>49% Male</td>
</tr>
<tr>
<td>BB</td>
<td>2001</td>
<td>440,354</td>
<td>22-36</td>
<td>50% Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50% Male</td>
</tr>
<tr>
<td>BNC2014</td>
<td>2012-2016</td>
<td>11,422,617</td>
<td>9-91</td>
<td>54% Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46% Male</td>
</tr>
</tbody>
</table>

The Norwegian Speech Corpus (NoTa)

NoTa is a 957,000-word monolingual corpus consisting of interviews and spontaneous conversations with 166 speakers from the Oslo-area recorded between 2004 and 2006. The corpus was compiled with funding from the Norwegian Research Council and the purpose of the corpus was to facilitate research within various sub-disciplines of linguistics, e.g. semantic descriptions of words, phraseology, sociolinguistic factors, comparative and contrastive studies, etc. (Johannessen & Hagen, 2008, p. 10). The corpus was accessed through the search interface, Glossa, (described in more detail below), which enabled me to restrict the search to the conversations in the corpus. The conversational part of the corpus contains 674,596 words. The conversations were mainly recorded at the speakers’ workplaces, in schools or in senior centres. Some of the recordings were also made in the speakers’ homes. In all recording situations, the facilities were made as comfortable as possible with some light snacks etc. to encourage an informal and relaxed atmosphere. Each conversation lasted for about 30 minutes and the speakers were encouraged to talk about whatever they wanted but to avoid sensitive topics such as third persons, religious and political preferences, etc. mainly due to privacy concerns. Both sound and video recordings were made of the conversations, and both are publicly available through the Glossa interface.

9 In the representative collection sample, which is used in this study, the number of female and male speakers are identical. In the corpus as a whole, 49% of the speakers are female and 51% are male. [http://www.tekstlab.uio.no/nota/oslo/index.html]

Metadata for the speakers such as gender, age, education and place of origin, as well as the relationship between the speakers, i.e. family, friends, strangers or acquaintances is available in addition to information about the recordings, e.g. place of recording, etc. The recordings were transcribed orthographically as closely as possible to what is actually said within the rules of the standard written language (Hagen, 2008, p. 3).12 The conversations are divided into turns which contain one or more utterances by one speaker. (See chapter 5.2.1 for a description of turns.) Extra-linguistic information about events that influence the speech, i.e. laughing, coughing, clearing the throat, sighing, etc., is indicated in brackets in the transcription. Such noises are only transcribed when they are thought to have consequences for the conversation (Hagen, 2008, p. 21). Interrupted utterances are marked with three dots in brackets, (...), and pauses are marked with one or more hash symbols, #, depending on the length of the pause (Hagen, 2008, p. 23). Incomprehensible passages are transcribed as [uforståelig] ‘[incomprehensible]’ (Hagen, 2008, p. 24), see example (4-1). Table 4-3 below summarises the most common transcription symbols used in the corpora.

(4-1)
A: ja det gjør e hun læreren vi har i matte og naturfag # hun blir rimelig sur i blant men noen så er det så kult å få dem til å bli skikkelig sure altså _uforståelig_ er til å grine av jo

‘A: yes our math and science teacher does that # she becomes pretty angry sometimes but some so it is so funny to make them really angry _incomprehensible_ makes you wanna cry’

NoTa 030><who_avfile 029-030 no. 997

Nordic Dialect Corpus (NDC)

The Nordic Dialect Corpus is a 2.79 million word multilingual corpus of Norwegian, Swedish, Danish, Icelandic, Faroese and Övdalian dialects initiated by the Scandinavian Dialect Syntax Network (ScanDiaSyn). It consists of spontaneous speech and interviews with speakers of various North Germanic dialects, and the primary purpose of the corpus was to study the Nordic dialects, mainly focusing on syntactic variation across the Scandinavian dialect continuum (Johannessen, Vangsnes, Priestley, & Hagen, 2014, p. 69). Therefore, the corpus is also equipped with a map function to pinpoint dialectal features geographically. Although the corpus was compiled with the purpose of studying syntax, it is also a source of natural language for general linguistic research.

---

12 The Bokmål standard was used as it is closer to Oslo speech than the alternative Nynorsk. For the Bokmål/Nynorsk distinction, see https://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk/
The Norwegian part of the corpus consists of approximately 2.2 million words produced by 564 speakers from 163 places all over Norway (Johannessen et al., 2014, p. 70). As with the NoTa corpus, only the conversational part of the corpus is used in this study. The recordings of the conversations were made in the speakers’ local environment and an effort was made to create comfortable and informal surroundings for the speakers. The speakers were asked to converse freely for about 30 minutes, but, as with the speakers in the NoTa corpus, they were asked to avoid sensitive topics. The conversations were transcribed orthographically based on the transcription guidelines of the NoTa corpus. Thus the principles of orthographic transcription are similar to those of NoTa. The conversations are also linked to video recordings. The Norwegian part of the corpus has also been transcribed phonetically in order to facilitate studies of dialectal variation. As with NoTa, metadata such as age, gender, place of origin, year of recording and relationship between the interlocutors has been registered.

*The BigBrother corpus (BB)*

The BigBrother corpus is a collection of transcribed sound and video recordings from the first Norwegian season of the reality show BigBrother which aired in Norway in 2001. The transcription and annotation of this corpus was completed in 2007. BigBrother was a reality show in which 12 contestants between 23 and 36 years old were isolated in a house for approximately 100 days. The contestants were monitored with cameras in every room 24 hours a day, and the recordings were broadcast live. In addition, an edited episode was sent once a day showing the most important events of the last 24 hours. Each week a contestant had to leave the competition based on viewer votes, thus not all contestants contributed equally to the corpus. The corpus contains transcriptions of nearly all 100 episodes of the show and contains nearly 440,300 words. The corpus only consists of the conversations between the participants, thus only unscripted informal speech. The data in the corpus is unique in the sense that it was not originally compiled for research purposes, and the subjects are not aware that their language is recorded for language research purposes, as would be the case with many other spoken corpus recordings. There were no restrictions on the topics the contestants were allowed to talk about, but because of the unusual situation they were in, living in a limited space with no contact with the world outside and competing with each other, the corpus typically contains episodes of bickering, discussions and expressions of strong emotions, which are unusual in most other language corpora. The orthographic transcription of the corpus is built on the same standard used in the NoTa corpus.
All of the Norwegian corpora were accessed through the Glossa interface.\textsuperscript{13} Glossa is a corpus search and results visualisation system which is open source and handles parallel and monolingual, spoken and written corpora. Glossa facilitates searches for linguistic or non-linguistic features and offers three levels of complexity of the search interface, “ranging from a simple Google-like search box for simple word or phrase queries, via an extended view that allows complex, grammatical queries, to a CQP query view that allows the user to specify the CQP query expression directly” (Nøklestad, Hagen, Johannessen, Kosek, & Priestley, 2017, p. 252). The Norwegian data was extracted according to the procedure described in 5.2.

\textit{Spoken British National Corpus 2014 (spoken BNC2014)}

The Spoken British National Corpus 2014 is an 11.5-million-word monolingual corpus of conversations between 668 speakers of British English recorded between 2010 and 2016. The Spoken BNC2014 is a result of a collaborative project between the ESRC Centre for Corpus Approaches to Social Science (CASS) at Lancaster University and Cambridge University Press and was made publicly available in September 2017. The data consists of spoken text which occurred only in informal contexts (Love et al., 2017, p. 324), and it is the speakers themselves who have made the recordings. The speakers used their own equipment and recorded conversations in their natural environments. In the collection of data, the corpus compilers made use of public participation in scientific research (Love et al., 2017, p. 328), meaning that anyone who was interested in participating was asked to register on a website and was contacted with further instructions about how to participate. The contributors gathered the requested metadata and informed consent from the speakers who were recorded. The corpus compilers applied an “opportunistic approach” to the data collection, meaning that they accepted the data that was made available, and when any holes in the demographic sampling were identified, they targeted that specific group of people (Love et al., 2017, p. 327). Still, the corpus ended up with a dearth of speakers from Scotland, Wales and Northern Ireland, whereas speakers from England were overrepresented. The compilers have accepted this dearth, as there exist other corpora of these varieties (Love et al., 2017, p. 327). The design and compilation process is described in more detail in Love et al. (2017).

The 1,251 recordings are orthographically transcribed. The starting point for the transcription was to employ a standard orthographic transcription (Love et al., 2017, p. 334) to facilitate quantitative studies of “morphology, lexis, syntax, pragmatics, etc.” (Atkins, Clear, & Ostler, 1992, p. 10). The data was

\textsuperscript{13} https://tekstlab.uio.no/glossa2/
transcribed according to a transcription scheme developed for the project. The development of the transcription scheme and the transcription process are outlined in Love et al. (2017), and only some features are described here. Speech phenomena that require a high degree of transcriber inference, e.g. false starts, non-verbal signals etc. have been normalised or disregarded to avoid having to spend too much time on ensuring consistency across transcribers, etc. Sounds such as *ah* and *um* have been transcribed according to their sound form and not their possible pragmatic function. Transcribers were instructed not to use most punctuation marks (full stops, commas, semicolons, exclamation marks); however question marks were used to mark questions ("The British National Corpus 2014: User manual and reference guide", 2018, p. 37). Non-linguistic vocalisations, such as coughing, laughing, etc. were recorded within [square brackets] ("The British National Corpus 2014: User manual and reference guide", 2018, p. 40). Table 4-3 summarises the transcription symbols used in the Norwegian and English corpora. Unlike the Norwegian spoken corpora described above, the data has not been transcribed phonetically and the sound files have not been made available.

The BNC2014 was accessed via Lancaster University’s CQPweb server (Hardie, 2012). The CQPweb is a web-based corpus analysis system through which a variety of corpora are accessible. When the chosen corpus is selected the researcher can perform a simple query or choose another menu option, e.g. restricted query and frequency lists. The English data was extracted according to the procedure described in 5.2.
Table 4-3 Overview of relevant transcription symbols.

<table>
<thead>
<tr>
<th>Transcription symbol</th>
<th>Meaning</th>
<th>Found in corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>short pause</td>
<td>NoTa, NDC, BB</td>
</tr>
<tr>
<td>(.)</td>
<td>short pause</td>
<td>BNC2014</td>
</tr>
<tr>
<td>…</td>
<td>pause</td>
<td>LINDSEI-no, LOCNEC</td>
</tr>
<tr>
<td>(eh)</td>
<td>filled pause</td>
<td>LINDSEI-no, LOCNEC</td>
</tr>
<tr>
<td>*</td>
<td>two or more speakers speaking at the same time</td>
<td>NoTa, NDC, BB</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td>two or more speakers speaking at the same time</td>
<td>BNC2014</td>
</tr>
<tr>
<td>(coughing)</td>
<td>vocalisations, such as coughing, laughing, etc.</td>
<td>NoTa, NDC, BB</td>
</tr>
<tr>
<td>[coughing]</td>
<td>vocalisations, such as coughing, laughing, etc.</td>
<td>BNC2014, LINDSEI-no, LOCNEC</td>
</tr>
<tr>
<td><em>uninterpretable</em></td>
<td>comments by the transcriber</td>
<td>NoTa, NDC, BB</td>
</tr>
<tr>
<td>(...)</td>
<td>Interrupted utterances</td>
<td>NoTa, NDC, BB</td>
</tr>
</tbody>
</table>

4.5 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4)

The intervarietal study of Norwegian learners builds on data from two comparable corpora of English by Norwegian learners and English by native speakers of English, LINDSEI-no and LOCNEC respectively.

Louvain International Database of Spoken English Interlanguage (LINDSEI)
The LINDSEI corpus is one of the most well-known corpora of learner English (Romero-Trillo, 2018, p. 115). The purpose of the LINDSEI corpus was to create a spoken counterpart to the International Corpus of Learner English (ICLE) which contains argumentative essays written by higher intermediate and advanced learners of English with different L1 backgrounds. LINDSEI was launched in 1995 by the Centre for English Corpus Linguistics at the Université catholique de Louvain and consists of spoken English produced by advanced learners with different L1 backgrounds. The learners are university undergraduates and learners of English as a Foreign Language (EFL) with an advanced
proficiency level (Gilquin et al., 2010, p. 7). The corpus consists of informal interviews with the non-native speakers and there are 50 interviews for each L1 group. Each interview consists of three speech tasks: “a warming-up activity, in which learners were given a few minutes to talk about one of three set topics, a free informal discussion which was conceived as the main part of the interview, and a picture description” (Gilquin et al., 2010, p. 3). The corpus therefore consists of both interviewer and interviewee speech, referred to as ‘A turns’ and ‘B turns’, respectively, the B turns representing learner language (Gilquin et al., 2010, p. 12). The tasks last for about 15 minutes and yield approximately 2,000 words from each learner, which means that each sub-corpus contains about 100,000 words of learner language (De Cock, 2004, p. 227). The interviews are also linked to metadata about the learner, the interviewer and the interview itself, which makes it possible to study the possible influence of certain factors on learner language. The corpus invites studies of various aspects of learner English such as lexis, syntax, phraseology, discourse or pragmatics.

LINDSEI is transcribed orthographically according to the transcription guidelines outlined in (Gilquin et al., 2010, pp. 13-18). In general, British spelling conventions are followed in the transcription. No punctuation marks are used to mark clause boundaries. Pauses are indicated with dots. The number of dots indicate the length of the pause. Filled pauses are indicated in brackets, e.g. (eh), (erm). Prosodic information, such as laughter, whispering voice, etc., is indicated in angle brackets at the start and end of the relevant stretch of text (see table 4-3 above).

The Norwegian part of the LINDSEI corpus was not a part of the original corpus launched in 1995, but has been designed and compiled in the same manner as the other sub-corpora, and thus is fully comparable to them.14 LINDSEI-no was compiled by a team of researchers from Hedmark University College of Applied Sciences (now Inland Norway University of Applied Sciences) and will be published in the second edition of LINDSEI (Aas & Nacey, 2019).15 LINDSEI-no contains 122,956 words of learner language. LINDSEI-no contains 50 interviews, made up of three tasks: set topic, free discussion and picture description, with advanced learners of English at Hedmark University College. The average age of the learners was 26, and there were 35 females and 15 males participating. Only the free discussion task has been used for this study, mainly because it is the part of the corpus which is most similar to the informal conversations in the corpora of native Norwegian and native English

---

15 The material has kindly been made available by the LINDSEI-no team at the Inland Norway University of Applied Sciences.
described in section 4.4. Table 4-4 presents an overview of the total number of words in the corpus and the number of words in the sample.

The learners in the LINDSEI corpus are labelled “advanced learners of English”; however, the notion of advanced learners is based on external rather than internal criteria, i.e. that they are all in their third or fourth year of studying English at university level. This means that there are differences in the proficiency levels across and within sub-corpora (Gilquin et al., 2010, p. 10). In her study of some discourse markers in the LINDSEI-no corpus, Sandal (2016) uses the label “advanced learners” when referring to the proficiency level of the Norwegian learners. This label will also be used here.

**The Louvain Corpus of Native English Conversation (LOCNEC)**

The LOCNEC corpus contains native speaker English and was compiled as a comparable corpus to LINDSEI. The purpose of the LOCNEC corpus was thus to enable the comparison of English interlanguage and native language. Like LINDSEI, the LOCNEC corpus contains informal interviews consisting of three speech tasks, i.e. set-topic discussion, free discussion and picture description. The interviewees were all British students, mostly undergraduates, but also some postgraduates, from Lancaster University, all native speakers of English. Their age ranged from 18 to 30, and there are 30 females and 20 males (Gilquin et al., 2010, p. 65). The corpus consists of 50 informal interviews totaling 170,347 words. Table 4-4 presents the total number of words of interviewee language in the corpus and the number of words in the sample. LOCNEC is compiled and transcribed according to the same principles as LINDSEI.

**Table 4-4 Total number of words in LINDSEI-no and LOCNEC.**

<table>
<thead>
<tr>
<th></th>
<th>Total no. of words of learner language</th>
<th>Total no. of words in free discussion (tokens)</th>
<th>Total no. of speakers (learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINDSEI-NO</td>
<td>122,956</td>
<td>43,431</td>
<td>50</td>
</tr>
<tr>
<td>LOCNEC</td>
<td>126,666</td>
<td>70,957</td>
<td>50</td>
</tr>
</tbody>
</table>

LINDSEI-no was accessed via the license provided by the University of Oslo and a corpus file of LOCNEC tagged for speech task was made accessible upon request from the Université catholique de Louvain.16

---

16 I am very grateful to Professor Sylvie De Cock at Université catholique de Louvain for sharing a version of the LOCNEC corpus tagged for speech task.
Strictly speaking, LINDSEI and LOCNEC do not comply with all definitions of a corpus (cf. Sinclair (1996)). One of the main corpus criteria is that the data is authentic. Sinclair (1996, p. 7) defines authentic data as material “gathered from the genuine communication of people going about their normal business”. In second language acquisition terms, authentic data can be defined as the result of learners trying “to use their L2 knowledge in communication” (Ellis, 1994, p. 671). In LINDSEI and LOCNEC, at least some of the language is elicited, rather than produced in a natural communicative setting. For example, the picture description task is not a natural communicative setting, but one where data is collected specifically for research purposes (Gilquin et al., 2010, p. 6). This is also reflected in the fact that the collections are referred to as databases rather than corpora. The free discussion task is closer to natural and authentic language use, although interview settings, such as that of the free discussion task, have also been criticised for preventing the production of authentic data (Ellis & Barkhuizen, 2005). Nevertheless, the data in LINDSEI and LOCNEC meet the criteria for learner corpora in the wide sense (Gilquin et al., 2010, p. 6) and are therefore referred to and treated as corpora here.

LINDSEI and LOCNEC are considered comparable corpora; however, their comparability has also been challenged. Gablasova et al. (2017) state that corpus comparability cannot be defined based on design criteria alone, but that corpus comparability is a complex issue dependant on a variety of factors. To illustrate their point, Gablasova et al. (2017) use LINDSEI and LOCNEC as an example of corpora which are comparable on the surface level, but which nevertheless contain differences. In my study, I have accepted the comparability of the LINDSEI-no and the LOCNEC corpora at face value; however, in the intervarietal study based on these corpora, I have considered both inter- and intraspeaker variation, which help nuance the results of the comparison. “If interspeaker patterns within one corpus vary considerably, this may make it difficult to interpret any difference found between the corpora” (Gablasova et al., 2017, p. 138). Another way of ensuring the validity of the results is to replicate the research in other L2 corpora with similar data; however, to this date, the LINDSEI-no is the only corpus of spoken English by advanced Norwegian learners.
Chapter 5 Methodological considerations

This chapter outlines the methodological considerations made throughout the phases of the project. A central objective has been to develop an approach within the form-to-function–function-to-form spectrum which would make it possible to study hedging strategies from a bottom-up perspective. (See section 5.1.1 for an explanation of how the term bottom-up is used here.) A significant part of this chapter is devoted to describing and discussing the process of this development. The chapter is structured according to the three phases of the project. Section 5.1 and its subsections describe how the methodological approach for the study of hedging strategies in English and Norwegian was developed and how a suitable probe was identified and tested. Section 5.2 and its subsections outline how the approach described in section 5.1 was applied to the Norwegian and English data in the contrastive part of the study in the second phase of the project. Section 5.2 also presents and discusses the procedure followed in the studies and the analyses of the data. Finally, section 5.3 and its subsections describe the methodological approach applied to the intervarietal study of English by Norwegian learners in the third phase of the project. The study described in section 5.3 builds on the results of the contrastive study described in section 5.2 and article 3.

5.1 Phase 1: Developing a method and identifying a probe (article 1)

The objective of the first phase of the project was to develop an approach which shared some characteristics with function-to-form approaches as they are described in e.g. Aijmer and Rühlemann (2015); O'Keeffe (2018). This partly experimental process is described in the following sections and in article 1.

5.1.1 Developing a method

As discussed in chapter 2, corpus pragmatics brings together two rather different subfields of linguistics; corpus linguistics, which is characterised by form-to-function approaches, and pragmatics, which is characterised by function-to-form approaches (Flögk & Geluykens, 2015, p. 9). As corpora are developed with the purpose of accessing linguistic forms in databases, the majority of corpus pragmatic studies have also used a form-to-function approach starting from one or more pre-defined forms and exploring their pragmatic functions in the corpus (see section 2.2.5). Thus, corpus pragmatics has been accused of merely using traditional corpus linguistics techniques under the label of corpus
pragmatics (Weisser, 2018, p. 10), and a need for exploring new methodologies in the study of pragmatic phenomena has been expressed (O'Keeffe, 2018, p. 588).

The form-to-function approach has also been favoured in studies of hedging strategies, mainly because of their unruly nature and large repertoire of realisations, (e.g. Aijmer, 1984; Farr & O'Keeffe, 2002; Gries & David, 2007; Baumgarten & House, 2010; and Hasund et al., 2012). Such studies give important insights into various forms and how they function in spoken or written discourse, but by starting with a set of pre-defined forms, other potential realisations of the same pragmatic meaning remain unexplored. This is especially true for hedging strategies, which can take almost any linguistic form, and the forms used to express hedging in one context do not necessarily express hedging in other contexts. Hedging is essentially a pragmatic function and corpora are generally facilitated for searching for language forms. Therefore, restricting a study to conventional realisations of hedging strategies may not show the whole picture. Moreover, comparing hedging strategies across languages from a function-to-form perspective might shed light on similarities and differences in hedging behaviour in different language cultures and pave the way for new form-to-function studies of less commonly described hedging strategies.

As opposed to the majority of corpus pragmatic studies, the contrastive study in this project aims to approach hedging strategies from a bottom-up perspective. The term bottom-up is polysemous and is often associated with corpus-driven studies (see section 2.2.1 for a discussion of bottom-up). In the current study, however, bottom-up is used to signify not pre-determined. When applying a function-to-form approach, the pragmatic function is pre-determined, but the forms realising the function are not. Thus the study is not bottom-up in the corpus-driven sense, but in the sense that it is used here.

In light of the methodological traditions in the field of corpus pragmatics, described in section 2.2.5, the contrastive part of this project applies a methodological approach which can be placed towards the function-to-form end of the scale (see figure 2-5 in section 2.2.5). However, it also shares some common features with form-to-function approaches, as is the case with most corpus pragmatic methodologies (O'Keeffe, 2018, p. 599). The approach is based on an established method in corpus linguistics at large, but is innovative in the sense that it has, to my knowledge, never been applied in this way to the study of pragmatic phenomena. In the contrastive studies, a probe will be used as a marker of a particular type of speech situation in which hedging strategies are likely to occur and can be retrieved. A probe may be defined as an element, i.e. a word, a string of words, or a tag, used to find other elements which cannot easily be found in a corpus (Hunston, 2002, p. 66).
Although most corpus linguistic studies, in the same way as corpus pragmatic studies, tend to investigate the word or phrase they search for, a search can also be used “to find sets of words or expressions that cannot easily otherwise be called to mind. These searches are known as ‘probes’” (Hunston, 2002, p. 62). For example, *would* could be used as a probe to retrieve ways of expressing possible or hypothetical events (Hunston, 2002, p. 64). However, a probe does not need to be a lexical word or a phrase. Depending on the type of corpus and whether it is tagged or not, a probe may also be a single POS-tag, several tags in combination or a combination of words and tags. The example given in Hunston (2002, p. 62) is from a study of how men and women are evaluated, and the probe is a combination of lexical words and tags: *something/nothing + [adjective] + about/in + him/her*. By searching for the adjective tag in this particular setting in a tagged corpus, the results would provide a list of adjectives that describe how men and women are evaluated in this particular structure. An important point to underline is that the probe itself is not the form under investigation. If *would* is used as a probe to investigate ways of expressing hypothetical events, as exemplified by Hunston (2002), it is the contexts of *would* which are under scrutiny. *Would* is merely a way to access these contexts. The same logic may be applied to the use of a probe to retrieve hedging strategies in this study. A probe is used to access contexts in which various forms of hedging strategies may be retrieved. Thus, although a form is used as a starting point, it is not the functions of that form which will be investigated. Formulaically this can be expressed in the following way:

\[
\text{Form}_1 \rightarrow \text{Function} \rightarrow \text{Form}_2
\]

\[
[\text{Probe}] \rightarrow [\text{Hedging}] \rightarrow [\text{Realisations of the hedging function}]
\]

The probe is used to find contexts where forms used to express hedging occur.

In addition to placing the study along the form-to-function–function-to-form spectrum, it may also be fruitful to place it on the corpus-based–corpus-driven continuum as suggested by Huang (2017a). (See section 2.2.1 for a discussion of corpus-based and corpus-driven studies.) Here, viewing the data within pre-existing frameworks will be referred to as corpus-based, and observing how a phenomenon occurs and behaves in the data will be referred to as corpus-driven. Based on the understanding of corpus-based and corpus-driven presented in section 2.2.1, the studies presented in article 2 and 3 may be
described as partly corpus-based and partly corpus-driven. It is corpus-based in the sense that it builds on previous research on motivational factors for using hedging strategies in the selection of a probe as well as previous research on hedging in the identification of realisations of hedging strategies. (See chapter 2 and appendices 1 and 2 for references to literature describing potential realisations of hedging strategies). It is corpus-driven in the sense that the object of study, i.e. hedging strategies, is studied from a bottom-up perspective, i.e. no forms are pre-determined, but whichever forms of hedging strategies emerging from the data will be studied. In an attempt to bring the form-to-function–function-to-form and corpus-based–corpus-driven spectrums together, Figure 5-1 illustrates where the study (marked by X) is situated along the two spectrums.

![Figure 5-1 Placing the approach within the corpus-based–corpus-driven/form-to-function–function-to-form spectrum.](image)

### 5.1.2 Identifying a probe

Choosing a probe that successfully retrieves hedging strategies is challenging. Hedging strategies can take almost any form and occur in nearly any lexical construction, which gives nearly endless choices to search for. Hedging strategies can occur in initial, medial and final position, which makes them hard to retrieve using syntactic tags. Finally, as there is a limited number of corpora annotated for pragmatic functions, hedging strategies cannot be found using pragmatic tags. Therefore, a probe which is successful in retrieving a high number of hedging strategies must identified using other types of criteria.
The criteria proposed here are based on motivational factors for hedging in informal spoken conversations. Hedging strategies may be applied for a number of reasons depending on a number of factors, such as mode (spoken or written), genre, interspeaker relationship, etc. For example, in scientific research articles, hedging strategies are often used to express an appropriate level of uncertainty with respect to a scientific claim to avoid unnecessary revisions (Hyland, 1996, p. 440). In spoken conversations, however, politeness is often regarded as the primary motivation (Nikula, 1997), i.e. hedging strategies may be applied as a remedial device in face-threatening situations (Brown & Levinson, 1987, p. 145) (see section 2.1.1). This assumption that hedging strategies are frequently used for politeness reasons is maintained e.g. in Markkanen and Schröder (1997, p. 7), Schröder and Zimmer (1997, p. 249) and Aijmer (2002). Therefore, the assumption is that a probe in the form of a word or phrase potentially introducing a face-threatening statement or characteristic of a face-threatening situation could give access to hedging strategies in the immediate context of the probe. Another criterion that the probe needs to fulfil is that it must be comparable across the two languages studied in this project. In order to elicit comparable data, the probe must behave similarly in the two languages, i.e. have similar semantic and syntactic properties. Furthermore, the probe should occur frequently in oral conversations, so that a substantial amount of occurrences can be found in the corpora. The overview below summarises the selection criteria.

The probe should
- be a characteristic of a face-threatening situation in the two languages
- display similar syntactic behaviour in the two languages
- have more or less the same semantic prosody in the two languages
- occur frequently in spoken everyday conversations in the two languages

The process of finding a probe that fulfilled these criteria started by identifying speech situations that can be perceived as face-threatening in both languages and subsequently considering how those speech acts are realised in the two languages and whether the realisations corresponded. Then the probes were tested in the corpora described in section 4.3 in a small pilot study to see whether they co-occurred with hedging strategies. Based on the results of these investigations, a probe was chosen and further
tested in a larger pilot study (see article 1). In this second pilot study 150 random instances of the probe were selected in each of the two languages respectively. Each instance was investigated in more detail to see whether it introduced a potentially face-threatening speech situation and whether any hedging strategies co-occurred. In this way both vertical and horizontal reading was applied (Aijmer & Rühlemann, 2015, p. 12) (see section 2.2.5 for an explanation of vertical and horizontal reading).

The objective of this process was to establish a tertium comparationis against which the two languages could be compared. (See section 4.2 for a discussion of other types of tertium comparationis.) To ensure reliability of a cross-linguistic comparison, the units compared must be comparable, i.e. when a phenomenon is compared across two languages, the first thing that needs to be done “is to make sure that we are comparing like with like” (James, 1980, p. 169). However, this does not mean that the units being compared must be identical (what would then be the point of the comparison? (Ebeling & Ebeling, 2013, p. 18)), but rather that they share some central attributes, i.e. the aim is to find “those properties which the compared items share, but which are outside the scope of comparison itself” (Krzeszowski, 1990, p. 117).

When discussing the tertium comparationis in contrastive studies, a distinction is often made between a formal tertium comparationis and a tertium comparationis based on semantic equivalence (Dupont, 2019, p. 11). A formal tertium comparationis is related to the comparison of corresponding grammatical systems, whereas a tertium comparationis based on semantic equivalence is related to the semantic meaning of the units compared. To find a probe that ensured a solid tertium comparationis, a thorough investigation of the formal and functional properties of the probe needed to be performed. As pointed out by James (1980, pp. 169-170): “A comparison based on formal tertium comparationis is only possible when the two languages have grammatical categories in common” and the grammatical labels used actually refer to the same phenomena in the two languages. The purpose of the investigation of the use of the probe was to also establish a tertium comparationis based on semantic equivalence. Two different probes were compared and tested. Ideally, a variety of probes could have been tested and compared, but due to the limited time and scope of this project, the more successful of the two were chosen. The probes were measured against the criteria outlined above.

5.1.3 No/nei

The first probe investigated was the interjection no and the corresponding Norwegian nei. No/nei is typically used to express negation, refusal or rejection of some sort, a speech act which can in many situations be perceived as face-threatening (Brown & Levinson, 1987). For example, when something
is refused, the hearer’s expectation is not met, i.e. his/her freedom of action is impeded (Demirkol, 2016). Acts preceding refusals can threaten the hearer’s face by putting pressure on her/him to do something, and refusals in turn threaten the hearer’s face and his/her wants and freedom. Thus speakers may choose to modify their refusals to avoid damaging the relationship with their interlocutor. *No* and the corresponding Norwegian *nei* can be seen as a conventional realisation of a refusal. When refusing something, however, speakers have several options as to how they choose to realise the refusal. In a study of how Japanese learners of English refused requests, invitations, offers and suggestions, Beebe, Takahashi, and Uliss-Weltz (1990) studied the way in which non-native speakers of English performed the speech act of refusal compared to native speakers. The study resulted in a taxonomy of refusal strategies which encompasses two direct refusal strategies and 11 indirect refusal strategies. Beebe et al. (1990) present *no* as an example of a non-performative direct refusal strategy. Speaker may choose a more elaborate or indirect strategy, but if they choose to perform the refusal using a direct refusal strategy, they can use *no/nei*, or *no/nei* in combination with hedging strategies. Furthermore, the use of *no*, and potentially also *nei*, has a high frequency in spoken language (Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 444).

### 5.1.3.1 The structural and semantic properties of *no/nei*

Formally *no/nei* is often described and classified as words that do not form an integral part of the syntactic structure (Biber et al., 1999, p. 56). They are however, labelled somewhat differently in different grammars. The Norwegian *nei* is classified as a special subclass of interjections, i.e. response forms (‘svarord’), alongside words like *ja* (‘yes’) and *jo* (‘yes’, ‘well’) (Faarlund, Lie, & Vannebo, 1997, p. 968). Similarly, *no* is classified as a part of the group of non-clausal units of inserts (Biber et al., 1999, p. 1082). The category of inserts can be further divided into interjections, greetings/farewells, discourse markers, attention signals, response elicitors, response forms, hesitators, thanks, the politeness marker *please*, apologies and expletives (Biber et al., 1999, pp. 93-94), in which *no* belongs to the subcategory of response forms. The category of response forms typically include responses to questions (yes, no, and their variants), responses to directives (e.g. *okay*), and responses to assertions (e.g. backchannels such as *uh huh, mhm*) (Biber et al., 1999, p. 1089). Another label put on the English *no* is *reaction signal* (Quirk et al., 1985, p. 444). The description is, however, similar to that of response forms. As opposed to *nei*, *no* is polysemous. *No* can also be classified as a determiner expressing ‘not any’ or ‘quite the opposite of’ ("Concise Oxford English Dictionary," 2011) and an adverb in the meaning “not at all” used with the comparative.
Semantically, *no/nei* has a core semantic meaning expressing negation, disagreement or rejection; however, its meaning is closely connected to the communicative situation in which it occurs (Faarlund et al., 1997, p. 968). *No/nei* as a response form can mainly be said to perform two tasks, either to negate positive content in a previous statement or to confirm a preceding negative statement (Faarlund et al., 1997, p. 1168), but its meaning may also vary between contexts. *Nei* is also described as a pro-form (‘pro-ord’), i.e. a word whose reference is decided by the context or co-text (Faarlund et al., 1997, pp. 25-26). Response forms such as *ja* and *nei*, fall into the pro-form category as they can replace clauses, e.g. *Kommer du? Nei, (jeg kommer ikke)*. (‘Are you coming?’ ‘No, (I am not coming)’) (Faarlund et al., 1997, p. 26). On a syntactic level, these words are mainly categorised as interjections, due to the fact that they are not syntactically connected to the other elements of the clause. Syntactically, this can be seen in parallel to the description of *no* as a non-sentence (Quirk et al., 1985, p. 849). Under this heading two categories of non-sentences are presented, *formulae* and *interjections* (Quirk et al., 1985, pp. 852-853). Interjections, according to Quirk et al. (1985, p. 853), are “purely emotive words” such as *Eh, Aha* and *Wow*. Formulae are used for stereotyped communication situations, e.g. greetings, thanks, etc. One of the types of formulae is reaction signals where one either expresses accept/agreement or denial/disagreement (Quirk et al., 1985, p. 852).

*Nei* and *no* also have a variety of functions where their core meaning is bleached. *Nei* may initiate new statements without negating or confirming previous statements (Faarlund et al., 1997, p. 969). In such situations, *nei* does not convey propositional content itself and thus shares the function of many pragmatic markers (Svennevig, 2001a, p. 144). This use of *nei* as a pragmatic marker can occur after a preceding wh-question or simply initiate a new statement (Svennevig, 2001a, p. 143). The function is also referred to briefly in Faarlund et al. (1997, p. 970), who say that this type of *nei* can be affirmation or negation of an implicit question. When *nei* is used as a pragmatic marker, it functions as a hedge, reducing the epistemic commitment of the speaker to what is said (Svennevig, 2001a, p. 151). It may be used to signal real uncertainty on behalf of the speaker, but also merely as a marker of politeness, thus the hedging function of *nei* has two sub-functions; it can express uncertainty on behalf of the speaker or be a marker of politeness expressing “modesty and cautioness” showing respect for one’s interlocutor (Svennevig, 2001a, p. 152). *Nei* can also be a discourse marker functioning as a means to organise the conversational interaction (Svennevig, 2001a, p. 164).

Correspondingly, *no* can also function as a pragmatic marker, i.e. mainly as a discourse marker (Lee-Goldman, 2011). (See section 2.1.2 for a discussion of pragmatic markers.) *No* can function as a discourse marker in a similar way as *yeah* and be used as a device to change the topic of a conversation,
particularly from non-serious discourse to more serious discourse, referred to as the serious-no (Lee-Goldman, 2011, p. 2632). The misunderstanding-no is another subtype of the discourse marker no which is used to manage and mitigate misunderstanding and disagreement by rejecting the mistaken conception (Lee-Goldman, 2011, p. 2634). A third subtype of discourse marker no can be found in the third-position repair-no, which typically includes a particle such as no, an acceptance of the prior turn, a rejection component and the repair proper (Lee-Goldman, 2011, p. 2638). Finally, a fourth subtype is the turn-negotiation-no (Lee-Goldman, 2011, p. 2640). This type can be illustrated with an example typically heard in spoken conversations, for example when speakers are speaking at the same time “No, no, go ahead” (Lee-Goldman, 2011, p. 2640).

5.1.3.2 No/nei in the ENPC

In order to determine the degree of comparability and correspondence of nei and no, a small-scale study was performed using data from direct speech in fictional texts from the English-Norwegian Parallel Corpus (ENPC) (described in section 4.3). The purpose of the study was to determine whether nei and no in fact correspond to such a degree that it would make sense to compare them across the two languages and to assume that they would be descriptive of the same types of situations. The advantage of using a translation corpus such as the ENPC for this purpose is that it “provides the researcher with a ready-made tertium comparationis: the linguistic units to be compared are directly accessible through the relation of translation equivalence that unites them” (Dupont, 2019, p. 131) (emphasis in original).

As described in section 4.3, only the fictional dialogues in the corpus were used. The degree to which fictional dialogue represents actual dialogue will not be discussed here; however, the choice of using only the direct speech part of the fictional texts was built on the assumption that fictional dialogues at least strive to be representative of natural dialogues. I am aware of the challenges of making this assumption, as pointed out by Page (1994, p. 7) in his book on dialogue in fiction: “there is an inevitable gap [...] between speech, especially in informal situations, and even the most ‘realistic’ dialogue in a work of literature”. Unfortunately, since there are no parallel corpora of spoken Norwegian and English, spoken-like texts seem to be the best alternative for this purpose.

The search for no in direct speech in fictional texts gave 426 hits in the English original. Of these, 216 were no as an interjection. In total, 89% of the 216 interjections were translated with nei. The other 210 occurrences of no were typically the determiner use of no, no as an adverb or the pronoun use of no, (no one) meaning ‘no person’. The search for nei in the Norwegian original texts gave 180 occurrences, of which 145 were translated with no. This gave a mutual correspondence of no and nei of 85%.
degree of mutual correspondence shows “the frequency with which different (grammatical, semantic and lexical) expressions are translated into each other” (Altenberg, 1999, p. 254). Table 5-1 presents the filters used when searching in the ENPC, the total number of instances for each word, the number of these which were used in their refusal sense and how many of these which were translated into no/nei.

Table 5-1 Filters applied in the ENPC and the number of instances of no/nei.

<table>
<thead>
<tr>
<th>Filters applied in the ENPC</th>
<th>Total no. of instances</th>
<th>Total no. of instances per 1,000</th>
<th>No. of instances used as refusals</th>
<th>No. of refusals per 1,000</th>
<th>No. of hits translated into no/nei</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No</strong></td>
<td>ENPC fiction, English original, direct speech</td>
<td>426</td>
<td>4.63</td>
<td>216</td>
<td>2.34</td>
</tr>
<tr>
<td><strong>Nei</strong></td>
<td>ENPC fiction, Norwegian original, direct speech</td>
<td>180</td>
<td>2.33</td>
<td>180</td>
<td>2.33</td>
</tr>
</tbody>
</table>

5.1.3.3 A pilot study of no/nei in the BNC2014, NoTa, NDC and BB

After establishing a high degree of mutual correspondence, a pilot study was performed using the spoken corpora selected for the full-scale study: NoTa, NDC, BB and BNC2014 (see section 4.4 for a description of the corpora). The purpose of this pilot was to see whether there was a substantial number of occurrences of the probes, i.e. no/nei used in their refusing sense in the corpora and whether the probes actually co-occurred with hedging strategies. The filters that were applied are summarised in Table 5-2.
Table 5-2 Filters applied in the NDC, NoTa, BB and BNC2014 (pilot study).

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDC</td>
<td>Country: Norway</td>
</tr>
<tr>
<td></td>
<td>Category: Conversation between strangers, family, acquaintances, friends</td>
</tr>
<tr>
<td></td>
<td>Extended search: Segment initial, 50 random results</td>
</tr>
<tr>
<td>NoTa</td>
<td>Representative selection: yes</td>
</tr>
<tr>
<td></td>
<td>Genre: Conversation between strangers, family, acquaintances, friends</td>
</tr>
<tr>
<td></td>
<td>Extended search: Segment initial, 50 random results</td>
</tr>
<tr>
<td>BB</td>
<td>Extended search: Segment initial, 50 random results</td>
</tr>
<tr>
<td>BNC2014</td>
<td>Show in random order(^{17})</td>
</tr>
</tbody>
</table>

A total of 150 random occurrences of *no* and *nei* respectively were retrieved from the corpora. Only 116 of the total of 300 were *no/nei* in their refusing sense, thus only about 38.6% of the instances were relevant for further study. Of these, 38.7% of *no* and 32.9% of *nei* co-occurred with hedging strategies respectively. Only hedging strategies occurring to the right of the probes were considered, because the probe was regarded as introducing a face-threatening situation, and thus hedging strategies were assumed to follow in the immediate context. Typically, the occurrences of *no* and *nei* in the data were uses confirming a previous negative statement as in (5-1) and thus not used in a refusing sense. Furthermore, *nei* was also occasionally used as a discourse marker, as in example (5-2) where speaker A changes the topic with *nei*. In example (5-3) *nei* is used as a hedge. *Nei* is not a response word indicating a refusal of something, but speaker B is disagreeing with speaker A and starts her utterance with *nei* to reduce her commitment to the proposition. Such instances were also disregarded. Thus only 38.6% of the occurrences of *no/nei* were relevant, and of these, only about a third included hedging strategies. Example (5-4) illustrates *no* used in a refusing sense co-occurring with hedging strategies (in italics).

(5-1)\(^{18}\)

- A: ja _clears-throat_ det gör jeg gå i skog og mark det er i grunn det som er # det kjækkest
- ‘A: yes _clears-throat_ I do that walk in the woods and fields that is really what is # the best’
- B: mm # men på jakt det går ikke du

\(^{17}\)The complete tag file for the BNC 2014 was not available at the time of the pilot study, thus planned filters such as excluding *no* as a determiner, looking at segment initial *no* and different types of dialogue were not applicable.

\(^{18}\)All examples from the corpora are included as they are transcribed including transcription symbols indicating pauses, interruptions, simultaneous speech, etc. See table 4-3 in section 4.4 for an explanation of the transcription symbols.
In summary, the results of *no/nei* as a probe were disappointing. Using *no/nei* as a probe would in practice mean that only a small number of instances, about 30–40% of those investigated as indicated by the pilot study, would be relevant uses of *no/nei*. In turn, these would yield a limited number of hedging strategies, resulting in a scarce foundation for the study of such strategies and the approach would not have the intended methodological rigor (possible reasons for this will be discussed in section 5.1.5 below). Table 5-3 presents the numbers of *nei* and *no* in their refusal sense and the numbers of
these refusals followed by a hedge. Therefore, no/nei was discarded as a probe and a new probe investigated.

Table 5-3 Numbers of no/nei in its refusing sense and co-occurring with hedging strategies.

<table>
<thead>
<tr>
<th></th>
<th>nei/no used in a refusal sense</th>
<th>refusal nei/no with hedging</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDC, NoTa, BB</td>
<td>31</td>
<td>12</td>
</tr>
<tr>
<td>BNC2014</td>
<td>85</td>
<td>28</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>116</strong></td>
<td><strong>40</strong></td>
</tr>
</tbody>
</table>

5.1.4 But/men

The second probe investigated was the connective but/men. But/men is typically used to express contrast or disagreement. The main function of connectives, such as but and men, is to indicate that “according to the speaker’s opinion there is a contrast between the second state of affairs in relation to the first one” (Rudolph, 1996, p. 33) Disagreeing or expressing something which is in contrast to something which has been said already by an interlocutor can be regarded as threatening to the hearer’s positive face (Brown & Levinson, 1987, p. 66). Contradictions and disagreements are closely associated with disapproval and thus may indicate that the speaker disprove or is inconsiderate of the hearer’s feelings. Even contradicting oneself may be considered threatening to the speaker’s own positive face because of the wish to be perceived as consistent. A second incentive for testing but/men as a probe is its high frequency in spoken text. Due to the high frequency of negatives in conversation and the fact that negation and contrast are two closely related concepts, a high frequency of but should be expected in the spoken material (Biber et al., 1999, p. 82).

5.1.4.1 The structural and semantic properties of but/men

The structural and semantic properties of but and men are described in article 1 and will be summarised here. Formally, both but and men are typically classified as belonging to the word class of coordinating (adversative) conjunctions (Faarlund et al., 1997, p. 25; Quirk et al., 1985, p. 920). Syntactically, they connect elements with the same syntactic role (Biber et al., 1999, p. 79). But and men typically connect constituents on the same level, e.g. words, phrases and clauses (Faarlund et al., 1997, p. 1116; Quirk et al., 1985, p. 928), but cannot link most subordinate clauses, like and and or (Quirk et al., 1985, p. 925), because such clauses are normally outside the scope of negation. The uses of the Norwegian men as a
Coordinating adversative conjunction can be divided into two main types (Faarlund et al., 1997, p. 1138). One type of use is when the first clause is negated and the second clause expresses what the speaker thinks is correct. The second type is when the contrast lies in the content of the proposition of the connected clauses. The same is true for but. A contrast may repudiate what has been said or implied by the first clause (Quirk et al., 1985, p. 935).

Semantically, as adversative conjunctions, but and men have a core semantic meaning of contrast (Biber et al., 1999, p. 79; Fraser, 2013, p. 322). The notion of contrast in general and how it can be expressed linguistically cannot be discussed in detail here, but the notion of contrast as expressed by but and men will be addressed in more detail. But/men typically signals a contrastive view either to something that is said already or to something implied or unexpected. Blakemore (1989, p. 15) distinguishes between types of contrasts, the so-called “denial of expectation” and the “contrast”, i.e. semantic opposition, use. According to Schiffrin (1987a, p. 152), but marks an upcoming unit as a contrasting action. It can also entail a notion of opposition and also potentially a broken causal chain (Rudolph, 1996). Although expressing contrast is one of the basic ways of connecting ideas, events and utterances (Rudolph, 1996, p. 32), “the notion of contrast in English, and presumably other languages, is not well-defined” (Fraser & Malamud-Makowski, 1996, p. 865). However, although there is no general agreement on how contrast is to be understood, there is consensus that it can be explained as “a general semantic category encompassing several, more specific subtypes” (Dupont, 2019, p. 43), for example as in Quirk et al. (1985, p. 634), who distinguish between reformulatory, replacive, antithetic and concessive types of contrasts. These subtypes are also exemplified in several dictionaries, for example the antithetic use is illustrated in a major Norwegian dictionary, NAOB, with a quote from the Norwegian playwright Henrik Ibsen’s Kongs-Emnerne (p.95): “kræfterne svigter, men sjælen friskner” (The powers are failing, but the soul is healing).

Another complicating factor when describing contrast is that it is culturally and individually conditioned. In Norwegian the sentence “Det var vinter, men varmt” (It was winter, but warm) would express a contrast, but it may seem strange for people in other parts of the world where the climate and seasons are different (Faarlund et al., 1997, p. 1137). The contrast may also lie in the unexpectedness of what is said, e.g. “John is poor, but he is happy” (Quirk et al., 1985, p. 935). This sentence implies that his happiness is unexpected in view of his poverty, which in turn is related to our presuppositions and our experience of the world, i.e. poor people are usually unhappy.
On a functional level, *but* and *men* can also function as pragmatic markers, mainly discourse markers (see section 2.1.2 for a discussion of pragmatic markers). Fraser (2013, p. 322) describes *but* as the most general of the contrastive markers signalling a variety of different relationships depending on the context and whether or not it occurs in combination with other pragmatic markers, e.g. contrast, contradiction, challenge, topic change or apology. Furthermore, *but* is said to be undergoing a grammaticisation process moving from a turn-continuing function in initial position to a turn-yielding function in final position (Mulder & Thompson, 2006). Some uses of *men* also have a discourse marking function (see section 2.4 in article 1 for examples of *men* as a discourse marker). As a discourse marker, *men* can be used to change the topic of conversation with a purely text-organising function. *Men* can also be used to express surprise or irritation. One could argue that *men* in expressions expressing surprise could be contrastive as well, in the sense that they express something which is in contrast to the speaker’s expectations, but such instances are classified as non-contrastive here because the *men* is more an expression of surprise in the same way as, e.g. *oh* or *wow*, and not a contrastive marker. *Men* can also function as a hedging strategy on its own, particularly in clause final position. (See article 1 for examples of various uses of *men*.)

5.1.4.2 But/men in the ENPC

The investigation of *but/men* in the ENPC was carried out in the same manner as that of *no/nei*. The occurrences of *men* and *but* in direct speech in Norwegian and English original fictional texts were checked with regards to their degree of correspondence. Table 5-4 presents the filters applied in the ENPC, the total number of instances, the number of contrastive instances and the number of contrastive instances translated into *but/men* in the respective translations.

<table>
<thead>
<tr>
<th>Filters applied in the ENPC</th>
<th>Total no. of instances</th>
<th>Total no. of instances per 1,000</th>
<th>No. of instances translated into <em>but/men</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Men</em></td>
<td>ENPC fiction, Norwegian original, direct speech</td>
<td>694</td>
<td>9.01</td>
</tr>
<tr>
<td><em>But</em></td>
<td>ENPC fiction, English original, direct speech</td>
<td>654</td>
<td>7.11</td>
</tr>
</tbody>
</table>

19 Mulder and Thompson (2006) use the term *discourse particle* rather than *discourse marker*.
Of the 694 occurrences of *men*, 617 (89%) were translated with *but*. In the majority of occurrences which were not translated with *but*, a corresponding expression was omitted altogether. Of the total of 654 occurrences of *but*, 600 (92%) were translated with *men*. This gave a mutual correspondence of 90%, which means that we can assume near-equivalence between these two items.

### 5.1.4.3 A pilot study of *but/men* in NoTa, NDC, BB and BNC2014

As with *no/nei, but/men* was subjected to a pilot study in the four spoken corpora. In all, 150 random instances of *men* were retrieved from the Norwegian corpora and 150 random instances of *but* from the English corpus. Table 5-5 presents the filters applied in the corpora.

*Table 5-5 Filters applied in the NDC, NoTa, BB and BNC2014 (main contrastive study).*

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Filters for <em>men</em> and <em>but</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>NDC</td>
<td>Country: Norway&lt;br&gt;Category: Conversation between strangers, family, acquaintances, friends&lt;br&gt;Extended search: 50 random results</td>
</tr>
<tr>
<td>NoTa</td>
<td>Representative selection: yes&lt;br&gt;Genre: Conversation between strangers, family, acquaintances, friends&lt;br&gt;Extended search: 50 random results</td>
</tr>
<tr>
<td>BB</td>
<td>Extended search: 50 random results</td>
</tr>
<tr>
<td>BNC2014</td>
<td>Show in random order, thin selection</td>
</tr>
</tbody>
</table>

The pilot study of *but* and *men* showed more promising results than that of *nei* and *no*. A total of 285 of the 300 instances of *men* and *but* were used in their contrastive sense and of these 68.2% and 59.3% co-occurred with hedging strategies in the two languages respectively. Table 5-6 presents these figures schematically.
All of the instances of *but* were contrastive, but some of the instances of *men* were considered irrelevant. The few uses of *men* that were disregarded were typically instances where *men* was used as a discourse marker, as in example (5-5). The contrastive use of *but* and *men* is illustrated in (5-6) and (5-7). Example (5-6) and (5-7) also show how they typically co-occurred with hedging strategies.

(5-5)

A: særlig norsken sånn du har ikke F1 i norsk?
‘A: especially Norwegian like that you do not have F1 in Norwegian?’
B: nei
‘B: no’
A: nei ## men e # ja
‘A: no ## but e # yes’
B: **Men** ellers da?
‘B **But** otherwise then?’

NDC darbu_01um_darbu_02_uk no.100x

(5-6)

A: and I ’m I ’m some privileged arsehole who should feel bad erm and I was just like well I ’m I ’m afraid I said to her if I could afford to send my kids to private school I would
B: mm
A: **but** one of them's *probably* gonna have to win the scholarship because it's very very expensive

BNC2014 SQ82 no. 132x

(5-7)

A: jeg er jo et prakteksemplar # jeg er jo_uforståelig_
A: ‘I am [pragmatic marker] a prime example # I am [pragmatic marker]_incomprehensible_’
B: *ja ja man kan man kan*
5.1.5 Comparing no/nei and but/men

From the outset, both no/nei and but/men seem to be eligible as probes. However, as the results from the pilot study show, only 38.6% of the occurrences of no/nei were refusals and thus a lot of corpus data had to be sifted through in order to find the relevant uses. Furthermore, only between 30 and 40% of the refusals co-occurred with hedging strategies, which led to a smaller number of hedging strategies than expected. One of the possible explanations for this may be that no/nei has a broader range of uses than but/men. Their polysemy and multifunctionality require advanced pragmatic tagging or manual processing of great amounts of data to remove irrelevant uses. In contrast, men and but do not have the same level of polysemy as no and nei. This makes it easier to locate the contrastive uses and subsequently the co-occurring hedging strategies.

Secondly, the low number of hedged refusals could also be related to the use of refusing strategies as discussed in Beebe et al. (1990). It could be that if speakers choose a direct refusal strategy, they may already be assessing the face-threat as minimal. As can be seen from several studies (e.g. Beebe et al., 1990; Chang, 2009; and Demirkol, 2016), refusals can be expressed in a number of ways. Thus, a speaker may choose to opt for a more indirect strategy than the direct non-performative no and nei. Such indirect and potentially heavily hedged utterances will be lost when using no and nei as probes.

The two sets of probes were measured against the criteria set up in section 5.2.1 and summarised in table 5-7. Of the two sets of probes tested, but/men expressing contrast fulfilled the criteria better. No/nei was discarded, and but/men was tested further to see if it could work as a probe. But/men had a very high mutual correspondence in the ENPC, which lays the foundation for a sound tertium
comparaison, when the two languages have grammatical categories in common and the labels are used for the same phenomenon, which is the case with but and men. Their high degree of mutual correspondence in the ENPC is promising when it comes to studying the same sort of contexts in both languages. Furthermore, but and men are very frequent in spoken discourse and have a neutral semantic prosody. When they are used, it is frequently in their contrastive sense making the sifting process less demanding, although there are occurrences deemed irrelevant.

Table 5-7 Summary of the selection criteria for the choice of probe.

<table>
<thead>
<tr>
<th></th>
<th>no/nei</th>
<th>but/men</th>
</tr>
</thead>
<tbody>
<tr>
<td>semantically similar/comparable</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(semantic prosodies)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>syntactically similar/comparable</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>operating at more or less the same</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>common in everyday spoken language</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>introducing a speech situation which</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>was likely to be perceived as face</td>
<td></td>
<td></td>
</tr>
<tr>
<td>threatening</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the majority of the instances were</td>
<td>÷</td>
<td>+</td>
</tr>
<tr>
<td>relevant (i.e. a potential face-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>threatening use)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the relevant instances gave a high</td>
<td>÷</td>
<td>+</td>
</tr>
<tr>
<td>number of hedging strategies</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although, but/men seemed to be the better alternative of the two probes tested, there were still some challenges with the use of this probe, which required further testing. First, one of the assumptions behind the choice of this probe was that expressing a contrastive view through the use of men/but could potentially be face-threatening and call for remedial hedging strategies. However, the majority of the contrastive utterances did not appear face-threatening. In example (5-8), hedging strategies are used, but the contrastive utterance does not seem to be a threat to the speaker or hearer’s face.

(5-8)

A: [...] kameratene mine de s- kan kommer egentlig fra Chile da
‘A: [...] my mates they s- can originally come from Chile [pragmatic particle]’
B: ja
‘B: yes’
A: så vi begge snakker spansk men når vi er sammen så snakker vi liksom norsk
uansett da ikke sant
‘A: so we both speak Spanish but when we are together we like speak Norwegian regardless [pragmatic particle] right’

NoTA 030> <who_avfile 029-030 no. 971

The *men* (‘but’) signals that what is said in the first clause is in contrast to what is said in the second clause, i.e. they are both from Spanish-speaking countries and speak Spanish, so one might expect that they would speak Spanish to each other, but they speak Norwegian. *Liksom* (‘like’) here, does not seem to fulfil any of the typical functions of clause-medial *liksom*, i.e. exemplification, approximation, quotative or intensifying (see appendices 1 and 2), but may still be described as having an interpersonal function similar to that of many of the Norwegian particles. Hasund (2003, p. 185) states that in contexts where *liksom* does not serve a clear textual or subjective function, its function may be described as primarily interpersonal and a measure to involve the interlocutor. In general, speakers often use hedging strategies to involve the interlocutor and facilitate discussion (Hyland, 1996, p. 433). This may be part of the explanation why hedging strategies also co-occurred with non-face-threatening uses of *but/men*.

Another challenge, with both of the probes, was related to the scope, i.e. how much context should be reviewed to find hedging strategies. In the pilot studies, the scope was the turn as it was transcribed in the corpora. However, as informal spoken language is fragmented, with a lot of interruptions, backchanneling, false starts etc. an utterance, in many cases, stretched beyond several transcribed turns. In example (5-9), the utterance of speaker B is an example of backchanneling and speaker A resumes the turn and continues. Sometimes, the topic also changed slightly within one turn, which meant that the use of some hedging strategies were not connected to the probe. (See section 5.2.1 for a discussion of utterances and turns.)

(5-9)

A: and they've both kind of forgotten about it erm but obviously I think it's been kind of
B: yeah it's been in the background
A: yeah in the back of his mind and and suddenly he's had a little bit of space to do it and he's actually sat down and done it
B: he's probably wanted to for a while

BNC2014 SVBH no. 65

The question of scope became particularly relevant when the probe *but/men* was compared to random pieces of text to see whether using the probe to access hedging strategies was more fruitful than simply studying random pieces of text, which was the purpose of the study reported in article 1. In this second
test, only hedging strategies within the first clause to the right of the probe were studied. This decision was made because of the need to compare the use of probes to randomly selected units of text to test the method. Identifying sentences or other grammatical units in spoken corpora is difficult as punctuation is often absent from the transcription. Furthermore, speakers rarely speak in complete sentences and spoken language is often characterised by false starts, interruptions and fragments of text and the only thing we can say for certain about spoken text is that the flow of speech is divided into some kind of units, in the form of pause units, tone units or both (Stenström, 1994, p. 7). Therefore the smallest grammatical unit that can express a proposition was chosen as the scope of study in order to make the pieces of text eligible for comparison. Thus, 150 clauses with contrastive men and 150 clauses with contrastive but were compared to 150 randomly selected Norwegian and 150 randomly selected English clauses from NoTa, NDC, BB and BNC2014. The randomly chosen clauses without probes were retrieved by searching for the verb tag, mainly because it is impossible to search for nothing in a corpus and because most utterances include some form of verb. The procedure and results of this test are reported in article 1. The decision to restrict the scope of study to the immediate clause to the right of the probe seemed, however, a bit too restrictive. The downside to limiting the scope to a clause was that hedging strategies occurring outside the immediate clause, but which also had an effect on the utterance as a whole, were disregarded. The challenges with restricting the scope to a clause are discussed in more detail in the concluding remarks of article 1 and in section 5.2.1.

The results of this second test showed that a significantly higher number of hedging strategies were retrieved with the probe than without in the Norwegian data, but the results were not significant for the English data (see article 1 for a description of a more detailed description of the results). The significance testing was done in R using a Pearson’s Chi Square test. Despite this result, but/men was chosen as a probe for the full-scale study of hedging strategies mainly for two reasons. First, the test gave statistically significant results for the Norwegian data. Second, due to the fact that the Norwegian and English corpora are only partly comparable (see section 4.2), the issue of a tertium comparationis with respect to the phenomenon studied becomes even more important. By using the probe to identify similar communicative situations in the corpora, i.e. speakers expressing contrasts, the units of comparison are more comparable and equivalent than if random utterances had been chosen. Thus, the choice of using the probe to elicit similar speech situations from the corpora ensures a tertium comparationis concerning the situations from which the hedging strategies are retrieved. This need to ensure comparability in the study of pragmatic phenomena is highlighted in e.g. Kasper (1990, p. 198) in her discussion of the universality of politeness theories. Kasper (1990, p. 198) warns against
assuming that conventionalised norms of usage are consistent across languages and points out that idiomatic expressions and other routine formulae tend to be language specific. It can therefore be difficult to find formal or functional equivalents across languages.

5.2 Phase 2: A contrastive study of hedging strategies in English and Norwegian (articles 2 and 3)

5.2.1 Procedure
The procedure builds on the results from phase 1 reported in section 5.1 and article 1. However, some refinements have been made in the way hedging strategies are identified and interpreted around the probe. Following the decision to use but/men as a probe, the full-scale study was carried out in the four corpora described in section 4.4. In all, 800 occurrences of but and men respectively were retrieved and the number of contrastive uses were determined through an integrated reading process (Aijmer & Rühlemann, 2015, p. 12). Integrated reading combines the vertical (quantitative) reading typical of corpus linguistics with the horizontal (qualitative) reading typical of pragmatics. The 1,600 random instances of but and men were retrieved in the Norwegian and English corpora (vertical reading) using the filters summarised in Table 5-2 (but increasing the number of hits to 800 instead of 50). Subsequently, the contexts to the right of the node, i.e. the probe, were read more closely to determine whether but/men was used in a contrastive sense and whether they co-occurred with hedging strategies (horizontal reading). But and men were mainly considered contrastive when they denoted denial of expectation, opposition of two elements (antithetic use), modification of a previous statement, restriction of a previous statement, or when a speaker objects to something said by another speaker. Examples (5-9) and (5-10) illustrate contrastive uses of but and men. In (5-9) the contrast signalled by men may be described as antithetic. The speakers are discussing a dialectal feature and where it is still being used. In (5-10) but signals an objection. The speakers are discussing the name of an old teacher, and speaker A objects to speaker B’s suggestion.

(5-9)

A: men jeg vet ikke om de sier det i Valle nå? # jo _uninterpretable_ gamle gjør vel #
ja
‘A: but I don’t know if they say that in Valle now? # yes _ uninterpretable_ old probably do # yes’
B: * ja * ja de gamle men ikke de e yngre kan jeg tenke meg
‘B: *yes *yes the old but not the e young I would imagine

NDC evje_03gm><who_avfile evje_03gm-04gk no. 725
(5-10)
A: Miss --ANONnameN was she was she Miss --ANONnameN?
B: no no her name was --ANONnameN
A: no **but** I think Miss --ANONnameN got married
B: no no she was a big woman with brown hair
A: oh right

BNC2014 S38F no. 472

*But* and *men* were not considered contrastive when they were used as discourse markers, as in (5-11), where *men* is used to reintroduce an earlier topic. The speakers are talking about football and that only women support FC Barcelona. Speaker A gives another example of a thing only women do, i.e. vote for SV (a Norwegian left-wing political party) and the speakers laugh at this stereotypy. Speaker A then returns to the topic of football and speaker B’s football career.

(5-11)
A: […] de mente de mente det er jentelag # bare damer heier på Barcelona i følge dem
‘A: […] they thought they thought it is a girl’s team # only women support Barcelona according to them’
[…]
A: å bare damer stemmer så- # SV og så _uninterpretable_
‘A: and only women vote so- # SV and so _uninterpretable_’
B: * ja jeg er
‘B: *yes I’
B: enig med dem
‘B: agree with them’
A: _laughter_nei nei # fotball **men** du sa du spilte for fotballag?
‘A: _laughter_ no no # football **but** you said you played for a team?’

NDC bodoe_02uk><who_avfile bodoe_01um-02uk no. 512

Several of the uses of *but* and *men* were difficult to classify, often due to lack of context. In (5-12), for example, the recording starts immediately before *men*, which gives limited context and makes it difficult to classify. This particular case was still included as a contrastive use of *men*, because it could be interpreted as an objection. In (5-13) the speaker stops speaking just after *men*. This example was also regarded as contrastive because of the negator *ikke* (’not’) which indicates contrast to the positive first clause. In (5-14) the speakers are talking about the NHS, and it seems as if the conversation is quite intense with a lot of overlapping speech and turn-switching. This makes it difficult to determine the function of *but* by speaker B, and thus this *but* is not regarded as contrastive.

(5-12)
A: **men** e det er sårne bondetøser det som er sånn _uforståelig_
‘A: **but** e it is those farm girls who are like that _incomprehensible_’
BB Ramsy><who_avfile 18 no. 40

(5-13)
A: um ja det er det òg men det er ikke noe så mye tilrettelagt for # folk som kommer utenifra for # det er det ikke
‘A: um yes it is that too but it is not very well facilitated for # people who come from other places for # it is not
B: nei
‘B: no’
A: der er vel et og anna skilt der som peker opp mot fjellet men ikke noe
‘A: there is well one or another sign there pointing up towards the mountain but not something’

NDC steigen_01um><who_avfile steigen_01um-02uk no. 490

(5-14)
A: the way people expect doctors to be when I see news reports of some tragic thing when the baby died because of something or other and the parents are going oh so this doesn't happen again in future and we're suing
B: yeah
A: the NHS for negligence
[...]
A: and and I just think well they're not
B: they're not gods
A: they're not gods
B: no
A: and
B: >> and they are you know they're fallible
A: exactly
B: >> yeah yeah
A: >> you can't expect to know
B: >> yeah yeah
A: everything
B: >> and actually you know but
A: >> and and you know and the misdiagnosis thing you d- I would not expect to er
B: >> get it right but did you know

BNC2014 SN64 no. 549

Following the identification of contrastive uses of but and men, all hedging strategies (in line with the definition of hedging in section 2.1.3) occurring within the same utterance as but/men were identified. Since hedging strategies can take on almost any linguistic form, it was challenging to identify hedging strategies. To make sure that the hedging strategies in fact were hedging strategies a significant amount of secondary literature had to be consulted. The identification of hedging strategies can be described
as a two-stage process. First, all potential hedging strategies were extracted into a list and grouped into types, e.g. tag questions, adverbs, general extenders, etc. Then previous research on these various types was consulted to see how their pragmatic functions had been described in the literature. This process resulted in a guide for analysis (appendices 1 and 2). The process of consulting existing literature on realisations of hedging strategies also opened my eyes for potential realisations which I had not discovered in the first analysis of the data. Therefore, the data was reviewed a second time in light of the guide to make sure that all potential hedging strategies were retrieved.

Of the 800 instances of *men*, 582 were classified as contrastive and 60% of these co-occurred with hedging strategies. Of the 800 instances of *but*, 659 were classified as contrastive and 53% of these co-occurred with hedging strategies. This left 677 hedging strategies to be examined in Norwegian and 762 hedging strategies to be examined in English. The numbers are summarised in table 5-8 (See articles 2 and 3 for a presentation of the results).

Table 5-8 Number of contrastive uses of the probe and number of the probe co-occurring with hedging strategies.

<table>
<thead>
<tr>
<th></th>
<th>No. of contrastive uses</th>
<th>No. of contrastive uses occurring with hedging strategies</th>
<th>No. of hedging strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>But</em></td>
<td>659</td>
<td>351</td>
<td>762</td>
</tr>
<tr>
<td><em>Men</em></td>
<td>582</td>
<td>352</td>
<td>677</td>
</tr>
<tr>
<td>Total</td>
<td>1,241</td>
<td>703</td>
<td>1,439</td>
</tr>
</tbody>
</table>

In the first pilot study of *but/men* (see section 5.1.4.3) a turn as it was transcribed in the corpora was used as the scope of study. In the second pilot study, the results of which were reported in article 1, only hedging strategies within the immediate clause following *but/men* were analysed. However, as discussed in section 5.1, both of these approaches caused some problems and in both cases, relevant hedging strategies were left out of the study. The first approach led to utterances going over more than one turn due to backchanneling, etc. being left out of the study. In the second approach, relevant hedging strategies within the same turn, but outside the immediate clause following the probe were also outside the scope of study. Building on the experiences from both of these pilot studies, the scope in the full-scale study was expanded to include all hedging strategies following the probe within an utterance, as illustrated in figure 5-2.
However, *utterance* is a fuzzy concept and there is no precise linguistic definition determining what an utterance is and when it starts or stops. Phonetically, an utterance has been described as a unit of speech bounded by silence, i.e. breath or other kinds of pauses. However, speakers use pauses for other purposes than ending an utterance, for instance to hesitate or simply to breathe or for more strategic purposes, such as to put focus on a particular part of the message or to mark off units of discourse, e.g. topic and subtopics (Stenström, 1994, p. 7). Pauses are generally marked in the transcription, typically by parenthesis with dots or the like (see section 4.4). However, the purpose of the pause is not always clear from the transcription, which makes it difficult to distinguish between types of pauses.

In dialogues, an utterance is often considered equal to a turn, but a turn can also consists of more utterances. A turn is generally transcribed as a separate unit in corpora and for that reason it can be easier to use as a unit of study. However, theoretically the term *turn* is also unclear. Linell and Gustavsson (1987) define a turn as the period when a speaker has the word and has the right or duty to speak. Turns go back and forth between speakers, and a speaker typically indicates that he/she is finished by pausing or through intonation. However, speakers may also say something without it being their turn and may also attempt to take over the turn by speaking loudly or simultaneously as the speaker who possesses the turn. Speakers may also speak at the same time as their interlocutor without attempting to take over the turn. For example, if speaker 1 is talking, speaker 2 can give feedback in the form of backchanneling (Stenström, 1994, p. 81), which is not an attempt to take over the turn, but to signal active listening and interest. Also frequently in dialogues, speakers are interrupted, but take back the turn and carry on where they left off. I have chosen to limit the study to hedging strategies that occur within the same utterance/turn, unless there is a clearly marked pause followed by a topic shift. If the utterance extends over more turns, e.g. because the speaker is interrupted, but takes back the turn at a later time and continues the topic of the original utterance, I have chosen to see this as a part of the original utterance. This was, however, not straight-forward and there were some challenging
5.2.2 Analysis of the data

The hedging strategies retrieved through the use of the probe were analysed for two purposes. First, the purpose was to discover differences and similarities in the types of hedging strategies used in Norwegian and English. To ensure a sound basis for the cross-linguistic comparison, an operational classificatory framework was needed to be able to account for the hedging strategies in the material. Therefore, the study also aimed to propose a framework based on the existing classificatory frameworks described in section 2.2.1. (See section 2.2.1 for a description of the frameworks and article 2 for a discussion of them.) Second, the hedging strategies were analysed with the purpose of comparing realisations in Norwegian and English, i.e. how Norwegian speakers express hedging in
Norwegian compared to how English speakers express hedging in English. The results of this study are reported in article 3.

5.2.2.1 Analysis of data in article 2

Based on previous research (see section 2.1.1), hedging strategies seem to have an effect within the proposition, on the relationship between the speaker and the proposition or on the relationship between the speaker and the hearer. This threefold distinction of types of hedging strategies was used as a point of departure for the analysis of the strategies. Following this broad classification, sub-categories from existing classificatory frameworks were used to describe each hedging strategy in more detail. Table 5-9 presents and describes the different categories and gives examples of each category from the data (See article 2 for a more detailed description of the categories).

Table 5-9 An overview of the categories applied in article 2.

<table>
<thead>
<tr>
<th>Within the proposition</th>
<th>Between the speaker and the proposition</th>
<th>Between the speaker and the hearer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rounders</strong></td>
<td><strong>Adaptors</strong></td>
<td><strong>Understaters</strong></td>
</tr>
<tr>
<td>(Prince, Frader, and Bosk 1982)</td>
<td>(Prince, Frader, and Bosk 1982)</td>
<td>(Hübler 1983)</td>
</tr>
<tr>
<td>Indicating approximation or that the precise term is not relevant</td>
<td>Indicating that a term is not necessarily the most appropriate term</td>
<td>Underrepresenting the state of affairs</td>
</tr>
<tr>
<td>but the girls surely would’ve noticed he’d been gone about three weeks (BNC2014 SN4N no. 143)</td>
<td>that’s teal is a very sort of greyish type taupe (BNC2014 SRKVK no. 171)</td>
<td>I just find it a bit draining (BNC2014 S48K no. 274)</td>
</tr>
<tr>
<td><strong>Plausibility shields</strong></td>
<td><strong>Attributions shields</strong></td>
<td><strong>Speaker-oriented hedges</strong></td>
</tr>
<tr>
<td>(Prince, Frader, and Bosk 1982)</td>
<td>(Prince, Frader, and Bosk 1982)</td>
<td>(Hyland 1996)</td>
</tr>
<tr>
<td>Expressing uncertainty on behalf of the speaker in relation to the proposition</td>
<td>Attributing the belief to another source (incl. hiding the source)</td>
<td>Reducing the commitment of the speaker</td>
</tr>
<tr>
<td>I might go on in September (BNC2014 SEGU no. 305)</td>
<td>according to this book […] they they were genuine (BNC2014 SVBJ no. 543)</td>
<td>that’s what every country says isn’t it ? (BNC2014 SAVN no. 680)</td>
</tr>
<tr>
<td><strong>Speaker-oriented hedges</strong></td>
<td><strong>Hearer-oriented hedges</strong></td>
<td></td>
</tr>
<tr>
<td>(Hyland 1996)</td>
<td>(Hyland 1996)</td>
<td></td>
</tr>
<tr>
<td>Establishing common ground between the speaker and the hearer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In order to show the distribution of each type within each language, the percentage of each type within each language was calculated. To compare the different types across the two languages, the frequencies of each type was compared using the log-likelihood and effect size calculator provided by the University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University.\footnote{http://ucrel.lancs.ac.uk/llwizard.html} The results are presented in article 2.

5.2.2.2 Analysis of data in article 3

For the investigation of similarities and differences between the two languages with respect to realisations of hedging strategies, each hedging form was analysed. The various realisations were grouped based on their form, which resulted in the 13 categories presented in table 5-10. The choice to categorise the hedging strategies based on their form was related to the desire to highlight differences between the two languages and to avoid combining functional and formal categories. (See article 3 for a presentation and a discussion of the categories). Originally, the realisations of hedging strategies I had retrieved using the probe were grouped and labelled according to the classification in appendices 1 and 2. However, the combination of formal and functional categories caused problems in the comparison of the various forms. For example, the category of pragmatic markers (discussed in section 2.1.2) is based on the pragmatic function of a form in a particular context and could in principle, depending on how the term is defined, encompass nearly all of the forms retrieved. By using formal categories, such as \textit{1/2 person pronoun + cognitive verb}, such problems were avoided.
To compare the different categories of forms across English and Norwegian, a Fisher’s Exact Test was performed in R to test whether there was a statistically significant association between language and form of hedging strategy. Each category of forms was also compared across the two languages using UCREL’s log-likelihood calculator. The results of the comparison are presented in article 3.

5.3 Phase 3: An intervarietal study of hedging strategies in English by Norwegian learners (article 4)

The third phase of the project builds on the results from article 3. The 10 proportionally most frequent forms among the hedging strategies identified in the English material in article 3 functioned as a point of departure for the study described in article 4. The purpose of the intervarietal study was to investigate potential over- and underuse of these 10 forms by Norwegian advanced learners of English compare to the use by native English speakers.
5.3.1 Procedure

The methodological approach in this part of the study can be placed towards the more traditional form-to-function end of the spectrum (see section 2.2.5). The reason for applying a form-to-function approach was related to the small size of the corpora. I feared that the form1-to-function-to-form2 approach would yield a too limited ground for comparison. The 10 forms were searched for in the LOCNEC and LINDSEI-no corpora (described in section 4.5), and their frequencies were compared. In order to select the 10 forms which were to function as the search terms for this study, the 30 forms with the highest frequency in article 3 were considered. The 10 most frequent of these were chosen as the forms to investigate. Table 5-11 presents the top 10 forms and table 5-12 presents some of their hedging characteristics. The descriptions are based on previous studies of native speaker usage, illustrated with examples from the LINDSEI and LOCNEC corpora.

Table 5-11 Overview of the 10 most frequent forms retrieved from the BNC2014 in phase 2.

<table>
<thead>
<tr>
<th>Hedging strategy</th>
<th>Raw frequency</th>
<th>% of total no. of hedging strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Like</td>
<td>167</td>
<td>23.36%</td>
</tr>
<tr>
<td>2 Just</td>
<td>78</td>
<td>10.91%</td>
</tr>
<tr>
<td>3 You know</td>
<td>62</td>
<td>8.67%</td>
</tr>
<tr>
<td>4 I think</td>
<td>38</td>
<td>5.31%</td>
</tr>
<tr>
<td>5 Thing(s)</td>
<td>26</td>
<td>3.64%</td>
</tr>
<tr>
<td>6 I mean</td>
<td>24</td>
<td>3.36%</td>
</tr>
<tr>
<td>7 A bit</td>
<td>20</td>
<td>2.80%</td>
</tr>
<tr>
<td>8 Might</td>
<td>16</td>
<td>2.24%</td>
</tr>
<tr>
<td>9 Probably</td>
<td>15</td>
<td>2.10%</td>
</tr>
<tr>
<td>10 Kind of / kinda</td>
<td>13</td>
<td>1.82%</td>
</tr>
</tbody>
</table>

Table 5-12 Overview of the hedging functions of the forms with corpus examples.

<table>
<thead>
<tr>
<th>Hedging expression</th>
<th>Hedging properties</th>
<th>Example from LOCNEC / LINDSEI-no</th>
</tr>
</thead>
<tbody>
<tr>
<td>A bit</td>
<td>- underrepresents the state of affairs (Markkanen &amp; Schröder, 1997, p. 7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- serves a face-saving function (Hasselgreen, 2004, p. 213)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I guess he's a bit. similar to Norwegians in his shy way of being (LINDSEI-no NO004)</td>
<td></td>
</tr>
</tbody>
</table>
I mean - negotiates meaning by nuancing a previous statement or softening an evaluative comment (Beeching, 2016, p. 185)
- signals hesitation

I think - indicates uncertainty
- softens or tones down a previous “head” (Aijmer, 1997, p. 25)
- marks hesitation/informality

Just - helps the hearer in interpreting the utterance (Aijmer, 2002, p. 158).
- downplays the force of an utterance (Aijmer, 2002, p. 163)
- modifies an assertion considered to be face-threatening (Aijmer, 2002, p. 169)
- tones down the speaker’s opinion
- functions as an “attitude-diminisher” (Quirk, Greenbaum, Leech & Svartvik 1985, p. 598)

Kind of / kinda - functions as an adjuster word (Aijmer, 2002)
- may indicate that an appropriate word is not cognitively available
- may also indicate approximation

Like - its core function is “to flag approximation and hedge discourse” (Beeching, 2016, p. 127)
- downplays “a potentially dogmatic delivery” (Beeching, 2016, p. 132)
- indicates imprecision or approximation, for example with numbers
- indicates speaker uncertainty
- introduces direct speech in combination with the verb BE, implying that what is being rendered is not the exact words which were said (Beeching, 2016, p. 131)

Might - expresses a degree of possibility
- *might* is associated with hypothetical situations, speaker tentativeness, and politeness (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 485)

Probably - reduces commitment and says something about the degree of reliability of the knowledge (Chafe, 1986, p. 263)
- expresses strong doubt about the truth of the proposition (Hübler, 1983, p. 124)

and *I mean* sort of like it it must have been awful (LOCNEC EN0018)

[...] that would be a good match for me *I think* (LINDSEI-no NO049)

it's just (eh) . it's typical sort of student night-club really (LOCNEC EN0037)

it's kind of . a bit south-east (LOCNEC EN0024)

[...] my few words is probably like twenty and their few words is probably like two hundred (LOCNEC EN0030)

I don't know I mean *I might* change my mind (LOCNEC EN0044)

yeah sixteen eighteen [...] probably [...] n= no (eh) perhaps: later (LINDSEI-no NO027)
**Thing(s)**
- conveys meaning in situations where speakers do not have the necessary expressions at their disposal (Channell, 1994)
- reduces speaker commitment and guards against criticism (Channell, 1994)
- general extenders, (e.g. *and things* (Overstreet, 1999)) may indicate shared knowledge and can appeal to the hearer

**You know**
- creates common ground between the interlocutors (Beeching, 2016, p. 97; Overstreet, 1999, p. 75)
- creates a sense of involvement (Nikula, 1997, p. 199)
- expresses intimacy with the hearer (Aijmer, 2002, p. 180)

you just get like this *thing* it just works it just flows

(LOCNEC EN0010)

[…] at the moment it depends what w= what my year out goes like you know

(LOCNEC EN0018)

WordSmith 7 (Scott, 2016) was used as a search tool to retrieve all occurrences of the 10 forms. Wordsmith 7 was chosen over other similar tools as it allows searches with a custom-made tag file, which enabled me to search within one speech task and for B-turns, i.e. learner turns, only.\(^{21}\) Furthermore, the WordSmith 7 wordlist statistics function was used to find the total number of tokens in the selection which served as the basis for the statistical analysis in article 4. The search for the 10 words in the two corpora gave in total 5,394 instances. Table 5-13 presents the distribution of the forms in both corpora.

Table 5-13 Overview of the number of forms in both corpora.

<table>
<thead>
<tr>
<th>Search word</th>
<th>No. of instances in the corpora</th>
<th>No. of hedging uses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOCNEC</td>
<td>LINDSEI-no</td>
</tr>
<tr>
<td>Like</td>
<td>816</td>
<td>391</td>
</tr>
<tr>
<td>Just</td>
<td>582</td>
<td>287</td>
</tr>
<tr>
<td>You know</td>
<td>432</td>
<td>68</td>
</tr>
<tr>
<td>I think</td>
<td>288</td>
<td>193</td>
</tr>
<tr>
<td>Thing(s)</td>
<td>265</td>
<td>101</td>
</tr>
<tr>
<td>I mean</td>
<td>347</td>
<td>32</td>
</tr>
<tr>
<td>A bit</td>
<td>142</td>
<td>57</td>
</tr>
<tr>
<td>Might</td>
<td>44</td>
<td>17</td>
</tr>
<tr>
<td>Probably</td>
<td>68</td>
<td>38</td>
</tr>
<tr>
<td>Kind of / kinda</td>
<td>53</td>
<td>75</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3,037</td>
<td>1,259</td>
</tr>
</tbody>
</table>

5.3.2 Analysis of the data

All 5,394 instances were read manually using the KWIC function in WordSmith 7. All instances which were clearly not hedging were disregarded, e.g. the use of *like* in example (5-16). The number of hedging uses of each word is presented in table 5-12 above. However, it is challenging to determine whether a particular expression is used as a hedging strategy or not due to the fact that these expressions may serve multiple purposes simultaneously. Example (5-17) illustrates the pragmatic use of *I mean*. According to Beeching (2016, p. 183), the functions of *I mean* range from canonical to fully fledged pragmatic uses, and there are many uses which are half way between pragmatic and canonical. The use of *I mean* in (5-17) could be classified as a form of justification (Beeching, 2016, p. 188), i.e. the speaker clarifies or explains why he/she thinks Brad Pitt is brilliant in a film role. Justifying a point of view does not necessarily need to entail hedging. However, in this case, the clause starting with *I mean* also modifies the content in the previous clause, i.e. the speaker thinks Brad Pitt is brilliant in this particular role, but underlines that he/she is aware that Pitt usually does not perform that well. Due to such challenges, an inclusive approach to the process of identifying hedging uses was chosen, meaning that if there was uncertainty whether an expression could be a hedging strategy or not, it was included. This inclusive approach was applied on the learner speaker data as well as the native speaker data to
ensure the validity of the comparison. Example (5-18) illustrates a clearly hedging use, where *I think* and *might* indicate uncertainty on behalf of the speaker.

(5-16)  
I don't *like* being inside sitting still anyway  
LINDSEI-no NO006_FD.txt no. 35

(5-17)  
A: and Brad Pitt and Morgan Freeman are brilliant *I mean* Brad Pitt's normally sort of quite light headed you know sort of quite light weight  
LOCNEC EN0015.txt no. 33

(5-18)  
*I think* it's Oxford it *might* have been Cambridge  
LOCNEC EN0021.txt no. 30

The number of instances of hedging strategies (tokens) and the range of expressions (types) were compared across the two varieties. Both interspeaker and intraspeaker differences were considered. The statistical significance of the differences between the groups, both for the total number (tokens) and for each expression, was calculated using the UCREL log-likelihood calculator. This calculator takes into account the size of the two corpora compared, so figures do not need to be normalised. It is also widely applied in other intervarietal studies using the LINDSEI and LOCNEC corpora, such as Aijmer (2011). Relying on significance tests of frequencies, such as log-likelihood, has been heavily criticised. Gablasova et al. (2017) state that such tests are not reliable because the elementary assumption of such tests, i.e. the independence of observations, is violated. However, the issue of independence is a complex issue, particularly in linguistics. It is difficult to determine to which degree a speaker influences him/herself when speaking. For example, if a speaker uses *like* as a hedging strategy in one utterance, it is difficult to determine whether that would influence his/her choice of hedging strategy in a later utterance. Although some speakers seem to prefer particular expressions over other, I would argue that the hedging strategies as they occur in the spoken corpora used here could be considered independent. Although this claim is made with reservations. In this study, the log-likelihood test was chosen because a relatively high number of instances of each word was expected in the two corpora and the log-likelihood test is less likely to overestimate the significance of high frequencies.

Furthermore, studies of particular words or structures in corpora have received criticism for using frequency as the only diagnostic of importance (Gries, 2006; Paquot & Plonsky, 2017). Looking at frequency data alone may be misleading because there is a risk of “relying on speakers' idiosyncrasies
or (register-specific) outliers” (Gries, 2006, p. 198). Gries (2006, p. 196) argues that there may be occasions where it is more useful to consider the dispersion of the words/structures in the corpus. Therefore, dispersion of the expressions and intraspeaker variation, in both corpora, were considered. The dispersion of the ten expressions and intraspeaker variation was illustrated through the use of boxplots created in R. The boxplots in article 4 showed individual variation in the total number of hedging strategies (tokens) per speaker, the total number of uses per expression and the total number of different hedging strategies (types) per speaker. The overlap of the plots served as the foundation for the discussion.

The dispersion of each word was also calculated using a statistical measure, i.e. the coefficient of variation (CV). Dispersion is a measure that quantifies the degree to which a word is distributed across a corpus, i.e. whether the occurrences of a word are distributed evenly or unevenly (Gries, 2006, Forthcoming), and the CV is one way of quantifying dispersion. The coefficient of variation is calculated by dividing the standard deviation by the mean. The result summarises the amount of variation as a percentage or proportion of the total. Gries (2006) suggests several different statistical methods of measuring dispersion, but the coefficient of variation was used because it is suitable for comparing the degree of variation between datasets with different means and different values. Another advantage is that it handles zero-occurrences, of which there were a lot in these datasets. However, disadvantage to using CV as a measure of dispersion is that it may indicate a relative high degree of relative variation when the standard deviation is low and the mean is low. The actual calculation of the CV was not included in article 4, but functioned as way ensuring the quality and correctness of the boxplots. The calculations are given in table 5-14.

<table>
<thead>
<tr>
<th></th>
<th>LINSEI-NO</th>
<th>LOCNEC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std dev</td>
<td>Mean</td>
</tr>
<tr>
<td>a bit</td>
<td>1.85</td>
<td>1</td>
</tr>
<tr>
<td>I mean</td>
<td>2.53</td>
<td>0.6</td>
</tr>
<tr>
<td>I think</td>
<td>2.79</td>
<td>2.96</td>
</tr>
<tr>
<td>just</td>
<td>3.28</td>
<td>4.22</td>
</tr>
<tr>
<td>kind of / kinda</td>
<td>1.61</td>
<td>1.3</td>
</tr>
<tr>
<td>like</td>
<td>4.66</td>
<td>2.84</td>
</tr>
<tr>
<td>might</td>
<td>0.69</td>
<td>0.34</td>
</tr>
<tr>
<td>probably</td>
<td>1.27</td>
<td>0.76</td>
</tr>
<tr>
<td>thing(s)</td>
<td>1.92</td>
<td>1.78</td>
</tr>
<tr>
<td>you know</td>
<td>2.22</td>
<td>1.26</td>
</tr>
</tbody>
</table>
The inter- and intraspeaker variation is discussed in more detail in article 4.
Chapter 6 Summaries of articles 1–4

**Article 1: Exploring the use of probes in a corpus pragmatic study of hedging strategies**

The majority of corpus studies of pragmatic phenomena deal with the functions of pre-defined forms. Moving in the opposite direction—searching for functions in order to identify the forms that can realise them—is impossible unless the corpus is annotated for pragmatic functions. This study explores a possible way around this problem: the use of probes. This strategy is tested as a means to identify hedging strategies in Norwegian and English spoken corpora. The probes *men* and *but*, signalling disagreement or contrast, are used as markers of face-threatening situations in which hedging strategies are likely to occur. The results show that clauses with *men/but* more frequently contain hedging than random clauses do, although the difference is statistically significant only for Norwegian. The use of probes thus seems to be a promising way forward, and future studies should aim at identifying even better probes with higher co-occurrence rates for the forms of interest.

Status: published

**Article 2: Classifying hedging strategies in Norwegian and English informal spoken conversations**

Hedging strategies, i.e. downtoning expressions or expressions of tentativeness or possibility, are ubiquitous in most, if not all, languages. However, hedging is often realised differently across languages, making such strategies particularly interesting to study from a contrastive perspective. The purpose of this study is twofold. Its primary aim is to compare the types of hedging strategies used in Norwegian and English informal spoken conversations. However, in order to enable a cross-linguistic comparison, the study also aims to propose a classificatory framework for hedging strategies. Most existing frameworks are based on specialised and formal forms of discourse, primarily in English, which may cause problems for contrastive studies of informal spoken discourse. The results show that a combination of categories from existing frameworks can be used to describe hedging strategies in informal conversations and that there are several significant differences in the type of hedging strategies used in the two languages.
Article 3: “I guess anyone would do that wouldn't they?” How do native speakers of Norwegian and English hedge in informal conversations?

Hedging is a complex phenomenon with an indefinite number of potential realisations. The complexity and versatility of hedging strategies make them particularly interesting to study across languages. This contrastive study compares the realisations of the pragmatic function of hedging in Norwegian and English everyday conversations by using data from four corpora of Norwegian and English informal spoken conversations (the Norwegian Speech Corpus, the Nordic Dialect Corpus, the BigBrother corpus, and the BNC2014). The results show that speakers of both languages mainly use pragmatic particles, adverbs, and first/second person pronoun + cognitive verb (1/2 pers. + Cog. V) to express hedging. Furthermore, English speakers use significantly more 1/2 pers. + Cog. V and modal verbs than Norwegian speakers, whereas Norwegian speakers use significantly more adjectives, prepositional phrases and clauses to hedge their utterances.

Article 4: Hedging in spoken conversations by Norwegian Learners of English

Hedging is an important aspect of pragmatic competence, but it is also a complex phenomenon that is difficult to master. This has resulted in underuse of hedging strategies by language learners, and many learners limit their hedging repertoire to a few strategies. This study compares the use of 10 hedging expressions commonly used in informal spoken English, *a bit, I mean, I think, just, kind of/kinda, like, might, probably, thing(s), and you know*, in data from LINDSEI-no, a corpus of Norwegian advanced learners, and LOCNEC, a comparable corpus of native English speakers. Norwegian learners typically show a high level of grammatical competence, but research on their pragmatic competence is limited. This study adds to the empirical research on Norwegian advanced learners by comparing the use of these expressions in native and learner language. The results indicate that as a group, Norwegian learners underuse hedging strategies, but when each expression and individual variation are considered, the picture is more nuanced. In fact, several of the Norwegian learners’ hedging practices partially overlap with several of the native speakers’ practices concerning hedging frequency and types of hedging strategies used.

Status: Submitted
Chapter 7 Summary of results and conclusions

This chapter presents a summary of the main findings of the project and the contributions of the thesis. Some limitations and suggestions for further work are also presented. The main findings are summarised in section 7.1. The contributions of the thesis to the fields of hedging and corpus pragmatics are presented in section 7.2. Limitations and suggestions for further research are outlined in section 7.3.

7.1 Summary of main findings

The overall aim of the thesis was to provide more insight into hedging as an aspect of pragmatic usage by studying linguistic hedging strategies in English and Norwegian informal spoken conversations. The thesis work was divided into three phases, each with its own operational objective and outputs.

The objective of phase 1 was to develop a methodological approach, i.e. a “workaround” (O'Keeffe et al., 2020, p. 9), to access expressions of the pragmatic function of hedging in Norwegian and English corpora of spoken conversations. Article 1 investigates whether the probe but/men, in its contrastive sense, can be used to identify speech situations in corpora where hedging strategies occur. It describes how the probe was chosen and compares the number of hedging strategies co-occurring with the probe to the number of strategies in randomly selected clauses. It was essential for this study that the same type of probe should be applicable across the languages studied (Norwegian and English) in identifying contexts that favour hedging strategies.

But/men is also register-neutral and frequent in oral conversations, which gave a substantial number of contrastive situations in the corpora. The probe was helpful in retrieving a variety of hedging strategies; however, the clauses with the probes only contained hedging strategies slightly more often than randomly selected clauses, and the difference was only significant for the Norwegian data. Nevertheless, but/men had the advantage over randomly selected clauses (without a probe) in eliciting the same type of contexts in
which hedging could be studied cross-linguistically. The results of the pilot study reported in Article 1 led to some refinements of the methodology, one of which was to expand the scope of study from clause to utterance/turn. The reason for this was that I had observed in the pilot study that there were hedging strategies modifying the contrast expressed outside of the immediate clause. The results of this refinement to the approach were that a larger number of relevant hedging strategies were included in the study.

The aim of phase 2 was to discover similarities and differences in the use of hedging strategies in Norwegian and English in order to increase our understanding of how hedging is expressed in the two languages. The results of these studies illustrated both similarities and differences between Norwegian and English hedging behaviour, concerning the types of hedging strategies used and the actual forms used. Article 2 compared types of hedging strategies in the two languages. In order to facilitate this comparison, a classificatory framework for hedging strategies in spoken informal discourse, which also took the two languages studied into consideration, had to be developed. Therefore, sub-categories from existing classificatory frameworks for hedging strategies were combined in a new way to account for the different types of strategies. The results reported in article 2 showed that both groups of speakers used a variety of different types of hedging strategies, ranging from hedging strategies with a propositional scope to hedging strategies with an effect on the relationship between the speaker and the proposition and the relationship between the speaker and the hearer. Native speakers of English used significantly more speaker-oriented hedging strategies than Norwegian speakers, whereas native speaker of Norwegian used significantly more hearer-oriented strategies. A potential explanation for this difference may be the frequent use of *like* to signal lack of commitment in informal conversations in English and the appellative aspect of many Norwegian modal particles often used for hedging purposes, such as *jo* (‘of course’, after all’).

The study in article 3 compared the different categories of forms used across English and Norwegian, and the results showed that both native speakers of Norwegian and native speakers of English use a range of different forms to express hedging. In all, 13 different categories of forms emerged from the data, and forms from all categories were used by both language groups. Forms from the category of pragmatic particles were particularly frequent in both languages. Approximately 40% of the Norwegian forms and approximately 35% of the English forms belonged to this category. The results also indicated several differences between the languages. For example, the Norwegian speakers used significantly more downtoning adverbs, such as *litt* (‘a bit’), and downtoning adjectives, such as *sånn* (‘such’) and *lille* (‘little’), than the English speakers did. The Norwegian speakers also used significantly more
prepositional phrases, such as på en måte (‘in a way’), and clauses det viste seg (‘it turned out’) than the English speakers. The English speakers, on the other hand, used significantly more first and second person pronoun + cognitive verb (1/2 pers. + Cog. V), such as I think and you know, and semi-/modal verbs, such as might, than the Norwegian speakers did. Such differences could cause problems for language learners. If the conventions for use of an expression in Norwegian are transferred into English, this could lead to over- or underuse of particular expressions.

The study in phase 3 of the project investigated the over- and underuse of frequently used hedges in a corpus of spoken English by Norwegian learners and a comparable corpus of English by native English speakers in order to gain insight into the pragmatic competence of Norwegian advanced learners. Previous research on learners of English with different first languages has indicated that learners tend to use fewer hedging strategies than their native speaker counterparts, and that they use a limited range of hedging strategies, but as leading researchers have pointed out more perspectives are needed to get the full picture. The purpose of the study reported in article 4 was to investigate whether Norwegian advanced learners behave in accordance with the results of previous research, or if their use of hedging strategies resembles that of native English speakers. By looking at both inter- and intraspeaker variation, the study provides more details than previous studies on Norwegian learners, in addition to focusing on a group of learners on which there has been limited research.

The 10 most frequently used forms identified in the English material in article 3, were used as the point of departure for the study. The results reported in article 4 showed that at group level, Norwegian learners hedged less frequently than native English speakers; however, this result was severely nuanced when individual variation was considered, both between the 10 forms and between the individual speakers. Only 5 of the 10 expressions were used significantly more frequently by native speakers, there were great differences in numbers and types of hedging used by individuals in both groups, and some learners and native speakers were similar in their profiles of hedging use.

In summary, the results of the four articles contribute to our understanding of how hedging is realised in informal conversations in Norwegian and English and by Norwegian advanced learners. The contributions of the project as a whole will be outlined further in section 7.2.
7.2 Contributions

The findings in this thesis offer important methodological and empirical contributions to the fields of corpus pragmatics, hedging research and learner corpus (pragmatic) research. The current section will address each of these contributions in more detail.

The most important methodological contribution of this project is the methodological “workaround” proposed to access hedging strategies in corpora. The need for studies on approaches that explore how pragmatic phenomena can be researched using corpora has been expressed by several leading researchers, such as Romero-Trillo (2008) and O'Keeffe (2018, p. 588), and most recently by O'Keeffe et al. (2020, p. 9). The form$_1$-to-function-to-form$_2$ is an attempt to create such an approach. The approach brings together aspects of (form-based) corpus linguistics and (function-based) pragmatics and builds on recent methodological developments within the field of corpus pragmatics (Huang, 2017b). As reported in article 1, the majority of corpus studies of pragmatic phenomena apply techniques which can be described as form-to-function, mainly because forms are easier to retrieve than functions in a corpus, unless the corpus is annotated for pragmatic functions. A downside to starting with a set of predecided forms is that one might miss out on other relevant forms used to realise a particular function, and thereby overlook important features of hedging in individual languages as well as cross-linguistic differences between them. The form$_1$-to-function-to-form$_2$ technique; however, circumvents the need for predecided forms and makes it possible to retrieve pragmatic functions from corpora, together with the forms that realise them.

The development of the form$_1$-to-function-to-form$_2$ approach facilitated a partly corpus-based, partly corpus-driven, contrastive function-to-form study which is the first of its kind. The results of the contrastive study are important for the understanding of differences between Norwegian and English. Kaltenböck et al. (2010, p. 2) emphasise that contrastive studies of hedging strategies are particularly relevant, because they “shine light on similarities and differences between languages and cultures.”

The results of the contrastive studies in articles 2 and 3 are important for our understanding of hedging in both Norwegian and English. The studies are, however, particularly important for our understanding of hedging in Norwegian, on which previous research is limited. The use of hedging strategies in Norwegian has so far been underexplored (Vold, 2006), and there is a need for empirical studies on how the pragmatic function of hedging is realised in the Norwegian language. The studies in articles 2 and 3 are the first ones to apply a function-to-form-inspired approach to the study of hedging strategies in spoken Norwegian. They also move beyond what has been the practice of pragmatic research in
Norwegian, which is the study of the meaning and functions of specific pragmatic markers mainly from a relevance-theoretical perspective. The importance of corpus-based contrastive studies in general has been underlined by Aijmer and Altenberg (1996, p. 6), who state that in addition to increasing our understanding of language-specific and cultural differences, they also provide “insights into the languages compared – insights that are likely to be unnoticed in studies of monolingual corpora.” By making use of the extensive literature and research on hedging in English, and compare the two languages, I have been able to describe the phenomenon of hedging in Norwegian better.

In addition to strengthening the body of research on hedging in Norwegian, studies of hedging strategies in other languages than English may also challenge the way hedging strategies are described. As discussed in article 2, the majority of classificatory frameworks used to describe hedging strategies are based solely on English data, which may reduce their general applicability. The results reported in article 2 illustrate that in Norwegian, the majority of hedging strategies are hearer-oriented, i.e. appeal to common ground between the speaker and the hearer. This type of hedging strategies is only covered by one of the frequently applied classificatory frameworks. The framework proposed in article 2 may be used to classify hedging strategies in spoken discourse in future studies, as well as be used as a point of departure for developing new and improved classificatory frameworks.

The empirical contributions of the studies reported in articles 2 and 3 also contribute to evening out the written-language bias in linguistics in general (Linell, 2005, 2019). As reported in section 4.2, the written bias in linguistic research remains, and there is still a need to explore properties of spoken language. Spoken language contains linguistic variables that are important for the accurate description of discourse, and there is a risk that such variables remain undiscovered if written language is the only mode studied (Adolphs & Carter, 2013, p. 1). Although hedging strategies are characteristic of both written and spoken discourse, the results reported in articles 2 and 3 indicate that the interpersonal aspect of hedging becomes increasingly important in spoken conversations. This thesis makes use of spoken data made available in recent years. The spoken BNC2014 is the first publicly-accessible corpus of its kind since the spoken part of the BNC was completed in 1994 and represents present-day spoken English. Similarly, the Norwegian corpora are from the early 2000s and can be taken to represent current usage.

Another empirical contribution concerns the investigation of hedging by Norwegian learners of English. The contribution of the study reported in article 4 differs from previous research on hedging in spoken English by Norwegian learners as it focuses on advanced learners of English. Previous
studies, such as Hasselgreen (2004) and Thomson (2016), have mainly focused on learners at lower secondary school level. The study in article 4 also provides insight into the use of more forms than previous form-to-function studies have investigated. Furthermore, the study gives a more detailed account of the hedging performance of advanced learners than many existing studies, as it looks at inter- and intraspeaker variation. The need to move away from only looking at frequency in learner corpus research studies has been emphasised by Paquot and Plonsky (2017) in their extensive assessment of quantitative research methods in LCR studies. The results reported in article 4 illustrate the importance of considering individual differences as well as the differences between individual expressions in the study of hedging strategies. The differences in over- and underuse of strategies indicate that there are various factors that influence learners’ use. The results also show that there is great variation in the way speakers hedge and that native-like is a concept which may need to be discussed or defined further.

The results presented in article 4 could have implications for how English is taught in Norwegian schools and universities. The broad spectrum of variation in hedging behaviour of native speakers is one aspect which needs to be considered in language teaching. The majority of Norwegian learners of English acquire English asynchronously through formal instruction, and although the teaching of pragmatic competence, including hedging, is difficult to implement in educational syllabi, cross-linguistic differences in hedging should be given more attention in language teaching.

7.3 Limitations and potential for further research

The main focus of this thesis has been to study the use of hedging strategies in informal spoken conversations through the application of a function-to-form approach, and the results of this work may provide opportunities for further work. Through the form1-to-function-to-form2 approach, I was able to retrieve a broad range of hedging strategies; however, through the probe I used, I only retrieved hedging strategies from one particular speech situation. There might be other strategies characteristic of other speech situations which can be retrieved using a different probe or a combination of probes. It might be worth looking for different probes that give a higher number of co-occurring hedging strategies, or to expand the study and look at more probes to give a better overview of the various realisations of hedging strategies in Norwegian and English. Furthermore, the approach involved a rather time-consuming, two-step manual sifting process, which meant that the size of the material had to be restricted. The starting point was 1,600 instances of but/men retrieved from the corpora, among which the contrastive uses had to be identified. Then, all co-occurring hedging strategies had to be retrieved
from the context of the probe. This two-step sifting process has to be taken into account when planning a study to ensure that it gives enough “value-for-money.”

This study has examined linguistic hedging strategies. As pointed out in section 2.1, hedging can also be signalled through the use of para-linguistic devices, such as gestures and facial expressions. To fully grasp how hedging in expressed in a particular language culture, one might need to study the full range of hedging strategies. Thus, provided that suitable multi-modal corpora can be found, it might be interesting to investigate how linguistic and para-linguistic strategies work together to create a hedging effect.

The study of hedging by Norwegian learners investigated the use of 10 expressions of hedging. Although these 10 expressions serve as an important starting point, more expressions should be studied to contribute to our understanding of the use of hedging strategies by advanced Norwegian learners. Furthermore, the study could be expanded to systematically investigate factors such as socio-pragmatic and pragmalinguistic transfer to contribute to the explanation of the Norwegian learners’ performance. The study could also be expanded by examining the LINDSEI corpora of learners with other L1s to see whether there are any differences or similarities across learner varieties.

Despite these limitations, the research presented has provided important contributions across several subfields of linguistics. The project has strengthened the body of empirical research on hedging and contributed to our understanding of how hedging is used in spoken informal conversations in Norwegian and English. Especially for Norwegian, it has provided a much fuller picture of hedging behaviour than previous studies. The study of learner language also gives much-needed nuances in our understanding of hedging by Norwegian learners of English. Such knowledge is beneficial for the development of new teaching material that may equip language learners with the necessary skills to avoid future communicative failure. The project has also contributed to the emerging field of corpus pragmatics, which is still in its infancy, by adding to the methodological toolbox.

I hope this thesis will encourage more investigations into hedging in spoken conversations in general, and hedging across languages and by language learners in particular.
References


Presentation at the International summer school – August 27–31, 2018, University of Bremen, Germany.


### Appendix 1 A working guide for analysis - English

<table>
<thead>
<tr>
<th>Form</th>
<th>Potential meaning/function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Just</strong></td>
<td>As an adverbial, not pragmatic marker:</td>
<td>Beeching (2016, pp. 77-78)</td>
</tr>
<tr>
<td></td>
<td>1. Precision with expression of time and place (<em>just before</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Recent past (<em>I've just described</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Adverbial <em>just</em> (<em>merely</em>) (<em>just a list of images</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Just</em> as a downtoner - related to the canonical sense of 'merely'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a, <em>can I just ask</em> - minimisation of additions, suggestions, criticisms and requests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b, <em>I just think that</em> - minimising assertions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c, <em>It's just that we have this issue</em> - minimising reasoning or justification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d, <em>Just remind me again</em> - hedged imperative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>e, <em>Just amazing</em> - Intensifier</td>
<td></td>
</tr>
<tr>
<td></td>
<td>f, <em>but it's just well I don't know I feel happier now</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of <em>just</em>: particulariser, temporal, emphatic (including downtoning) and planning.</td>
<td>Aijmer (2002)</td>
</tr>
<tr>
<td></td>
<td><em>just sort of</em> and <em>sort of really</em> show that <em>sort of</em> can have a polite downtoning function</td>
<td>Aijmer (2002, p. 189)</td>
</tr>
<tr>
<td></td>
<td>The most frequent functions of <em>just</em> in academic discourse is with a minimising or mitigating function.</td>
<td>Lindemann and Mauranen (2001)</td>
</tr>
<tr>
<td></td>
<td>Speaker may use <em>just</em> to minimise the significance of some process</td>
<td>Lee (1987, p. 378)</td>
</tr>
<tr>
<td></td>
<td>Pragmatic uses of <em>well</em>:</td>
<td>Beeching (2016, p. 51ff)</td>
</tr>
<tr>
<td></td>
<td>a, Hesitation = Could mean that the response is not the most relevant one or prefacing a dispreferred response.</td>
<td>Aijmer (2011)</td>
</tr>
<tr>
<td></td>
<td>b, Transitional <em>well</em> = used to gather up the consequences of what the previous speaker has just said and move on to the consequences of that remark. (Also referred to as narrative <em>well</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c, Changing the topic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d, Raising an objection</td>
<td></td>
</tr>
</tbody>
</table>
e. Prefacing a dispreferred response (replies that are not consonant with the hearer's expectations) related to c and d.

f. Taking over a turn/polite interruptions = used mid-conversation to interrupt and take a turn. *Well* eases the transition and makes the interruption less face-threatening.

e. Other-correction = used to introduce a correction of a misconception which is perceived by an interlocutor.

f. self-correction

g. quotative *well*

The main function of pragmatic marking *well* in ordinary conversation can be summed up as flagging a demurral, i.e. to hesitate, pause in uncertainty, or object.

*Well* is retrospective and prospective

Watts (1989)

A distinction is often made between left-hand markers and right-hand markers, *well* is an example of a left-hand marker.

*w*ell often introduces turns

Sacks, Schegloff, and Jefferson (1974)
Schiffrin (1987b)

*w*ell prefaces disagreements, alternating in this environment with *yes* and silence.

Schiffrin (1987b, p. 102)

a, exemplifying *like* = most akin to canonical *like* (similar to). However, the exemplifying use of *like* is weakening which leads to the possibility of *like* being used in other ways to hedge inadequacy of expression. Exemplifier *like* shares some of the features of a more purely hedging *like*.

Beeching (2016, pp. 127-134)
Andersen (2001)

b, approximative *like* = (glossed about) e.g. with numbers. In association with verbal expressions it resembles *sort of* which makes it possible for the speaker to distance him/herself from what is said. Andersen (2001) refers to this as the metalinguistic use of *like*.

*c*, quotative (be) *like* = suggesting that the person whose speech or thought is reported said something along the lines of what is then reported, indicating that the speaker does not take responsibility for what is said. This face-saving potential is almost lost as *be like* has become so common.

*d*, focuser *like* = discourse marker *like* may have a textual function, but it also modalises the text, downplaying
potentially dogmatic delivery and thus playing a face-saving role.

e, hedging *like* = is similar to focusing/discourse marking, but its main purpose is to hedge its potentially critical and emphatic stance the could be perceived as taking.

"we are left with the conclusion that *like* indicates looseness of meaning, or focus, or both"

Fuller (2003, p. 369)

Initial *like* may be used to convey a hedging stance in relation to the content or form of a following element. In these cases there is a focus on the imprecision of the comparison expressed by the markers. […] As hedges, the markers can be assigned a number of different paraphrases, including *approximately* and *for example*.

Hasund (2003, p. 123)

Clause-medial *like* - can be hedging, typically approximative, but also loose commitment on the part of the speaker in relation to the content/quote. It may also serve a metalinguistic function, i.e. focus. *Like* and *liksom* serve a primarily interpersonal function much more often in clause-medial position than in initial position.

Hasund (2003, pp. 171-177)

*Like* in final position can serve a hedging function, conveying a stance to the preceding expression, also reduced commitment on behalf of the speaker. (Not as markers of hesitation and speaker planning as initial and medial, naturally. Clause-final *like* can also have an intensifying function, but not as often as initial and medial

Hasund (2003, p. 199)

The core function of the pragmatic *you know* is to create common ground between the speakers

Beeching (2016, p. 97ff)

a, hesitation and appeal to common knowledge: invite the hearer to fill the gaps and co-construct meaning

b, word search and appeal to the interlocutor to fill in the gap: invite the collaboration of their interlocutor to find the right words.

c, clarification and appeal to common knowledge: *you know* can preface an explanation or clarification

d, attention-getting/launching a new piece of information: in initial position, *you know* attracts the listener's attention and points out a new piece of information. *You know* launches the proposition and does not appeal to shared communication.

e, direct appeal to shared knowledge/initiating a topic: Similar to d, but more direct and closer to the propositional meaning. Appeals directly to the knowledge of the interlocutor as a prelude to e.g. suggesting something new.

f, repair: *you know* is also used to repair syntactic reformulations where the speaker stops mid-flow and changes tack to reformulate a construction. Can appeal to common knowledge, initiate a clarification, etc.
In final position: pointing out a self-evident truth/impositional: point out a self-evident fact (with falling intonation). This you know is a powerful argumentative tool and can be impositional

You know is primarily used to express intimacy and rapport with the hearer.

The fact that general extenders and the discourse marker you know frequently co-occur seems to support the notion that, in using a general extender, a speaker is relying on an assumption of shared knowledge or experience.

Initial you know tends to be used to attract attention in general terms or to draw attention to something in particular which acts as a prelude to the main proposition.

Medial you know tends to be used to repair and to invite co-construction either of the content or the formulation of the message.

Final you know (when used with fallen intonation) tends to reinforce the message, implying that the proposition is self-evident. In final position with rising intonation, the message is portrayed as self-evident, but agreement is sought from the interlocutor.

Non-pragmatic/canonical uses: expressing an opinion, to mean is inflected

Emphatic and tag-forms which exhibit some of the characteristics of both non-pragmatic marker and pragmatic marking usages: quasi-canonical use (e.g. if you know what I mean) Such usages both have a signifying sense (canonical) and are considered to elicit agreement by appealing to presupposed shared knowledge.

Fully fledged pragmatic usages: I mean is used to establish and negotiate meaning with the hearer. Can serve multiple functions at the same time.

a, self-repair: where sequence 1 is followed by I mean + a corrected sequence 2

b, hesitation: frequently used in initial position as a filler and staller. Frequently with other pause fillers (um, er) when thinking of what to say next. It can also serve to introduce a clarification in these settings.

Beeching (2016)
c, clarification, exemplification, elaboration, reformulation: *I mean* is often used to link two sequences in which S2 clarifies, exemplifies, elaborates or reformulates S1.

d, justification: the speaker provides justification in S2 for the attitude expressed in S1, articulated by *I mean*. Can be used to move the argument to the speaker's advantage thus modifying his/her intentions.

e, concession and nuancing: *I mean* ... *but* - constructions. i.e. the speaker doesn't disagree introducing concession by using *I mean*, but comes back to his/her original position using *but*.

f, hedging: *I mean* can be used to soften "the strength of an evaluative comment", typical examples that start with *but I mean*:

> **you know what I mean** = generally considered to elicit agreement by appealing to presupposed shared knowledge

**Holmes (1995, p. 87)**

**Beeching (2016, p. 184)**

**initial I mean**: b, f

**medial I mean**: a, c, d, e

*Sort of* is considered to be a hesitation marker or a way of marking attitude and to function as an approximator, mitigator or as a lexical hedge similar to *kind of*.

*Sort of* can be used to avoid disagreement with the hearer because it avoids preciseness when this would be inappropriate.

*Sort of* may be an adjuster which can be used evidentially and affectively to flag imprecision and self-repair, and to tone down and hedge strong opinions.

*Sort of* "softens a strongly voiced opinion or the exaggerated expression of affect" and can be used to make opinions "safely vague" and to refer to common ground. It can be used as a negative politeness strategy to soften a request or a suggestion. It can also be used as a positive politeness device - a conventionalised mechanism for maintaining identity and establishing group affiliation.

*Sort of* differs from other adverbials expressing approximation, because its rock-bottom literal meaning is metalinguistic. [...] Hedges (e.g. *sort of, kind of, about*) involve a constant comparing of ongoing knowledge with expectations.

*Sort of* is a full-fledged discourse particle when it has the speech act within its scope and has downtoning function. Its main function in informal speech is interpersonal and it can be analysed in terms of politeness.

**Beeching (2016, p. 156)**

**Brown and Levinson (1987, p. 117)**

**Aijmer (2002, pp. 201-205)**

**Aijmer (2002, pp. 178, 181)**

**See also** Fetzer (2010)
The large number of examples of the collocation *sort of you know* provides strong evidence for saying that *sort of* can have interpersonal or affective meaning.

*Sort of* is regarded as ‘a particularly interesting pragmatic particle because of the wide range of meanings it may convey’.

A distinction can be made between the epistemic modal (evidential) and affective (interpersonal) function of *sort of*. In the evidential function *sort of* is used when the speaker avoids being precise for example because he does not know the right word or because there is a lexical gap. The affective *sort of* is hearer-oriented and closely associated with claiming common ground and avoiding disagreement.

*Sort of* modifies a range of different syntactic constituents: noun phrases (most common), verb phrases, prepositional, adjectival and adverbial phrases as well as single words such as numerals or onomatopoetic words.

The normal position for *sort of* was medially.

*Kind of* and *sort of* seem to have the same meaning.

“As far as I have been able to discover, these two items mean exactly the same thing”.

*Kind of* is not used to the same extent as *sort of* at least not on the British English scene.

*Kind of* and *sort of* may function as vague stance adverbs. More frequently used in spoken discourse compared to written discourse.

However, of these two, *sort of* was significantly more common in written mode than *kind of*, while the reverse was found to be true of the spoken mode.

*kind of* and *sort of* are both used more frequently in formal, academic prose than in other written registers.

**KIND OF**

*Modal adverbs*

The modal adverbs behave in a similar fashion to parenthetical verbs and serve approximately the same function that is to reduce commitment.

Modal adverbs can be subdivided into those expressing certainty and those expressing doubt, but they have in common that they express attitude of the speaker to the propositional content.

Change their position freely within the sentence.

**PROBABLY**

Meaning: ‘almost certainly’, ‘very likely’.
**Probably** expresses the speaker's strong doubt about the truth of the proposition.

Meaning: very likely, but not known for certain, what you think is the likely situation. *Presumably* expresses degree of doubt about the truth of the proposition. It is not that the speaker's attitude to the truth of the propositional content is not affirmative, but that he considers negation by the hearer to be a real possibility here.

**Presumably**

Meaning: that something is possible or that you are not certain about something.

**Perhaps**

*Perhaps* expresses the speaker's strong doubt about the truth of the proposition.

**May** expresses the speaker's strong doubt about the truth of the proposition.

Parenthetical verbs modify or weaken the claim of truth implied by the assertion.

When the verb is used parenthetically, it has no propositional status.

---

(Parenthetical) first person pronoun + verb

*I think*/*I believe* can be regarded as evidentials with unspecified evidence.

*I think* has gone a step further in the pragmatization process and developed into a discourse marker or modal particle which is syntactically a speech act adverbial.

A distinction can be made between two senses of *think*: ‘belief’ and ‘attitude’

*I think* as a pragmatic element expresses epistemic modality.

*I think* may express either certainty or uncertainty.

The proposition expressed in connection with ‘I’ will always be directly related to the person of the speaker by the hearer(s). Therefore, I + predicate combinations in discourse are self-revelations.

*I think* can occur in different positions in the utterance. The phrase is flexible with regard to tense, aspect and modality.

Its position seems to be important for the function of *I think*. When *I think* is placed at the end of an utterance, less certainty is signalled.
I think in medial or final position seems to be an effect of the speaker’s planning.

Parenthetical I think functions as an afterthought qualifying some information which has been given in the preceding part of the utterance.

The clitic I think is also a polite way of softening or downtoning a previous "head".

In initial position, I think can mark hesitation or informality.

Non-factive verb

expresses a stronger degree of commitment to the truth than I guess (Aijmer 1997:18)

I suppose corresponds to väl and jag antar in the Swedish parallel corpus (Aijmer 1997:32)

I FEEL / I FELT

I GUESS

I FIND IT

I AM NOT SURE

Understating adverbs/adjectives

Understaters "are used to underrepresent the state of affairs denoted by the proposition"

Expressing a matter of degree.

a bit of (a + singular quantifier + (of + noun) denote all quantities from a little to a lot.

Two meanings of a bit: 1, one which is neutral as to amount and 2, where there is a sense of small quantity. Where a bit of seems to indicate a large quantity it can be analysed as an understatement.

Vague quantifier + countable noun

a few = some, several

can have both positive and negative connotations (few vs. not many)

A BIT

A FEW

ALMOST

LITTLE

NOT REALLY

ABOUT
<table>
<thead>
<tr>
<th><strong>General extenders</strong></th>
<th><strong>Reference</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>General extenders constitute a class of expressions, pervasive in ordinary conversation, typically occur in clause-final position and have the basic form of conjunction + noun phrase. They are called <em>general</em> because they are non-specific and <em>extenders</em> because they extend otherwise grammatically complete utterances.</td>
<td>Overstreet (1999, p. 4)</td>
</tr>
</tbody>
</table>

**And** + NP = adjunctive general extenders.

**Or** + NP = disjunctive general extenders

**General extenders** typically occur in final position.

“Endocentric general extenders”: They orient the addressee towards previously agreed common ground.

“Exocentric general extenders”: They orient the addressee towards an alternative meaning outside the speaker-hearer cognitive agreed realm, allowing the addressee to complement the verbalised information with additional or alternative information.

Referred to as *set-marking tags* in Andersen (2010): Several pragmatic functions can be associated with this usage. The tag can indicate a wider semantic scope and a loose interpretation of the nominal phrase. Moreover that tag can be used to emphasise the impressive nature of the concept talked about.

Referred to as *vague tags* in Channel (1994)

Referred to as *vagueness tags* in (De Cock et al. 1997).

**AND ALL**

**THAT**
<table>
<thead>
<tr>
<th>Neustic indetermination as a device for forming hedges:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presuppositions are almost impossible to negate, everything else to do with the propositional content is negatable. [...] Where the speaker is virtually sure of what he says and does not therefore expect to be challenged by the hearer. On a formal level the simple assertory sentence corresponds to this (This performance was outstanding). The modality can also be marked, e.g. This performance, I think, was outstanding, and It was an outstanding performance, wasn't it? where they express reservations.</td>
</tr>
<tr>
<td>Hübler (1983, pp. 97-109)</td>
</tr>
<tr>
<td>See also</td>
</tr>
<tr>
<td>Fraser (1980, p. 349)</td>
</tr>
<tr>
<td>Quirk et al. (1985) for a contrastive view of tags.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tag questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>To the extent that questions are assertory, they express a qualified attitude on the part of the speaker to the (objectively-motivated) adequacy or the (subjectively/hearer-motivated) acceptability of the propositional content under consideration. This attitude can be described in general terms as doubt.</td>
</tr>
</tbody>
</table>
A distinction is made between two groups of tag questions, asymmetrical tags and symmetrical tags, i.e. those which have an opposite polarity to the preceding clause and those which have the same polarity.

Symmetrical tags are rare in English and they are characterised in the literature in a way which rules out their suitability as hedges right from the start.

The asymmetrical tags can be further subdivided into two groups:

a, those with (weak) rising intonation: imply a comparatively greater degree of uncertainty as to the validity and therefore requires express confirmation by the hearer.

b, those with (weak) falling intonation: the validity claim seems to be so great and confirmation anticipated to such an extent that such questions require no express confirmation.

Both can be used to form hedges? Assertory questions and tag questions the negatability of the sentence is clearly marked by the speaker (as opposed to in declaratives).

Positive politeness: redress directed to the addressee's positive face, i.e. his desire that his wants are thought of as desirable. Three main positive politeness strategies; claim common ground, convey that s and h are co-operators, fulfil H's wants for some X. Claim common ground can be done in three ways, of which intensify interest to H is one of them. Tag questions, as a way of drawing the hearer into the conversation.

Also presuppose/assert/raise common ground: personal-centre switch: where if H were S or if their knowledge was equal. (tag questions with falling intonation) include both S and H in the activity.

Negative politeness: Don't presume/assume: make minimal assumptions about H's wants --> question / hedge

Some pragmatic particles, e.g. väl (in Swedish) can be translated into tag questions to keep the appeal to the hearer.

A softer way of asserting something - depends on the intonation

Tag questions may be used to seek confirmation from the hearer

wasn't I?
didn't she?
don't you?
didn't you?
doesn't she?
wouldn't they?
isn't it?
weren't you?
dosen't it?
are you?
do you?
hasn't it?
wasn't it?
weren't they
didn't it
don't they?
doesn't it?
is it?
was he?
(you would think that) would you
doesn't he?

<table>
<thead>
<tr>
<th>Vague language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choosing a vague expression instead of a more precise one may have several motivations and does not necessarily entail any uncertainty on behalf of the speaker. Other possible motivations: as part of a cooperative strategy or of politeness reasons.</td>
</tr>
</tbody>
</table>

Andersen (2010, p. 35)

<table>
<thead>
<tr>
<th>Reasons for using vague language:</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Speakers tailor their utterances so that the right information is given, in accordance with Grice's maxim of Quantity</td>
</tr>
<tr>
<td>b. using an approximation may have the effect of focusing attention towards the most important part of the utterance - foregrounding</td>
</tr>
</tbody>
</table>

Channell (1994, p. 173ff)
c, deliberately withholding information: could be that the speaker does not want to commit himself, i.e. a defensive tactic. Could also be to deliberately be unhelpful.

d, use language persuasively e.g. presenting statistical data in a way that give the best support to the argument, e.g. up to 100%

e, to convey meaning in situations where they do not have the necessary words or phrases for the concepts they wish to express at their disposal. Vagueness is something speakers use when they cannot find the words they need

f, lacking specific information: when speakers do not remember

g, displacement: speakers use vague expressions when there is uncertainty about what they want to say

h, self-protection: vagueness being used as a safeguard against being later shown to be wrong

i, power and politeness: adhering to politeness rules of a culture and not threatening face

j, informality and atmosphere

k, characteristic for women's language ?? (not controlled for in the data)

Intentional vagueness occurs for a variety of reasons, e.g. the non-vague contribution might offend someone, the speaker doesn’t know the precise details, it doesn’t matter, or the intention is to evade.

Worry of being charged with being wrong, might give rise to the hedged responses.

Intentional vagueness also arises to accommodate memory loss, or when there is a desire to imply shared knowledge. Finally, there are occasions when the speaker deliberately increases the vagueness, using terms like and so on, in a way, or something like that, more or less, et cetera, in a sense, and things like that, or whatever, so to say, and so forth, somewhat, and everything, and stuff like.

One way of being vague is by means of nouns whose meaning is maximally inspecific, that is to say, they are essentially void of lexical features but can be used to refer to any object or phenomenon, e.g. thing.

Three functionally distinct types of use:

a, the place-holder: replacing a more specific item, cannot be removed without affecting the grammatical integrity of the
sentence/proposition. The speaker is signalling that the utterance is in need of pragmatic enrichment.

b, a set-marking tag: (og greier) external to the proposition (see general extenders, what is proposed here is the traditional understanding) the function is to signal that the noun phrase refers to a semantically open collection of things.

c, discourse marker: the thing is that has a textual discourse-structuring function

The most common structural realisation of vague nouns is as the head of a noun phrase or as the right most parts of a compound. Functionally these types of use are similar in that the noun is a lexically void substitute for a more specific referent. The vague nouns provide an incentive for the hearer to semantically enrich the proposition in order to arrive at the intended meaning of the utterance.

Stuff appear less formal than thing. 

Lots of / loads of: (plural quantifier + of + noun): approximating quantities with non-numerical vague quantifier

Some (weeks, money, etc.): Vague quantifier + countable noun: (some + noun): while some is semantically neutral, pragmatically it is not a neutral choice. Implying not a lot

Somewhere
at some point
for some reason
people / a lot of people
ages
a while
a number of
other bits and pieces

Modal (and semi-modal) verbs

Semantically, modal verbs are rather more complex (than grammatically) in that their individual readings can also be of a non-modal nature. Only those meanings which are modal can be considered for forming hedges.

Andersen (2010, p. 37) 

Channell (1994, p. 101)

Channell (1994, p. 114) 

Hübler (1983, p. 127)
<table>
<thead>
<tr>
<th>Might</th>
<th>may express doubt</th>
<th>Hübler (1983, pp. 133-134)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would say</td>
<td>can have a detensifying function</td>
<td>Hübler (1983, p. 132)</td>
</tr>
<tr>
<td>They seem to</td>
<td>Indicate moderate doubt.</td>
<td>Hübler (1983, p. 137)</td>
</tr>
<tr>
<td></td>
<td><em>seem</em> may indicate that knowledge is arrived at by inference or it can be analysed as a marker of hearsay.</td>
<td>Aijmer (1996, p. 398)</td>
</tr>
<tr>
<td></td>
<td><em>seem</em> also refers to general belief or opinion</td>
<td>Chafe (1986, p. 268)</td>
</tr>
<tr>
<td>Most people would</td>
<td></td>
<td>Aijmer (1997, p. 13)</td>
</tr>
<tr>
<td>Tend to be</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Necessarily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technically</td>
<td>Moving the responsibility over to technical specifications and away from the speaker</td>
<td>Gries and David (2007)</td>
</tr>
<tr>
<td>Basically</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apparently</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeah</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Though</td>
<td></td>
<td></td>
</tr>
<tr>
<td>You see</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No one I know</td>
<td></td>
<td></td>
</tr>
<tr>
<td>According to this book</td>
<td>The speaker prefaces his main thought with the possibility that he may be incorrect in what follows. What follows is nearly always a declarative with the force of a claim, judgement, diagnosis, etc. which in that context conveys an unwelcome effect.</td>
<td>Fraser (1980, pp. 347-348)</td>
</tr>
<tr>
<td>If I remember correctly</td>
<td>Another type of disclaimer. The force of the following sentence is generally restricted to a directive.</td>
<td></td>
</tr>
</tbody>
</table>
I am no expert

A mitigation marker: the message that follows is typically disadvantageous to the addressee and thus susceptible to mitigation

Fraser (1996, p. 184)

In our eyes

From my point of view

Really

Normally
Appendix 2 Working guide for Analysis – Norwegian

<table>
<thead>
<tr>
<th>Form</th>
<th>Glossing</th>
<th>Potential meaning/function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pragmatic markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>It is a well-known fact that the modal particles in German and Scandinavian languages typically occur in the middle field and that they may serve an interpersonal function in this position</td>
<td>Hasund (2003, p. 177)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pragmatic particles do not add to the content of the proposition, but to aspects relating to the relationship between the proposition and the interlocutors or to other propositions. They contribute to the procedural meaning.</td>
<td>Borthen (2014, p. 261)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pragmatic particles may indicate that what is being said is in contrast to what an interlocutor may presume, or that the speaker wants the interlocutor to confirm what is said.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modal particles have no effect on the truth condition of a sentence.</td>
<td>Andvik (1992, pp. 9, 12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modal adverbs may occur at the beginning of sentences, while modal particles may not</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Modal particles are however problematic. Their meanings shift from one modal domain to another with polysemy and fuzziness as a result. In addition, modal particles can have contrary meanings and express either weak or strong commitment to the proposition&quot;.</td>
<td>Aijmer (1996, pp. 395, 399)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Typical discourse meanings of the modal particles are 'evidence', 'emphatic assertion', 'counter-assertion' and 'expectation'</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Jo</em> signals that the proposition is a joint manifest between the interlocutors and that the proposition may be used as a premise for a decision.</td>
<td>Berthelin, Borthen, and Knudsen (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>According to the existing literature, <em>jo</em> indicates that the proposition is ‘given’ in some sense or other. Or that it indicates that something is ‘uncontroversial’.</td>
<td>Borthen (2014, p. 271)</td>
<td></td>
</tr>
<tr>
<td>JO</td>
<td>‘of course’, ‘after all’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The semantics of *jo* includes a constraint such that the utterance’s proposition is interpreted as a premise for deriving a conclusion.

*jo* is often used in contexts where the speaker argues against someone’s view; however, the data of Berthelin & Borthen does not support the hypothesis that a *jo*-warranted conclusion always counters an idea in the air. They propose that the semantics of *jo* includes a constraint on the interpretation such that the propositional content is interpreted as a premise for drawing a conclusion (in addition to the constraint on mutual manifestation).

A procedural semantics for sentence-internal *jo*: In an utterance that contains sentence-internal *jo* and is used to express the proposition *p*, interpret the utterance in light of the following procedures:

a. Constraint on higher-level explicature: interpret *p* as mutually manifest to speaker and hearer.

b. Constraint on implicature: interpret *p* as a premise for deriving and supporting an available conclusion *q*.

*Jo* conventionally implicates that there is consensus about the facts conveyed by the *jo*-clause.

*Jo* and *nå*: The claim is made that these constitute an attempt by the speaker to make recourse to an assumed or implied consensus, or commonality, between the speaker and the addressee.

The obvious basic feature of the *jo* particle is that it conveys that the information contained in the utterance is somehow already known to the addressee.

Perhaps the most frequent occurrences of *jo* are in clauses which in some way support a prior utterance by the same speaker. These are typically
used in arguments to advance the position of the speaker by giving support to his assertions.

There are four main categories of ways in which the *jo* particle may be used:

1. supportive
2. oppositional
3. concessive
4. contra-expectational

*Jo* expresses that the speaker considers the truth of the expressed proposition p to be ‘mutually known’ to the interlocutors.

The giveness aspect of *jo* is best understood as “mutual manifestness”.

*Da, nå* and *jo* may also function as boosters

The modal particle *jo* signals that the content of the proposition is familiar to the interlocutors and thus functions as a booster.

*(The Swedish) ju* and *väl* have developed evidential meanings indicating that the hearer is appealed to as the source of knowledge.

*Ju* and *väl* are extremely frequent in conversational interaction. This is explained by the fact that they have developed interpersonal or polite functions.

The modal function of *ju* is usually interpersonal. By emphasising that the speaker and hearer have some knowledge in common, *ju* may create a feeling of intimacy and rapport.

*da* may signal

a. that the propositional content is a conclusion on the part of the speaker
b. that the speaker questions and disagrees with something previously said or implied
c. that a request is perceived as more motivated than without *da*
d. modification of something which has been said

See also Fretheim (2014)
e, that the utterance is to be interpreted as an argument in an ongoing discussion.

*da* often signals a request by the speaker for confirmation from the hearer.

*da* is often used to express disagreement with a perception held by the hearer.

The booster *da* can occur in medial and final position. *Da* as a booster reinforces the speaker’s commitment to the truth of the proposition.

Final *da* does not necessarily indicate epistemic certainty.

*Da* in final position may also function as a turn signal.

<table>
<thead>
<tr>
<th>Da så</th>
<th>‘then so’</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ja</em></td>
<td>‘yes’</td>
</tr>
<tr>
<td>På en måte / på måte</td>
<td>‘In a way’, ‘sort of’</td>
</tr>
</tbody>
</table>

Quotative *liksom* in initial position: an indication that the following unit is an approximate rendering of what is in the speaker’s mind. Hasund (2003, pp. 123-125)

Initial *liksom* may be used to convey a hedging subjective stance in relation to the content or form of a following element. Hasund (2003, pp. 170-172)

<table>
<thead>
<tr>
<th>LIKSOM / LISSOM</th>
<th>‘like’</th>
</tr>
</thead>
</table>

a, serving a hedging function in relation to the following linguistic form.

b, serving a hedging function in connection with hesitation and speaker planning.

c, to soften a potential face-threat

medial *liksom*:
a, can be used to express hedging stance in relation to the content or form, usually an immediately following constituent, phrase or part of a phrase.

b, signals approximation e.g. in front of numbers.

c, expresses reduced commitment on behalf of the speaker when quotational.

d, serving a hedging function in connection with hesitation and speaker planning.

e, hedging for politeness purposes

**final liksom:**

final liksom conveys epistemic stance:

a, in relation to the content or form of the preceding expression.

b, expresses reduced commitment on behalf of the speaker in relation to the content of a quote.

c, politeness device in connection with sensitive topics or potentially face-threatening acts. (very common with liksom).

*Liksom* as a discourse marker has the same element of comparison and similarity inherent in non-discourse liksom. However, discourse marker liksom may indicate a comparison between entities that are not necessarily on the same syntactic level in the utterance. One of them might not be explicitly expressed but implicitly referred to e.g. ‘what’s in the speaker's mind’.

Quotative liksom frequently occurs after the quotation.

Quote-final liksom should be considered a hearer-oriented, appealing marker, inviting the hearer to visualize or imagine the dramatised situation.

May occur unstressed post-verbally

*Vel* signals uncertainty and in certain contexts an utterance containing *vel* may function in the same way as a question.

*Vel* is an alter-particle in medial and final position.

*Vel* is an alter-hedge.
Swedish väl is associated with an appeal to the hearer to confirm something or to agree in addition to the epistemic meaning of expressing some reservation with regard to the truth of the proposition.

The Swedish väl is placed immediately after the verbal operator but before other adverbs.

Väl may appeal to the interlocutor and thus often corresponds to the English tag questions.

While vel in post-verbal position carries a subtle appeal to the listener to confirm the statement, this appeal is much stronger, almost demanding when the particle is in sentence-final position.

*Vel* has two main functions: attenuate the content of a proposition and appeal to the interlocutor. The attenuation is often related to speaker uncertainty. The appeal to the interlocutor may have many explanations. These functions often overlap and it may be difficult to distinguish between them.

*Vel* may also be used to invite the interlocutor to take part in the conversation.

In final position *vel* may express emotions such as surprise, annoyance and impatience. In these cases *vel* does not express hedging.

*Vel* may also indicate that the speaker does not have enough information or knowledge to comply with Grice’s maxim of relevance.

*Vel* is an alter- and hearer-oriented hedge.

Väl signals not only uncertainty or weak commitment but may serve as a politeness in indirect requests. [...] Väl may show respect for the other person’s negative face-needs.

Nok signals uncertainty, but unlike vel, there is no appeal to the interlocutor when using nok.

NOK ‘probably’ Nok is an ego-hedge and is used to make the utterance less categorical.
By using *nok*, the speaker hedges the utterance, but simultaneously takes full responsibility for the utterance.

The modal particle *nok* is used as hedge to signal uncertainty on the part of the speaker. It may be described as a subjective particle, i.e. it attenuates subjective assumptions (as opposed to *visst*).

*Nok* and *vel* may be used to prepare the interlocutor for objections.

The Swedish *nog* may serve as an emphatic discourse marker (‘certainly’) or non-emphatic (‘probably’).

See also Aijmer (2016)

Aijmer (1996, p. 408)

Solberg (1990, pp. 50-53)

Solberg (1990, pp. 50-53)

The Swedish *nog* may serve as an emphatic discourse marker (‘certainly’) or non-emphatic (‘probably’).

The discourse use of *bare* can be seen as related to its non-discourse use as an adverb that is historically derived from an adjective meaning ‘bare’ or ‘naked’.

*Bare* can be used as a downgrading focus adverb indicating that something is low on a scale, either quantitatively (no more than) or qualitatively (no better than).

As a discourse marker *bare* can have an epistemic hedging function as well as an affective intensifying function. As a hedge *bare* reduces the speaker's commitment to the proposition expressed. This function is often exploited for politeness reasons, for example in order to reduce the imposition of requests. The function of *bare* as a hedge is considered as derived from its use as a downgrading focus adverb.

*Bare* shares certain similarities with the discourse marker *just* in spoken English, which is used to intensify an adjective or a noun.

Quotative *bare* typically signals heightened speaker involvement, may function to highlight a dramatic peak in a narrative, evaluative intensifier. There may be signs of semantic bleaching and *bare* having lost its intensifying function.

*Bare* indicates that something is low on a quantitative or qualitative scale.

*Bare* can be used as a focusing adverb and in a pragmatic sense.

Opsahl (2002, pp. 27-114)

Svennevig (2001b, p. 147)

Hasund et al. (2012, pp. 41-46)

Opsahl (2002, p. 111)
*Bare* may have an attenuating function and communicates the speaker's attitude towards the content of the proposition. *Bare* may also express the speaker's attitude towards the hearer or other parts of the world.

*Bare* is connected to modality. It may express reservations, but also boost certain expressions.

*Bare* may serve a contrastive downgrading function. *Bare* often follows a negative proposition. («det er ikke det at P, det er bare det at Q»). The sentence starting with bare does not necessarily follow immediately after the negative sentence.

This function is similar to that of *just*, i.e. when the speaker uses *just* to 'minimise the significance of some process' Lee (1987, p. 378).

*Bare* may express both epistemmic and deontic modality. (primarily deontic)

*Bare* as a hedge attenuates requests, which may be face-threatening.

*Bare* attenuates the content of utterances by underlining the fact that the utterance should not be interpreted as anything stronger than X. This attenuation is often connected to a form of contrast.

*Bare* may also attenuate the content of the proposition (and not just the force).

*Bare* may be used to re-establish and maintain symmetrical relationships between speakers. By claiming to know something about a certain topic, the symmetrical relationship becomes asymmetrical and hedges such as *bare* may be used to regain the symmetry in the relationship.

*Bare* occurs rarely in final position (unlike in Swedish)

*Bare* as a booster is typically found in evaluative utterances and rendering of events in a narrative.

---

*Det er bare det* ‘it is just that’
<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vet du, du vet, du veit, det vet</td>
<td>'you know’</td>
<td>In final position</td>
</tr>
<tr>
<td>Vet</td>
<td>'know'</td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>'you might say’</td>
<td></td>
</tr>
</tbody>
</table>

**Modal adverbs (expressions of uncertainty)**

<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visst</td>
<td>'surely'</td>
<td>Visst is an objective hedge. It is not speaker-oriented as nok or hearer-oriented as vet. It only affects the content of the proposition.</td>
</tr>
</tbody>
</table>
| | | The modal particle visst, at first glance is equivalent to the adverb tilsynelatende ‘apparently’.
| | | Visst may signal that the content of the utterance is something inferred by the speaker.
| | | Visst may also indicate that something is hearsay and that the speaker is not responsible for the content of the proposition.
| | | Visst may signal irony.
| | | Visst is also used when the speaker is talking about something he/she has done, seen or heard at an earlier point in time. |
| | | The Swedish visst can mark that the source of information is a third person.
| | | Visst can have a variety of modal and evidential functions. Discourse function: indicate the strength of an earlier assumption or expectation.
| | | Sikkert may occur unstressed post-verbally. Sikkert and liksom indicate the speaker's attitude towards the proposition. |
| | | In the Swedish parallel corpus säkert was often translated with surely. When surely is translated with an adverb (säkert), no appeal is made to the hearer. |
| | | Everything that is said about antagelig is also valid for the hedge sikkert. |
| | | May express possibility, signal politeness, and express a need for confirmation (in negative interrogative sentences) |
| KANSKJE | 'maybe’ | Can be in initial and medial position. |

Source:
- Fretheim (1981, pp. 91-92)
- Andvik (1992, p. 8)
- Solberg (1990, pp. 46-48)
- See also Borthen and Slinning (2014)
- Aijmer (1996, pp. 399, 409)
- Andvik (1992, p. 14)
- Aijmer (1997, p. 32)
- Fretheim (1981, p. 90)
- "NAOB – Det Norske Akademis ordbok"
**ANTAKELIG**

*Antagelig may be classified as an ego-oriented word (but with great uncertainty)*

Fretheim (1981, p. 90)

<table>
<thead>
<tr>
<th>Metacommments</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Det vil ikke si</strong></td>
<td>‘That wouldn’t say’</td>
</tr>
<tr>
<td><strong>Jeg vil ikke si</strong></td>
<td>‘I wouldn’t say’</td>
</tr>
<tr>
<td><strong>Har jeg tenkt på i ettertid i hvert fall</strong></td>
<td>‘I haven’t thought about [that] in retrospect in any case’</td>
</tr>
<tr>
<td><strong>Kan jeg tenke meg</strong></td>
<td>‘I can imagine’</td>
</tr>
<tr>
<td><strong>Kan du si</strong></td>
<td>‘You could say’</td>
</tr>
<tr>
<td><strong>For å si det sånn</strong></td>
<td>‘To put it like that’</td>
</tr>
<tr>
<td><strong>I det store og det hele</strong></td>
<td>‘overall’</td>
</tr>
<tr>
<td><strong>I og får seg</strong></td>
<td>‘In and of itself’</td>
</tr>
<tr>
<td><strong>For så vidt</strong></td>
<td>‘Supposedly’</td>
</tr>
<tr>
<td><strong>Det syns jeg ikke</strong></td>
<td>‘I don’t think’</td>
</tr>
<tr>
<td><strong>For meg</strong></td>
<td>‘For me’</td>
</tr>
</tbody>
</table>

(Parenthetical) first person pronoun + verb

<p>| <strong>JEG TROR / JEG TRUR / TRUR</strong> | ‘I think’ |
| <strong>TROR JEG / TRUR JEG</strong> | ‘Think I’ |
| <strong>Jeg trudde</strong> | ‘I thought’ |
| <strong>Jeg tror ikke</strong> | ‘I don’t think’ |
| <strong>Trur ikke det</strong> | ‘don’t think so’ |
| <strong>Vil jeg tru / vil jeg tro</strong> | ‘I would think’ |</p>
<table>
<thead>
<tr>
<th>Norwegian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeg synes / Jeg syns / Syns</td>
<td>‘I think, ‘I find’</td>
</tr>
<tr>
<td>Syns jeg</td>
<td>‘think I’</td>
</tr>
<tr>
<td>Jeg s-</td>
<td>‘I s-’</td>
</tr>
<tr>
<td>Tenker jeg</td>
<td>‘think I’</td>
</tr>
<tr>
<td>Jeg tenkte</td>
<td>‘thought I’</td>
</tr>
<tr>
<td>Jeg vet ikke / jeg vet ikke jeg</td>
<td>‘I don’t know’, ‘I don’t know I’</td>
</tr>
<tr>
<td>Jeg føler / følte jeg</td>
<td>‘I feel’, ‘felt I’</td>
</tr>
<tr>
<td>Jeg mener</td>
<td>‘I mean’</td>
</tr>
</tbody>
</table>

**Understating adverbs/adjectives**

<table>
<thead>
<tr>
<th>Norwegian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litt</td>
<td>‘a little’, ‘a bit’</td>
</tr>
<tr>
<td>Liten</td>
<td>‘little’</td>
</tr>
<tr>
<td>Lille</td>
<td>‘little’, ‘small’</td>
</tr>
<tr>
<td>Liten</td>
<td>‘small’</td>
</tr>
<tr>
<td>Litt (sånn)</td>
<td>‘a bit like that’</td>
</tr>
<tr>
<td>Ikke stort mer</td>
<td>‘not much’</td>
</tr>
<tr>
<td>Til en viss grad</td>
<td>‘to a certain degree’</td>
</tr>
<tr>
<td>Nesten</td>
<td>‘almost’</td>
</tr>
<tr>
<td>Ikke helt</td>
<td>‘not quite’</td>
</tr>
<tr>
<td>Så mye</td>
<td>‘so much’</td>
</tr>
<tr>
<td>Så</td>
<td>‘so’</td>
</tr>
<tr>
<td>Så å si</td>
<td>‘so to speak’</td>
</tr>
</tbody>
</table>

*Så we might eliminate (as modal particle) on the grounds that, even unstressed post-verbally, it is a degree adverb which modifies the predicate adverb or adjective, and not the clause as a whole.*

Andvik (1992, p. 13)

**Tag questions**

<table>
<thead>
<tr>
<th>Norwegian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ikke sant</td>
<td>In spoken Norwegian conversations it is common to attach <em>ikke sant</em> at the end of an utterance to</td>
</tr>
</tbody>
</table>

Svennevig (2008, pp. 127, 131-133)
appeal to the interlocutor by seeking confirmation, agreement or understanding from him/her.

In recent years, *ikke sant* (with fallen intonation) has also been used as a response in itself and as an independent utterance.

This use indicates acceptance and agreement. It also implies that the speaker uttering *ikke sant* has some knowledge or thoughts about the topic. It can confirm a previous utterance, confirm an implicit point or confirm the interlocutor’s opinion.

A similar development has been seen for *isn’t it / innit* with fallen intonation (Andersen 2001)

And for *eller hur* in Swedish (Brumark 2008)

### General extenders

<table>
<thead>
<tr>
<th>General extenders</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set marking tag (<em>og greier</em>) whose function is to signal that the noun phrase</td>
<td>Andersen (2010, pp. 38, 43)</td>
</tr>
<tr>
<td>refers to a semantically open collection of things, including the item explicitly mentioned.</td>
<td></td>
</tr>
</tbody>
</table>

Several pragmatic functions can be associated with the set-marking tag: a, indicate a wider semantic scope and a loose interpretation of the nominal phrase, or it can be used to emphasise the impressive nature of what is talked about.

- **Og sånn / og sánt**
  - ‘and stuff’,
  - ‘and things like that’

- **Og alt sammen**
  - ‘and everything’

- **Og alt mulig**
  - ‘and everything’

- **Eller sånt noe**
  - ‘or something like that’

- **Eller noe sånn**
  - ‘or something like that’

- **Eller noe sånt noe**
  - ‘or something like that’

- **Eller Andre ting**
  - ‘or other things’

- **Eller noe**
  - ‘or something’
Choosing a vague expression instead of a more precise one may have several motivations and does not necessarily entail any uncertainty on behalf of the speaker. A speaker may opt for a vague description as part of a cooperative strategy to avoid providing more information than is required in a communicative situation or as a politeness strategy to avoid sounding superior or pedantic.

Vague expressions generally can be used in a variety of communicative contexts, including, but not limited to:

a, the speaker does not know or wish to specify the exact nature of the concept.

b, the speaker knows the exact nature of the but does not know or remember the term.

c, the speaker knows both the nature of the concept and the term, but nevertheless chooses an inexact rendering of it. In such cases the vagueness marker may act as a means of marking a psychological distance to a linguistic expression.

One way of being vague is by means of nouns whose meaning is maximally unspecific, i.e. they are essentially void of lexical features but can be used to refer to any object or phenomenon, e.g. ting.

As placeholders, the nouns signal that the utterance is in need of pragmatic enrichment.

Maximally vague in meaning denoting 'an entity of some kind'.

Andersen (2010, pp. 37-41)

Ting / tingene ‘Thing / things’ Can be a conceptual placeholder.

Ting is virtually always used with the same vague sense as greie (in Andersen’s 2010 study).

Andre ting ‘Other things’

Dusteting ‘Stupid things’

Mellomting ‘between things’

Andersen (2010, pp. 35-42)
<table>
<thead>
<tr>
<th>Norwegian</th>
<th>English</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noe greier</td>
<td>‘Some stuff’</td>
<td>For <em>greie</em>, see <em>ting</em>. Maximally vague in meaning denoting ‘an entity of some kind’. <em>Andersen (2010, pp. 36-41)</em></td>
</tr>
<tr>
<td>Hele greia</td>
<td>‘the whole thing’</td>
<td><em>Greie</em> is less formal than the stylistically neutral <em>ting</em>. Can be a conceptual placeholder.</td>
</tr>
<tr>
<td>Folk</td>
<td>‘people’</td>
<td></td>
</tr>
<tr>
<td>Noe</td>
<td>‘stuff’</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>‘one’</td>
<td></td>
</tr>
<tr>
<td>Det meste</td>
<td>‘most’</td>
<td></td>
</tr>
<tr>
<td>En god del</td>
<td>‘a great deal’</td>
<td></td>
</tr>
<tr>
<td>Ditten_og_datten</td>
<td>‘this and the other’</td>
<td></td>
</tr>
<tr>
<td>Et eller annet</td>
<td>‘something or other’</td>
<td></td>
</tr>
<tr>
<td>En</td>
<td>‘one’</td>
<td></td>
</tr>
<tr>
<td>En plass</td>
<td>‘a place’, ‘somewhere’</td>
<td></td>
</tr>
<tr>
<td>Rundt omkring</td>
<td>‘round about’</td>
<td></td>
</tr>
<tr>
<td>Her og der</td>
<td>‘here and there’</td>
<td></td>
</tr>
<tr>
<td>Folk tror</td>
<td>‘people think’</td>
<td></td>
</tr>
<tr>
<td>Noe (annet, særlig, spesielt, etc.)</td>
<td>‘some’, ‘something’</td>
<td></td>
</tr>
<tr>
<td>Noen (andre, sånne, mennesker, etc.)</td>
<td>‘someone’</td>
<td></td>
</tr>
<tr>
<td>Den der</td>
<td>‘that thing’</td>
<td></td>
</tr>
<tr>
<td>Det / den / han derre</td>
<td>‘that thing’</td>
<td>Deictic adverb. When combined with <em>sann</em>, The pointing character of <em>sann</em> is further accentuated. <em>Hasund et al. (2012, p. 58)</em></td>
</tr>
</tbody>
</table>
Sånn is a speaker-oriented focus marker with a demonstrative or pointing function. The discourse marker sånn derives from a deitic demonstrative meaning 'such' or 'like this/that'.

As a discourse marker, sånn may be used as a hedge, particularly in situations when the speaker expresses some degree of uncertainty about whether the phrase used is the correct formulation or whether the facts themselves are correct.

The hedging function of sånn can be glossed as 'approximately like this' where 'approximately like' contains a comparative element as well as deictic element.

Quotative sånn can be characterised here as a quotative marker with a demonstrative or pointing function. The marker points to the quotation as an embodied reenactment of the dramatised event and is used with quotations that involve extensive prosodic staging devices.

Both discourse sånn and liksom share the same double semantic potential of expressing approximation and exactness.

The Swedish equivalent såhör has a mainly demonstrative function, inviting the hearer to get involved in the narrative and to imagine the reported events - an interpretation which is fully compatible with our discussion of sånn above.

Öqvist (2000) claims that sär is a rheme marker, introducing new or important information. It also has the pragmatic function of inviting the hearer to cooperate by mobilising socially shared knowledge in order to arrive at an interpretation of the following constituent.

Sånn has multiple pragmatic uses, e.g. signaling approximation and reinforcing something.

Sånn may indicate speaker uncertainty, both uncertainty concerning the choice of words and concerning the content.

Sånn may also be used to attenuate an utterance and make it less categorical. Sånn, sånt and sånne have overlapping functions.
Norwegian has some special particles that appear in several of the examples and that indicate some psychological distance from the referents mentioned. One is sånn, which literally translates as ‘such’.

According to Lie (2008), the particle sånn is used as a hedge either to indicate reservations regarding whether the lexical expression is fully correct or appropriate, or to signal that the referent or the lexical item may not be completely familiar to the speaker or the hearer.

<table>
<thead>
<tr>
<th>Sånn (kvote, voldsomt, etc.)</th>
<th>‘like’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modal (and semi-modal) verbs</strong></td>
<td>Epistemic meanings – the space between &quot;yes&quot; and &quot;no&quot;</td>
</tr>
<tr>
<td><strong>Kunne</strong></td>
<td>‘could’</td>
</tr>
<tr>
<td><strong>Måtte</strong></td>
<td>‘must’</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
</tr>
<tr>
<td>Når det er sånn</td>
<td>‘when it is like that’</td>
</tr>
<tr>
<td>Spesielt</td>
<td>‘especially’</td>
</tr>
<tr>
<td>For en gutt</td>
<td>‘for a boy’</td>
</tr>
<tr>
<td>Det visste seg</td>
<td>‘it turned out’</td>
</tr>
<tr>
<td>Det lønner seg ikke</td>
<td>‘it does not pay off’</td>
</tr>
<tr>
<td>Heller</td>
<td>‘either’</td>
</tr>
<tr>
<td>Uansett</td>
<td>‘anyway’</td>
</tr>
<tr>
<td>Folk jeg har snakka med</td>
<td>‘people I have spoken to’</td>
</tr>
<tr>
<td>Svennevig (2010, p. 178)</td>
<td></td>
</tr>
</tbody>
</table>
Appendices references


165
Article 1

Title: Exploring the use of probes in a corpus pragmatic study of hedging strategies
Author: Stine Hulleberg Johansen
Exploring the use of probes in a corpus pragmatic study of hedging strategies

Abstract

The majority of corpus studies of pragmatic phenomena deal with the functions of pre-defined forms. Moving in the opposite direction—searching for functions in order to identify the forms that can realize them—is impossible unless the corpus is annotated for pragmatic functions. This study explores a possible way around this problem: the use of probes. This strategy is tested as a means to identify hedging strategies in Norwegian and English spoken corpora. The probes men and but, signalling disagreement or contrast, are used as markers of face-threatening situations in which hedging strategies are likely to occur. The results show that clauses with men/but more frequently contain hedging than random clauses do, although the difference is statistically significant only for Norwegian. The use of probes thus seems to be a promising way forward, and future studies should aim at identifying even better probes with higher co-occurrence rates for the forms of interest.

1. Introduction

In corpus linguistics, the default approach to studying various linguistic phenomena is to move from forms to their functions, whereas in pragmatics, the norm is to move from pragmatic functions to forms (O'Keeffe 2018: 588). The challenge of studying pragmatic functions using corpora is that there rarely is a one-to-one relationship between form and function. Therefore the majority of corpus pragmatic studies have taken on a form-to-function approach, the danger of which can be that possible realisations are not discovered because the search is limited to the items decided on prior to the search. There have been some attempts to work in the opposite direction, but there is still a need to “consider how, whether, and how best” pragmatic phenomena can be studied using corpus linguistic methodologies (O'Keeffe 2018: 588). Investigating one such method is the purpose of this study.

The reason for preferring forms as a starting point in corpus pragmatic studies is mainly that “core features of pragmatics studies [...] are harder to catch with corpus methodology than lexical or morpho-syntactic features” (Taavitsainen and Jucker 2015: 12). One example of a core feature which is hard to catch is that of hedging (see further section 2.1). Hedging strategies can take almost any linguistic (or paralinguistic) form and hedging is not an inherent property of words or phrases (Stenström 1994). Thus, identifying hedging strategies in a corpus is challenging without it being annotated for pragmatic
functions, and the existence of pragmatically annotated corpora is still rather limited (Aijmer and Rühlemann 2015). This study explores the use of probes, here defined as a search to find other expressions “that cannot easily otherwise be called to mind” (Hunston 2002: 62), as a means of identifying hedging strategies (see section 2.3).

The English contrastive conjunction *but* and the corresponding Norwegian conjunction *men* will be tested as probes. The reason for choosing *but/men* is the assumption that expressing something in contrast or disagreement to what has been said, either by the speaker or by an interlocutor, is threatening to the speaker and hearer’s face (Brown and Levinson 1987: 66, 68) and potentially calls for some remedial action (see further section 2.3). By searching for a characteristic of this typically face-threatening situation in corpora of spoken conversations, hedging strategies are identified without limiting the search to pre-defined typical hedges. If the suggested approach proves successful, it could potentially open up possible pathways for more studies from the functional perspective and thus be a way around the problem of going from function to form. Furthermore, it could pave the way for more bottom-up contrastive studies.

The following research questions will be addressed in this paper:

RQ1a: Can a marker of a face-threatening situation, i.e. expressing contrast using *men/but*, be used as a probe to retrieve hedging strategies in corpora?

RQ1b: Do hedging strategies occur significantly more often in clauses with the contrastive *men/but* than in randomly selected clauses?

RQ2: Will this functional approach to retrieving hedging strategies work across languages (Norwegian and English)?

The research questions will be addressed in light of recent developments within the fields of hedging research and corpus pragmatics in section 2: Section 2.1 discusses the concept of hedging and how hedging has been studied previously, section 2.2 describes common approaches to corpus pragmatics, section 2.3 presents the use of probes and section 2.4 describes the probes selected for this study in more detail. Section 3 describes how probes have been applied in the present study, whereas sections 4 and 5 present and discuss the results of this application. Some concluding remarks are presented in section 6.
2. **Background**

2.1 **Hedging strategies**

Ever since hedging strategies became a field of interest in the early 1970s, there have been several attempts both to define and to classify them, but their unruly nature has made it challenging, and to this date there is no general agreement on either an exact definition or an appropriate classification system although researchers have expressed the need for such a system (Kaltenböck, Mihatsch, and Schneider 2010). Still, the conceptual understanding of hedging has changed since it first attracted scholarly attention. Hedging was originally seen as a semantic concept, and the initial focus was on establishing a separate class of hedges. The earliest studies considered hedges as words whose job it was to make things more or less fuzzy (Lakoff 1972: 195). This type of hedging, affecting the truth value of the proposition, has later been referred to as *propositional hedging*. Propositional hedging was later contrasted with *speech act hedging* (Fraser 1975), which refers to hedging on the illocutionary force of the speech act, i.e. modifying the speaker’s intention in producing an utterance. This twofold distinction gave rise to taxonomies accounting for hedging in both spoken and written text, e.g. Prince, Frader and Bosk (1982), who distinguished between two types of hedging strategies: hedging within the proposition (*His feet are sort of blue*) and hedging between the speaker and the proposition (*I think his feet are blue*). Similarly, Hübler (1983) distinguished between *understatements* and *hedges*. In his taxonomy, understatements concern the propositional content whereas hedges concern the speaker’s attitude.

Soon hedging shifted from being considered a semantic concept to a pragmatic one, and today most researchers agree that there are no restrictions on the forms that can be used as hedges (Clemen 1997: 242). Thus the topic of study has moved from *hedges* to *hedging*. This development is reflected in many of the definitions applied in current studies (see e.g. Farr and O'Keeffe 2002 and Fraser 2010). In this study, the definition proposed by Kaltenböck, Mihatsch, and Schneider (2010) is adopted. They define hedging as “a discourse strategy that reduces the force or truth of an utterance” (Kaltenböck, Mihatsch, and Schneider 2010: 1). *Discourse strategy* is understood here as a (linguistic) means of bringing about a desired result (Sanders 2015: 1). Hedging strategies can thus take almost any form and signal non-prototypicality, uncertainty on behalf of the speaker or mitigation to lessen the impact of the utterance. However, using a broad definition is not without its challenges. If hedging is regarded as a discourse
strategy, it may be difficult to determine exactly what in an utterance gives the hedging effect (Stenström 1994). Furthermore, discourse strategies may also entail gestures, body language, stress and intonation. Since this study uses corpora of transcribed spoken language, only linguistic elements that are transcribed in the corpora and that are used to express e.g. politeness, mitigation or vagueness (Gries and David 2007) will be considered. Such elements could typically be, but are not limited to, pragmatic markers (e.g. well/vel), adverbs expressing uncertainty (e.g. probably/muligens), epistemic modal verbs (e.g. may/kan), parenthetical verbs (e.g. I think/jeg tror), vague expressions (e.g. thing/ting) and general extenders (e.g. and stuff/og sånn). It is worth noticing, however, that hedging does not only occur on the word or phrase level. Even clauses or combinations of words, phrases and clauses may be used to create a hedging effect (Fraser 2010: 24; Salager-Meyer 1994: 154). The range of possible realisations which comes as a result of applying a broad definition is the main challenge when it comes to retrieving hedging strategies in corpora.

2.2 Corpus pragmatics

When describing types of corpus linguistic studies, a distinction between corpus-based and corpus-driven studies is typically made. In corpus-based studies the researcher typically forms hypotheses based on pre-existing theories which in turn are tested using corpus data (top-down), whereas in corpus-driven studies, corpus data is the source of new hypotheses (bottom-up)22 (Tognini-Bonelli 2001). In the field of corpus pragmatics, studies can be placed along the same continuum, but an additional distinction is made between form-to-function and function-to-form. The form-to-function approach starts from pre-defined lexical words or constructions (forms) whose potential pragmatic uses (functions) are examined (Aijmer and Rühlemann 2015). The function-to-form approach starts from a function and investigates the forms performing that function. Both the form-to-function and the function-to-form approach can be corpus-based and corpus-driven as shown in Table 1.

22 The author is aware of the challenges of using controversial terms such as corpus-based and corpus-driven, e.g. as discussed in McEnery and Hardie (2012). Here the terms are used to refer to ends on a continuum representing in broad terms either a top-down or bottom-up approach to the use of corpora.
Table 1 Form-to-function and function-to-form matrix

<table>
<thead>
<tr>
<th></th>
<th>Form-to-function</th>
<th>Function-to-form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Corpus-based</strong></td>
<td>Using pre-defined forms and investigating their functions using corpus data</td>
<td>Identifying pre-defined functions in a corpus and studying the forms which realize them</td>
</tr>
<tr>
<td>Testing and exemplifying theories and descriptions that were formulated before/without the use of corpora (Tognini-Bonelli 2001: 65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Corpus-driven</strong></td>
<td>Using corpus data to identify forms e.g. word lists etc. and then study their functions</td>
<td>Using corpus data to identify functions and then study their realizations, the “holy grail” (O’Keeffe 2018: 599)</td>
</tr>
<tr>
<td>Observations of the data lead to hypotheses (Tognini-Bonelli 2001: 84)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Up until now, the vast majority of studies of pragmatic functions using corpora have taken lexical items or morpho-syntactic structures as their starting points, e.g. Aijmer (1984) and Farr and O’Keeffe (2002). The tendency to use forms as a starting point is not surprising as corpora have traditionally been developed with the aim of electronically accessing linguistic forms in large language databases (Flöck and Geluykens 2015). Additionally, pragmatic functions “are not readily amenable to corpus linguistic investigations” (Jucker 2009: 273), as they are defined by their illocutionary force, i.e. the speaker’s intention, or their perlocutionary effect, the effect on the hearer, neither of which can be searched for directly in a corpus. Typically, pragmatic functions can only be identified automatically when they appear in routinized forms or in conventionalized combinations with Illocutionary Force Indicating Devices (IFIDs), i.e. devices that guide the hearer in understanding the intended illocutionary force, such as word order, performative verbs, stress, etc. (Flöck and Geluykens 2015), or as surface forms orbiting the function they perform, such as thank you as an expression of gratitude (Aijmer and Rühlemann 2015).

The few studies which have taken a function-to-form approach have typically either used close horizontal reading of small corpora or small samples of larger corpora, as in Tagliamonte and Hudson (1999) and McCarthy and O’Keeffe (2003), or studied pre-defined forms occurring in the form of IFIDs, as in Deutschmann’s (2003) study of apologies focusing on expressions containing words such as sorry, pardon, forgive, etc. Others have searched for metacommunicative expressions, such as variants of the word compliment in the study of compliments, e.g. Jucker and Taavitsainen (2014), or used output from Discourse Completion Tasks (DCTs) as starting points for corpus searches, e.g. Schauer and Adolphs
These studies range from purely function-to-form approaches to borderline form-to-function. Deutschmann’s study is an example of the latter. He searched for explicit apologies in the form of IFIDs in a sub-corpus of the BNC, went through all occurrences manually to identify the ones which were actual apologies, and finally studied the contexts of these apologies more closely. Using IFIDs to access the contexts of the apologies has led this study to be classified as applying a function-to-form approach by O’Keeffe (2018: 607). The study of Schauer and Adolphs (2006) is another example. They used a DCT to elicit gratitude expressions from native speakers of English. They used the results from the DCT as a starting point for a corpus investigation to study the expressions in actual language use. Jucker and Taavitsainen (2014) applied a different approach than the other studies mentioned here. By searching for variations of the word *compliment* they were able to study how compliments were talked about. Through the study of the extended context of the node they were also able to identify compliments, record information about complimenter/complimentee, types of compliment and compliment responses (O’Keeffe 2018: 611).

Although they have been classified as function-to-form, several of the above-mentioned studies use forms as a point of departure. Deutschmann and Jucker and Taavitsainen both started with forms that were part of those they were interested in. Similarly, the corpus part of Schauer and Adolphs’ study started with the forms they had found through the DCT. The present study differs from these studies by searching for forms (probes) that are not part of the hedging strategies I am looking for, but indicators of situations within which hedging is likely to occur. By using a probe to retrieve hedging strategies, the strategies themselves are not pre-defined and not restricted to routinized forms or surface forms. Thus the approach can provide examples of hedging strategies from a bottom-up perspective and can be described as form1-to-function-to-form2. The notion of probes will be further discussed in section 2.3.

### 2.3 Probes

The frequent mismatch between form and function and the lack of corpora annotated for pragmatic functions make it challenging to take on a function-to-form approach to the study of hedging strategies. Aijmer and Rühlemann (2015: 9) argue that the only way to locate realizations of functions in corpora is to search for surface forms *orbiting* the function in question (see section 2.2). However, when searching for an orbiting form or any conventionalised realization, you are moving towards a form-to-function approach again. Nevertheless, such conventionalised expressions may serve an additional
purpose as they can be used, not to study their own function, but to study other functions that tend to co-occur in their context, i.e. they can work as probes. A characteristic of form-to-function approaches is that the form being searched for is the form being investigated. A probe, however, is a search to find other expressions (Hunston 2002: 62). The use of probes is well-established in corpus linguistics and the probes used can be quite elaborate. Hunston (2002: 62) gives an example of how a probe can be used to investigate how men and women are evaluated, i.e. something/nothing + [adjective] + about/in + him/her. This probe would give a list of adjectives used in this particular phrase.

Syntactically, hedging strategies can occur in front, medial and final position and they are not restricted to a particular word class or phrasal structure. It could therefore be relevant to look at motivational factors for the use of hedging, rather than syntactic or semantic features, to identify an appropriate probe. Speakers use hedging strategies for a variety of purposes. However, politeness is often regarded as the primary motivation, particularly in spoken conversations (Nikula 1997). Therefore it is probable that hedging occur in situations where politeness measures are called for. Politeness is often linked to the concept of face, as it was originally described by Goffman (1955) and further developed by Brown and Levinson (1987). According to Brown and Levinson (1987: 62), face can be understood as basic wants of a person. They distinguish between two types of wants, i.e. components of face. Positive face is a person’s wish that her wants are desirable to others, whereas negative face is a person’s wish that her actions are unimpeded by others. A Face Threatening Act (FTA) is a speech act which runs contrary to the speaker and hearer’s wants (Brown and Levinson 1987: 65). The threat of an FTA can be mitigated in various ways, such as by performing the act indirectly or by applying positive or negative politeness strategies. Positive politeness is strategies that minimize the threat to the addressee’s positive face, whereas negative politeness is first and foremost redressive strategies to save the addressee’s negative face (Brown and Levinson 1987: 129). Hedging strategies may belong to both types of politeness, especially when defined in its broadest sense. Avoiding disagreement and asserting common ground, e.g. by using the pragmatic marker you know as a hedging strategy, is an example of positive politeness. Expressing uncertainty, e.g. through expressions like I think, could be used to minimize threat to both negative and positive face. An example of a face-threatening situation from the material investigated here is given in (1). In this situation, the speakers are talking about the weather and speaker B, in the third turn, mentions that a big rock has collapsed from the mountainside, implying that this is due to poor weather conditions and that this is problematic. Speaker A, however, objects to this. He/she does not object to the rock collapsing (thus the partial agreement), but the underlying assumption that
this is a weather-related problem. By expressing disagreement, the speaker is threatening the hearer’s positive face (Brown and Levinson 1987: 66).

(1)

A: [...] nei nå har det vært en fin periode altså

B: * ja det har _uninterpretable_ nå

B: men denne steinen som har drevet og rasa ned uti e Nautvika _uninterpretable_

A: * ja

A: **men** det er nå litt e # overdrevet trur jeg […]

A: [*] no now it has been a good period [pragmatic particle]

B: * yes it has _uninterpretable_now

B: but this rock which have been collapsing down in e Nautvika_uninterpretable

A: * yes

A: **but** that is a bit e # overrated I think

NDC volda_04gk><who_avfile volda_03gm-04gk

The concept of Face Threatening Acts has been instrumental in selecting probes in this study. Disagreeing or saying something that is in contrast to what has previously been said can threaten the interlocutor’s positive face, i.e. by disproving the interlocutor’s thoughts or opinions on some issue (Brown and Levinson 1987: 66). Even contradicting oneself is considered threatening to the speaker’s own positive face (Brown and Levinson 1987: 68). Consequently, identifying conventionalised realisations of contrasts may be instrumental in retrieving hedging strategies in a corpus. Due to the cross-linguistic nature of this study, the contrastive use of *but* and the corresponding Norwegian *men* have been chosen as probes. *But* and *men* have corresponding meaning, overlapping use and can be regarded as prototypical markers of contrast in both languages. Furthermore, the chosen probes behave syntactically similarly in the two languages, have more or less the same semantic prosody and occur frequently in spoken everyday conversations thus ensuring that the contexts of the probes are comparable.

2.4 But and men

Expressing contrast is one of the basic ways of connecting ideas, events and utterances (Rudolph 1996: 32), and it entails a notion of opposition and potentially also a broken causal chain. The relation of contrast can be divided into a variety of different subtypes which in turn can be expressed in a range of ways (see e.g. Quirk et al. 1985: 634 and Halliday and Matthiessen 2004: 541). The notion of contrast in general will not be discussed here, but the use of *but* and *men* as ways of signalling contrast will be
discussed in more detail. *But* and *men* have been assigned different labels depending on the perspective of study, e.g. conjunctions, discourse markers, connectives, etc. (e.g. Becher 2011; Fraser 1999). From a grammatical point of view, *but* and *men* are commonly classified as conjunctions connecting clauses, phrases or words that stand in contrast to each other (Faarlund, Lie, and Vannebo 1997: 25; Biber et al. 1999: 79). The contrast is typically either expressed explicitly or lies in the content of the connected clauses (Faarlund, Lie, and Vannebo 1997: 1138). It can also be inferable if the proposition violates the speaker’s expectations (Schiffrin 1987: 156).

The nature of the contrast implied may vary to a great extent. Blakemore (1989: 15) distinguishes between two main types of contrasts, the so-called “denial of expectation” and the “contrast”, i.e. semantic opposition, use. Blakemore (1989: 15) illustrates these uses with two examples, (2) being the denial of expectation use and (3) being the contrast use.

(2) John is a Republican but he's honest.
(3) Susan is tall but Mary is short.

In (2) there is no direct semantic opposition. The speaker assumes that all Republicans are dishonest, but the second part of the sentence rejects this conclusion by pointing to an exception. In (3) the speaker points to a difference in height between two people. In addition to marking an upcoming unit as contrastive, *but* and *men* can be used to modify or restrict a previous statement. They can express hesitation or an explanatory circumstance or reason (‘NAOB – Det Norske Akademis ordbok’). They can also be used to express an opinion that runs contrary to that of the interlocutor or to refute a statement or reject a suggestion (Biber et al. 1999).

From a discourse perspective *but* and *men* are often described as contrastive discourse markers (Fraser 2013; ‘NAOB – Det Norske Akademis ordbok’). When *but* is not used in combination with other discourse markers, it typically conveys contrast, contradiction, challenge, topic change or apology (Fraser 2013: 322). Furthermore, *but* has been said to be undergoing a grammaticization process moving towards an even broader spectrum of discourse marking uses, i.e. from having a turn-continuing function to having a turn-yielding function (Mulder and Thompson 2006). *Men* can also have a purely text-organizing function, e.g. *Men ellers da?* (gloss. *But otherwise then?* = a way of moving away from the topic just discussed). Other non-contrastive uses of *men* typically include expressing irritation, e.g. *Men se deg for da!* (gloss. *But watch out!* or *men* used to express surprise, e.g. *men i all verden!* (gloss. *But in all the world* = expression of surprise). The *men* in this latter example is similar to the English expressions *oh* or *wow*. *Men* can also function as a hedging strategy
on its own, particularly in clause final position, e.g. *Det var ikke det jeg mente da, men* (gloss. *It wasn’t what I meant then, but*). In my investigation I have rarely found *but* in final position performing the same function. However, Mulder and Thompson (2006) argue that, in Australian and American English, *but* in final position can signal that a turn is completed. Such non-contrastive uses are not relevant for this study.

Although the degree of correspondence is high, there are also several challenges of using *but* and *men* as probes. First, they do not always express contrast, as they both serve a variety of pragmatic functions as well. This means that non-contrastive uses have to be removed manually, which can be a strenuous task in large datasets. Moreover, the contrast expressed by *but* and *men* is not always easy to identify. In this study, the co-text of each use of *but/men* was carefully studied in order to identify the nature of the contrast, i.e. what was contrasted, modified, objected to, etc. thus vertical and horizontal readings were combined (Aijmer and Rühlemann 2015). The occurrences of *but/men* which were either clearly non-contrastive or where it was impossible to determine from the co-text were excluded from the study. This extensive horizontal reading could vote in favour of using a more clearly contrastive probe, such as *however*. Still, *however* poses a new problem as it is much less frequent in spoken conversations (Biber et al. 1999: 565). The corresponding Norwegian expression, *imidlertid*, does not occur at all in any of the spoken corpora in this study. *But* expressing contrast, on the other hand, is more common in conversation than any of the other registers studied by Biber et al. (1999: 82). Another challenge with using *but* and *men* as probes is that the contrast expressed is not always perceived as face-threatening. Whether something is perceived as face-threatening or not depends on various different factors, such as the relationship between the interlocutors, the situation in which the utterance is being expressed and the content of the utterance. I will return to this issue in section 5.

In this study, *but* and *men* were mainly considered contrastive when they denote denial of expectation, opposition of two elements (antithetic use), modification or restriction of a previous statement, or when a speaker objects to something said by another speaker. The examples in section 4 illustrate some of these types of contrasts. For example, in (4) speaker B objects to something speaker A is saying by modifying her initial agreement *ja/yes*. *Men* in (4) introduces a view which disagrees with what A is saying. Disagreement is associated with disapproval (Brown and Levinson 1987: 66) which is threatening to the hearer’s positive face. In example (8), the *men* signals denial of expectation similar to that expressed in (Blakemore 1989). The expectation is that everyone who was there saw what was put in the pot, but contrary to this expectation, the speaker did not see it.
3. Material and method

3.1 Material

The corpora used in this study were chosen based on their degree of comparability and their availability, as well as the conversational nature of the language they contain. All of the corpora include spoken dialogues between family members, friends, acquaintances and strangers and are sources of natural conversations on everyday topics which are relevant for the study of pragmatic phenomena.

The English data is from the Spoken British National Corpus 2014 (BNC2014) (Love, Dembry, Hardie, Brezina, & McEnery, 2017). The BNC2014 is an 11.5 million word corpus publicly released in 2017 and contains transcribed informal British English conversations recorded between 2012 and 2016. The situational contexts of the recordings are mainly casual conversation among friends and family members in various settings, recorded by the speakers themselves in their natural environment.

The Norwegian data is collected from the Norwegian part of the Nordic Dialect Corpus (NDC) (Johannessen, Priestley, Hagen, Åfarli, & Vangsnes, 2009), the Norwegian Speech Corpus (NoTa) ("Norsk talespråkskorpus – Oslodelen,")) and the BigBrother corpus (BB) ("BigBrother-korpuset,"). These three corpora are the only available corpora of spoken Norwegian conversations. Since they are smaller than the BNC2014 (the relevant parts used here amount to about 2.1 million words), and to better reflect the composition of the English data, they were all used as sources of data.

The NDC is a corpus of Norwegian, Swedish, Danish, Faroese, Icelandic and Òvdalian spoken language. It consists of spontaneous speech data from dialects of the North Germanic languages across all of the Nordic countries. The Norwegian part of the corpus consists of interviews and spontaneous conversations between family members, friends, strangers and acquaintances. In order to make the data as comparable as possible, only the data from the conversation part is used. The recordings were made between 2006 and 2012, involve 422 different speakers from different parts of the country and total approximately 1,120,000 words.

NoTa was compiled between 2004 and 2006 and contains 957,000 words. The corpus is made up of spontaneous spoken conversations and interviews from which again only the conversational part is used. The participants are between 16 and 85 years of age and are all from the Oslo region. The
conversation part of the corpus involves 127 different speakers and contains approximately 540,000 words.

The BB corpus consists of transcribed spoken data from the first season of the BigBrother reality show in Norway in 2001. Although the setting is somewhat unusual, the corpus contains naturally occurring language over an extensive period of time. There are 12 participants between 23 and 26 years old from different parts of Norway. The corpus contains approximately 440,000 words.

Table 2 summarises some of the metadata for the clauses investigated in the corpora to show their degree of comparability.

Table 2 Metadata for the data analysed in this study

<table>
<thead>
<tr>
<th></th>
<th>BNC2014</th>
<th>NDC</th>
<th>NoTa</th>
<th>BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age distribution</td>
<td>9-91</td>
<td>15-88</td>
<td>15-85</td>
<td>22-36</td>
</tr>
<tr>
<td>Gender distribution</td>
<td>61 % female / 39 % male</td>
<td>57 % female / 43 % male</td>
<td>59 % female / 41 % male</td>
<td>42 % female / 58 % male</td>
</tr>
<tr>
<td>Relationship between the interlocutors</td>
<td>71 % close family and friends / 24 % friends and wider family / 13 % acquaintances / 3 % colleagues / 1 % strangers</td>
<td>9 % family / 26 % friends / 59 % acquaintances / 6 % strangers</td>
<td>12 % family / 64 % friends / 19 % acquaintances / 5 % strangers</td>
<td>The participants are strangers who get to know each other and establish various kinds of relationships throughout the recording period</td>
</tr>
</tbody>
</table>

3.2 Method

In order to test the probes, 150 clauses with *men* and 150 clauses with *but* were compared to 150 randomly selected Norwegian and 150 randomly selected English clauses from the corpora. Each instance of *men* and *but* was controlled manually to ensure that it was of a contrastive nature and thus potentially introducing a face-threatening situation (see section 2.4 for criteria). When an instance of
men or but was not classified as contrastive, it was replaced by another randomly chosen instance with the probe. This was done with 11 occurrences of but and 35 occurrences of men. The reason for simply excluding non-contrastive uses was that only contrastive uses of men and but were relevant as probes. But and men themselves were not the subject of study. The randomly chosen clauses without probes were retrieved through searching for the verb tag, first of all because it is impossible to search for nothing in a corpus and secondly because most utterances include some form of verb. In this way, minimal utterances, such as yes/no-answers etc. were mostly excluded, which made the stretches of text more comparable.

The Norwegian data was extracted through the Glossa interface\textsuperscript{23} whereas the English data was extracted through the CQPWeb interface\textsuperscript{24}. Following the retrieval of the 300 instances of contrastive men and but, co-occurring hedging strategies were identified. Only hedging strategies to the right of the node and within one clause were considered. It is difficult to identify sentences and other grammatical units in spoken corpora as punctuation is often absent, thus the smallest grammatical unit that can express a proposition was chosen as the scope of study. Restricting a unit of study was also crucial to establishing the control units; otherwise it would have been impossible to know what to compare. As the Norwegian data came from several corpora, 50 random units with men and 50 random units with [verb] were chosen from each corpus, making the total 150 clauses with probe and [verb] respectively.

4. Results

4.1 Raw frequencies and qualitative analysis

The raw frequencies show that the clauses with both men and but contained hedging more frequently than the randomly selected clauses did. However, the difference was greater for the Norwegian data than the English data. The raw frequencies are given in Table 3.

\textsuperscript{23} http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/glossa/index.html
\textsuperscript{24} https://cqpweb.lancs.ac.uk/usr/index.php?thisQ=login\&uT=y
Table 3 Raw frequencies of clauses with and without hedging in Norwegian and English

<table>
<thead>
<tr>
<th></th>
<th>men</th>
<th>random Norwegian</th>
<th>but</th>
<th>random English</th>
</tr>
</thead>
<tbody>
<tr>
<td>with hedging</td>
<td>85</td>
<td>66</td>
<td>73</td>
<td>67</td>
</tr>
<tr>
<td>without hedging</td>
<td>65</td>
<td>84</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>sum</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
</tbody>
</table>

Although the raw frequencies show that the clauses with probes overall contained hedging strategies more often, only 56.7 % and 48.7 % of the contrastive instances of *men* and *but* respectively co-occurred with hedging strategies. Examples (4) and (5) show how hedging strategies typically co-occurred with *men* and *but* respectively. Only the immediate clause following *men* and *but* was considered in the categorisation of the clause, i.e. with or without (a) hedge(s), but in the examples given below, more context is included to create a better understanding of the dialogue.

(4)²⁵

A: [...] # det lønner seg egentlig å være litt streng på IT-en for da # lærer du deg du har jo uttelling for det # nå

B: *ja

B: ja # *men* a- de var *jo* d- m de var *jo litt* streng det var *bare* at # gikk ikke så bra liksom _latter_ sånn karaktermessig

A: [...] # it pays off to be a bit strict in IT-class because then # you learn you benefit from that # now

B: *yes

B: yes # *but* a – it was [pragmatic particle] d - m they were [pragmatic particle] a bit strict it was just that # didn’t go that well like _laughter_ like grade-wise

NoTa 020><who_avfile 019-020

In (4) speaker B (partially) objects to what speaker A is saying by using the objecting or intervening *men* and modifies her objection with various hedging strategies. The pragmatic marker *jo* can function as a way of indicating that what is said is shared knowledge between the speaker and the addressee. *Litt* (*a bit*), *bare* (*just*) and *så* (*that*) function as modifying expressions reducing the impact of parts of

²⁵ The # symbol indicates a short pause when speaking. The * symbol indicates overlapping speech. The symbols used in the original transcription are kept in these examples.
or the objection as a whole. *Liksom (like)* reduces the commitment of the speaker to the proposition, whereas *sann (like)* has an approximative function, indicating that the term *karaktermessig (grade-wise)* might not be the right term.

In (5), the expression *kind of* also has an approximative function indicating non-prototypicality.

(5)²⁶

A: [...] I found the local people there
B: >> they were just really trying to rape you
A: >> to be
B: >> mm
A: real assholes compared to
B: >> and then we found them
A: Mexico sou- there were really nice people but Guatemala we actually got *kind of* robbed at machete point in [...]  

BNC2014 SUVQ

Examples (6) and (7) show hedging strategies in the random instances in Norwegian and English respectively.

(6)

A: at n ikke er sur?
B: mm
C: ja at han ikke er *sann e sær* # som han pleier å være
D: *sann grinete og sær

A: that he is not moody?
B: mm
B: yes that he is not like e weird # like he usually is
D: *like cranky and weird

B B Anette><who_avfile 72

*Sann (like)* and *e* in (6) are interpreted as hedging within the proposition. Speaker C is either not certain that the term she is using to describe the person is the correct one, or there simply is no appropriate term, so she chooses the closest one in meaning and marks this non-prototypicality with the approximative *sann (like)* and the hesitation marker *e.*

²⁶ The >> symbol indicates overlapping speech. The symbols used in the original transcription are kept in these examples.
(7)\(^{27}\)
A: oh I know (...) I do need to finish it (.) I need to get into it
B: \textit{I just think er} if you if you did give it \textit{like} a couple of hours you you 'd
A: yeah I know

BNC2014 SESD

In (7) speaker B is suggesting that speaker A does something, which is a face-threatening act, as it threatens the negative face of the hearer, i.e. obstructing his freedom of action (Brown and Levinson 1987). The speaker thus uses the expression \textit{I think}, which says something about the speaker’s commitment to the truth of the proposition and \textit{just} which has a downtoning effect. \textit{Like} in this example is somewhat ambiguous and can either be used in an exemplifying sense or in a hedging sense (Beeching 2016: 128).

4.2 Statistical evaluation and quantitative analysis

Although the raw frequencies indicate that there is a difference, a Pearson’s Chi-squared test was performed in R to test whether clauses with the contrastive \textit{men/but} contained hedging more often than random clauses did. First the total of \textit{men} and \textit{but} co-occurring with hedging strategies were compared to the total of random Norwegian and English clauses with hedging.

\begin{table}[h]
\centering
\begin{tabular}{l|c|c}
\hline
 & \textit{Men + but} & random (Norwegian + English) \\
\hline
with hedging & 158 & 133 \\
without hedging & 142 & 167 \\
sum & 300 & 300 \\
\hline
\end{tabular}
\caption{Total of \textit{men} and \textit{but} compared to the total random units}
\end{table}

The difference between the two totals proved to be significant ($X^2 = 3.8435$, $p < 0.05$) with a p-value of 0.04994. This could indicate that the approach of using a probe to retrieve hedging strategies is successful. However, the significance is marginal and the two languages need to be considered separately before anything can be concluded.

\(^{27}\) The (...) symbol indicates a short pause.
For the Norwegian data, the difference between the number of men-clauses and random clauses containing hedging strategies was significant ($X^2 = 4.3202, p < 0.05$) with a p-value of 0.03766. This indicates that using the contrastive men as a probe to retrieve hedging strategies is a valid methodological approach. However, although the English data also showed a difference in the number of but-clauses and random clauses containing hedging, this difference was smaller than for the Norwegian raw data and was statistically non-significant ($X^2 = 0.33482, p > 0.05$) with a p-value of 0.5628.

![Figure 1 Frequencies of clauses with and without probes containing hedging in Norwegian and English](image)

5. Discussion

As seen in section 4, clauses with the probes but and men more frequently contained hedging than clauses without the probes. Research question 1a can thus be answered in the affirmative: these probes can be used to retrieve hedges. However, the difference between clauses with and without probes was rather small, and the statistical analysis showed that it was significant only for Norwegian, thus leaving us with a somewhat inconclusive answer to the other two research questions. It could be that the study is too small to make any firm conclusions and that the difference does not reach statistical significance for English with only 150 instances, but that it might do so in a larger dataset. Still, there are many
instances of hedging even in the randomly selected clauses, and the gain from using probes was not as great as expected. Moreover, one of the criteria for the choice of probes in this study was that they were comparable across the languages investigated, in order to establish a tertium comparationis. The degree of correspondence between but/men here makes sure that the study is comparing like with like. In a mono-lingual study, there would be fewer such restrictions on the choice of probe.

It is not surprising that hedging strategies can be found with random clauses as well as with the probes men and but, as hedging is a characteristic of spoken interaction overall (Aijmer and Stenström 2005) and can be used in various settings. However, as the primary motivation for hedging in spoken discourse is said to be politeness, it was expected that the number of hedging strategies following a contrastive men or but would be higher. One explanation for the relatively high number of hedging strategies in the random clauses could be that there were several instances of other face-threatening acts in the data. As seen in (7), speaker B suggests or recommends that speaker A should set aside a couple of hours to finish a task. This can be perceived as a face-threatening act, restricting the hearer’s freedom (Brown and Levinson 1987:66). At the same time, the contrastive uses of men and but were not always as face-threatening as one might assume. In example (8), speaker B uses men in a contrastive sense, expressing denial of expectation, but does not use any hedging strategies, presumably because she perceives no serious face threat. This illustrates the important point that although expressing an opinion that is in contrast to something you yourself or your interlocutor has said is face-threatening according to Brown and Levinson (1987), the magnitude of the threat is context-dependant, and expressing contrast is not necessarily face-threatening in all contexts. Similarly, in example (4) speaker B expresses partial agreement, i.e. yes, but. In this utterance, speaker B chooses to use hedging strategies to modify her partially contrastive opinion, but in other cases the speaker may deem them unnecessary. The degree of agreement or disagreement could determine whether or not the speaker chooses to opt for hedging strategies.

(8)

A: bra den e brukte dere olje eller brukte dere smør når dere # skulle # woke det der ?

B: jeg veit ikke hva det var ikke jeg som gjorde det # men jeg var der _laughter_ men jeg så ikke hva de hadde i

A: good it e did you use oil or did you use butter when you # were to # wok that ?

B: I don’t know what it wasn’t me who did it # but I was there_laughter_ but I didn’t see what they put in
Another reason for the small difference between clauses with and without probes might have to do with the scope of analysis. In this study, only instances with hedging strategies within the same clause were registered. This was to ensure comparability between the clauses with *men/but* and the random clauses. However, as can be seen from several of the examples above, hedging strategies can also appear outside a clause or in a following clause and still have an effect on the utterance as a whole. Had the scope of analysis been expanded, there might have been an effect on the statistical analysis. In example (9), the hedging expression *I mean* was not counted because it was not part of the clause with the contrastive *but*. Still, this expression has an effect on speaker A’s utterance as a whole. This is also evident from example (4) above, where the speaker uses several hedging strategies in the clauses following the clause with *men*, which can be said to be relevant for the contrastive statement as such.

(9)

A: *[…] when dad died they brought her up to the funeral and erm you know they look after her it’s not they’re not there sort of every day or anything
B: no but still yeah that’s
A: >> **but** they do they do look after her and *I mean* --ANONnameF was there in oh September

BNC2014 SRBZ
This illustrates the close interdependence between hedging and context (Kaltenböck, Mihatsch, and Schneider 2010), which causes challenges when studying pragmatic phenomena with the use of corpora. As Adolphs (2008: 3) points out,

> It seems questionable that the same techniques developed for written corpus analysis should be sufficient or appropriate for exploring spoken corpora, not least because discourse is an essentially collaborative event which is co-constructed by a number of participants in a discourse sequence where one contribution may directly influence the next (Adolphs 2008: 3).

Thus it is not necessarily the case that the span on each side of the node can be determined in the same way as it may be when studying certain written phenomena. In (10), the speakers are talking about a school assignment and agree that the way it is organised is somewhat unsatisfactory. The use of *liksom* (*like*) by both speakers can be a way of co-constructing meaning. Co-construction of meaning across different speakers’ contribution could be one reason to expand the co-text in each speech situation.

(10)

A: * nei men det er så innmari
A: * no but it is so freaking
Although this study does not investigate the types of hedging strategies co-occurring with the probes, one interesting observation should still be mentioned: there is a large variety of hedging strategies in the material, and several of the strategies are rather un-typical compared to those that are used to illustrate hedging in the literature and that are often used as points of departures in form-to-function studies, e.g. *sort of, kind of, I think*, etc. This shows the value and necessity of studies that go from function to form, in terms of discovering a more extensive range of hedging strategies. These hedges should be studied more closely to evaluate existing classification system and potentially challenge them.

6. **Concluding remarks**

The purpose of this study was to explore an approach to the study of hedging strategies which moves in the direction of corpus-driven function-to-form rather than the typical form-to-function. The ambition was to test whether the use of a probe could be advantageous in retrieving hedging strategies in a bottom-up fashion. The need for such a bottom-up analysis is amplified by the change in the understanding of what constitutes hedging since it became a topic of research interest. When hedging is defined as a discourse strategy, communication strategy or rhetorical strategy like in many of today’s studies, e.g. Kaltenböck, Mihatsch, and Schneider (2010), Fraser (2010) and Prokofieva and Hirschberg
(2014), it is no longer clear what forms should be searched for in a corpus. But nor is it easy to search for the hedging function and this study thus investigates the use of probes. Previous research on the motivation behind the use of hedging strategies indicates that they will be used to attenuate face-threatening acts. Since the conjunctions *men* and *but* can be signals of face-threatening contrastive situations, these words were selected as probes to locate hedging strategies. *But* and *men* were chosen because of their register neutrality, their frequency in oral conversations and their core contrastive meaning.

The clauses with probes contained hedging strategies slightly more often than randomly selected clauses. Although the gain was limited and the results were only significant for the Norwegian data, the use of probes seem to be a promising technique that should be investigated further. It might be worth looking for even better probes that give a higher number of co-occurring hedging strategies. If hedging strategies occur more frequently as a remedial strategy in face-threatening situations, there might be range of various signals that could function as probes in what we may call a “form\_1-to-function-to-form\_2” approach.

It might also be worth considering an increase of the unit selected as the scope of the analysis. In this study, only the immediate clause following *but* and *men* was considered, which might have limited the number of co-occurrences, seeing as hedges often have scope over several clauses. A potential alternative would be to study hedging strategies at utterance or turn level. If good probes can be identified, this will make it easier to retrieve a large number of hedges, something that is needed if we want a full overview of how hedging can be realised. The hedging strategies identified through this study were extremely varied, which shows that such bottom-up approaches will be fruitful in terms of extending and modifying the existing taxonomies.

Finally, although the difference in the number of hedging strategies was significant only for Norwegian, the use of similar probes may still be a way of ensuring that cross-linguistic studies compare like with like. In this way it is possible to compare hedging strategies in two or more languages even when their realizations are different.
References


Gries, Stefan Th., and Caroline V. David. 2007. “This is kind of/sort of Interesting: Variation in Hedging in English.” in Päivi Pahta, Irma Taavitsainen, Terttu Nevalainen and Jukka Tykkö


**Corpora**

BigBrother-korpuset. [http://www.tekstlab.uio.no/nota/bigbrother/](http://www.tekstlab.uio.no/nota/bigbrother/)


Norsk talespråkskorpus - Oslodelen. [http://www.tekstlab.uio.no/nota/oslo/index.html](http://www.tekstlab.uio.no/nota/oslo/index.html)
Article 2

Title: A contrastive approach to types of hedging strategies in Norwegian and English informal spoken conversations

Author: Stine Hulleberg Johansen

Status: Under review. Revised according to reviews and re-submitted to *Contrastive Pragmatics* (24 April 2020)
A contrastive approach to types of hedging strategies in Norwegian and English informal spoken conversations

Abstract
Hedging strategies, i.e. downtoning expressions or expressions of tentativeness or possibility, are ubiquitous in most, if not all, languages. However, hedging is often realised differently across languages, making such strategies particularly interesting to study from a contrastive perspective. The purpose of this study is twofold. Its primary aim is to compare the types of hedging strategies used in Norwegian and English informal spoken conversations. However, in order to enable a cross-linguistic comparison, the study also aims to propose a classificatory framework for hedging strategies. Most existing frameworks are based on specialised and formal forms of discourse, primarily in English, which may cause problems for contrastive studies of informal spoken discourse. The results show that a combination of categories from existing frameworks can be used to describe hedging strategies in informal conversations and that there are several significant differences in the type of hedging strategies used in the two languages.

Keywords: hedging, English/Norwegian, classification, classificatory framework

1. Introduction
The understanding of the concept of hedging has changed dramatically since it first became a topic of research interest in the 1960s. Hedging has gone from being understood as a semantic feature of a class of words to a broad pragmatic concept encompassing nearly any expression of tentativeness or possibility, or with a softening or downtoning function. With this definition, hedging strategies are ubiquitous in both written and spoken discourse in most, if not all, languages, and hedging is considered to be an essential part of our pragmatic competence, i.e. the ability to “communicate your intended message with all its nuances in any socio-cultural context and to interpret the message of your interlocutor as it was intended” (Fraser 2010: 15). Pragmatic phenomena, such as hedging, are especially interesting to study from a contrastive perspective, because their realisations often do not correspond across languages (Romero-Trillo 2018). For example in Norwegian, hedging may be expressed through modal particles, which, in most cases, do not have a direct translation in English (Johansson and Løken 1997). This makes contrastive studies of hedging particularly relevant, because they “shine light on similarities and differences between languages and cultures” (Kaltenböck, Mihatsch, and Schneider 2010: 2).
In studies of hedging strategies, classificatory frameworks, such as those of Prince, Frader, and Bosk (1982) and Hyland (1996, 1998), are often used to describe and distinguish between strategies. Such frameworks say something about the type of hedging strategy, i.e. whether it concerns the relationship between parts of the proposition, between the speaker and the proposition, or between the speaker and the hearer. However, a challenge with many of the prevailing classificatory frameworks is that they are based on specific and often very formal types of discourse, such as physician-physician interaction (Prince et al. 1982), physician-patient interaction (Caffi 1999), scientific articles from medical journals (Salager-Meyer 1994) or scientific papers in molecular biology (Hyland 1996). Most frameworks are also based exclusively on English data and developed for analysing hedging in English. As a result, challenges arise when they are applied in studies of other discourse types and languages.

The purpose of this study is twofold. First, the study aims to identify similarities and differences in the types of hedging strategies used in Norwegian and English informal spoken conversations. Hedging in Norwegian remains underresearched (Vold 2006: 63), and the few studies on Norwegian either investigate hedging in written discourse (e.g. Vold, 2006), or study hedging as one of several pragmatic functions of various pragmatic markers (e.g. Andvik, 1992, and Berthelin and Borthen, 2019). In English, hedging has been studied from various perspectives and in various types of discourse (see further section 2); nevertheless, the majority of these studies are devoted to hedging in written discourse (e.g. Markkanen and Schröder, 1997, and Hyland, 1996). Thus, there is still a need to investigate hedging in informal spoken language. Moreover, the access to newly released corpora of spoken informal conversations in English, i.e. the spoken British National Corpus 2014, enables research on hedging in present-day informal speech in English. This newly compiled corpus of contemporary spoken discourse provides a unique source for investigations into pragmatic phenomena such as hedging in authentic everyday conversations.

Second, to ensure a sound basis for the cross-linguistic comparison, an operational classificatory framework is needed to account for hedging strategies in spoken Norwegian and English informal conversations. To the author’s knowledge, there is no established classificatory framework for hedging strategies in informal spoken language. Therefore this study also aims to propose such a framework. Informed by corpus data, the proposed framework will be based on four of the most prevailing classificatory frameworks of hedging strategies, i.e. Prince et al. (1982), Hübler (1983), Hyland (1996,
The bilingual dataset is particularly important from the perspective of contrastive pragmatics to make sure that the same classification scheme is equally applicable to the languages compared. The aims of the study can be summarised in two research questions:

1. What are the differences and similarities in the types of hedging strategies used in Norwegian and English informal spoken conversations?

2. To what extent can categories from existing classificatory frameworks be combined (and/or revised) to account for hedging strategies from Norwegian and English corpora of informal spoken conversations?

Although RQ1 is the overarching research question for this study, RQ2 must be addressed first in order to describe and classify the hedging strategies in the two languages. Both research questions will be discussed in the light of previous research on hedging in English and Norwegian outlined in section 2. The existing classificatory frameworks on which the proposed framework is built are described in section 3. The approach and results of the contrastive study are outlined in section 4. The material and methodological approach will be presented in section 4.1 and the results discussed in sections 4.2.2 to 4.2.4. Some concluding remarks will be offered in section 5.

2. Previous research on hedging

The start of hedging research in the mid-1960s is often seen in connection with the introduction of fuzzy set theory (Zadeh 1965) and the concept of metalinguistic operators (Weinreich 1966). In fuzzy set theory, the boundaries between classes of objects are considered to be fuzzy, i.e. one can talk about degrees of class membership. For example a dog is a prototypical member of the class animals, but bacteria would be a less prototypical member (Zadeh 1965: 338). The term metalinguistic operators refers to words which signal how phenomena should be interpreted (Weinreich 1966: 26). For example the word about can in some contexts indicate that something is an approximation. Based on this work, G. Lakoff (1972: 471), introduced the term hedge to describe words whose job it is to modify the category membership of a word or phrase. The initial concept of hedging included both the attenuation and the reinforcement of class membership. Lakoff was mainly interested in the properties of this class of words and how they indicated non-prototypicality of objects, i.e. made things more or less fuzzy.

---

28 Caffi (1999) does not study hedging, but rather mitigation. However, mitigation overlaps to a great extent with hedging, when both concepts are defined in a wide sense.
This type of hedging has later been referred to as *propositional hedging* in contrast to *speech act hedging*, which refers to hedges attenuating the strength of a speech act (Fraser 2010). The speech act aspect of hedging was developed further through the work of Brown and Levinson (1987) who studied hedging to account for politeness phenomena in language use. They explained hedging as applying a set of strategies which minimised the threat to the speaker’s and hearer’s face, i.e. their wish to be desired by and not imposed on by others (Brown and Levinson 1987: 62). The binary distinction between propositional hedging and speech act hedging has served as the foundation for several of the classificatory frameworks for hedging strategies developed later.

The concept of hedging was broadened further as a result of the growing influence of pragmatic research from the 1980s and onwards. The reinforcement aspect of hedging, introduced by Lakoff (1972), was pretty much laid aside (Fraser 2010: 26), and hedging research became centred on the attenuation aspect. From the 1980s and onwards, there were studies of the properties of individual hedging expressions (e.g. Aijmer, 1984). There were also several attempts to sub-classify hedges based on various criteria (Fraser 2010; Clemen 1997) (see further section 3). The use of hedging expressions was studied in various genres of written discourse, particularly in academic texts, e.g. Salager-Meyer (1994) and Hyland (1998, 1996), in learner language (e.g. Yang, 2013) and across languages and cultures (e.g. Kranich, 2011 and Tchizmarova, 2005). Hedging has also been studied as a pragmatic function of certain expressions such as pragmatic markers (e.g. Jucker and Ziv, 1998, Andersen, 2001 and Hasund, 2003), and as an aspect of related phenomena such as vague language (e.g. Channell, 1994, Jucker, Smith, and Lüdge, 2003, Overstreet, 2011), modality (e.g. Holmes, 1982, Coates, 1987 and Farr and O'Keefe, 2002), and mitigation (e.g. Fraser, 1980 and Caffi, 1999).

In the late 1970s and early 1980s, hedges also became an area of interest in Norwegian pragmatic research. The concept of *dempere* (‘attenuators’, ‘hedges’) was first mentioned in Fretheim (1979) and further investigated in Fretheim (1981). Fretheim (1981) states that his mapping of various hedges in Norwegian is an attempt to make sense of a linguistic topic on which close to no research has been done (Fretheim 1981: 95). Since the 1980s, Norwegian hedging research has overlapped with research on pragmatic particles (e.g. Andvik, 1992, Hasund, 2003 and Berthelin and Borthen, 2019. In such studies, hedging is regarded as one of several pragmatic functions of such particles and not investigated as a phenomenon as such. To the author’s knowledge, the only study devoted primarily to hedging in Norwegian is that of Vold (2006), who compares the use of epistemic modality markers in Norwegian, English and French scientific articles.
The development reflected in this extensive body of research has resulted in a much wider definition of hedging than the one proposed by Lakoff (1972). Today, hedges are mainly regarded as “realizations of an interactional/communicative strategy” (Markkanen and Schröder 1997: 5). The concept of “fuzziness” in the Lakoffian sense has been expanded to also include elements such as uncertainty, lack of precision and lack of commitment of the speaker to the propositional content (Prokofieva and Hirschberg 2014: 1). Furthermore, the idea of a grammatical class of hedges is abandoned and there is now general agreement that “no linguistic items are inherently hedgy” (Clemen 1997: 6) and that “hedging devices are drawn from every syntactic category” (Fraser 2010: 23).

Although most researchers now agree that hedging is some kind of strategy, the definitions applied in studies of hedging still vary in some respects. Some definitions are formulated from a speaker-oriented perspective, for example as in Fraser (2010: 22), where hedging is described as a “rhetorical strategy, by which a speaker […] can signal a lack of commitment to either the full semantic membership of an expression […] or the full commitment to the force”, whereas others also include the perspective of the hearer. For example, Holmes (1982: 9), states that hedging is a linguistic device used by speakers to express their attitudes to the proposition and “to reflect their perceptions of their relationship to those listening”. Hyland (1998: 1) defines hedging as any linguistic means indicating a lack of commitment to the truth of a proposition, or a desire to “not to express that commitment categorically”. Although the interlocutor is not explicitly mentioned in this definition, Hyland (1996: 446) explains how hedging has an interpersonal function and is used to appeal to the reader. Similarly, Tchizmarova (2005: 1146) understands hedging as a strategy that “mitigates the harshness or hostility of the force of one’s actions, softens the force of utterances and makes them more acceptable to the hearer.” This appellative aspect of hedging has also been discussed in connection with the functions of Norwegian pragmatic markers. Fretheim (1981: 87-88) distinguished between ego-dempere (‘ego-hedges’) and alter-dempere (‘alter-hedges’), arguing that certain expressions used to express the speaker’s uncertainty also implicitly seek confirmation from the hearer.

In this study, hedging is understood as discourse strategies that reduce the force, truth or perceived negative effect on the hearer of an utterance. This view encompasses the hearer-oriented focus mentioned in Fretheim (1981), Hyland (1996), and Tchizmarova (2005) and thus stretches beyond the distinction between hedging within the proposition and hedging between the speaker and the proposition, which characterises many hedging studies.
3. Classificatory frameworks for hedging strategies

The first exhaustive attempts to create a classificatory framework for hedges were developed in the early 1980s. The purpose of such frameworks was to divide the class of hedges into subcategories to better understand the nature of scope of various hedging strategies. Fraser (2010) mentions some 20 suggested sub-categories from various classificatory frameworks developed from the 1980s and onwards. However, only some of these frameworks, described in further detail below, have been and continue to be widely applied in studies of hedging strategies. In this paper, I will attempt to combine categories from existing classificatory frameworks into a revised framework for hedging in spoken conversations in order to account for hedging strategies in Norwegian and English informal spoken conversations.

Prince et al. (1982) developed the first widely acknowledged framework for hedging in spoken interaction. The framework arose from research on a corpus of interaction between physicians working in a paediatric ward. Prince et al. (1982) base their understanding of hedges on Lakoff’s (1972) concept of fuzziness, but distinguish between two types of fuzziness, i.e. within the propositional content and in the relationship between the speaker and the content. These two types of hedging strategies are further subdivided into four categories: adaptors, rounders, plausibility shields and attribution shields. Adaptors indicate non-prototypicality and may be used when the speaker adapts an existing word to a new situation, e.g. *His feet were sort of blue*. Rounders are used when the speaker does not want or need to give exact information, e.g. *His weight was approximately 3.2 kilograms*. Both are used to indicate that the actual situation is not identical with, but close to the prototypical situation (Prince et al. 1982: 11). Plausibility shields indicate uncertainty on behalf of the speaker towards the content of the proposition, e.g. *I think his feet are blue*. Attribution shields are used when the speaker wants to attribute the belief expressed in the proposition to someone else, e.g. *According to her estimates, the heart rate was back in two minutes*. The system of Prince et al. (1982) has also inspired later attempts of establishing classificatory frameworks, such as that of Salager-Meyer (1994), who use labels such as adaptors, rounders and shields in her description of hedging strategies in English medical discourse.

The same binary distinction between hedging within the proposition and hedging between the speaker and the proposition as in Prince et al. (1982) is found in Hübler (1983) who authored the first monograph on hedges. Hübler (1983: 11) argues that hedges may be used to express “the attitude of the speaker to the hearer regarding the proposition” and distinguishes between understatements and hedges, where understatements concern the propositional content, i.e. the phrastic part of the sentence,
and hedges concern the speaker's claim to the validity of the proposition, i.e. the neustic part of the sentence. However, Hübler’s distinction has received little support as understatements are commonly regarded as a sub-category of hedges (Clemen 1997: 241).

The related phenomenon of mitigation strategies can also be subdivided into different types depending on their scope in the speech act. Mitigation was originally defined as the reduction of unwelcome effects of a speech act on the hearer (Fraser 1980: 341), but is now regarded as a synonym for weakening, downgrading and downtoning (Caffi 1999: 884). Although mitigation and hedging have been regarded as two different concepts (Fraser 1980: 344), the broadening of the concepts has made them overlap to a large extent, making mitigation frameworks relevant for the study of hedging strategies. Caffi (1999) developed a framework for describing mitigation strategies which bears resemblance to the framework of Prince et al. (1982). Caffi (1999) distinguishes between mitigation on the proposition, bushes, mitigation on the illocution, hedges, and mitigation on the utterance source, shields, i.e. hiding the operator of mitigation. The category of shields is related to what Prince et al. (1982) call attribution shields. However, whereas Prince et al. (1982) include lexical expressions attributing a belief to someone else in this category, Caffi (1999: 896) also includes complete deletion of the source, for example as in passive clauses where the agent responsible for the action is hidden in the syntactic structure. Like Prince et al. (1982), Caffi (1999) also uses medical discourse data. Caffi (1999) studies data from a corpus of transcripts of doctor-patient and psychotherapeutic conversations, but in Italian rather than English.

Although Hyland (1996) studied hedging in written discourse, his framework is also relevant when studying spoken conversations, as it contains a category which neither Prince et al. nor Caffi include, i.e. reader-oriented hedges. Hyland (1996) distinguishes between content-oriented and reader-oriented hedges where the former concern the relationship between the proposition and a representation of the world, and the latter are hedges that seek acceptance by the reader for a claim made by the writer and have an interactional function in this respect (Hyland 1996: 446). Content-oriented hedges are further subdivided into accuracy-oriented and writer-oriented hedges (Hyland 1996: 436-37). The former type concerns hedges expressing uncertain scientific claims with appropriate caution to reduce the risk of refutation, e.g. The EGTA clots are possibly comprised of thinner fibres than [...] . The latter type concerns hedges that enable the writer to guard him-/herself against criticism, e.g. The figures suggest that [...] . Accuracy-oriented strategies are further sub-divided into attribute hedges indicating the extent to which propositional content is precisely described, and reliability hedges indicating the degree of writer confidence. Reader-oriented hedges attend to the acceptability of statements to the hearer and
share some characteristics with the alter-hedges proposed by Fretheim (1981). The categories proposed by Hyland (1996, 1998) have been criticised for being overlapping and difficult to distinguish (Diewald 2006); nevertheless, the framework is widely adopted in studies of hedging in scientific discourse.

Overall, the classificatory frameworks identify five main areas in which hedging takes place. Table 1 summarises the labels used in the different frameworks and how they overlap. These five areas and the corresponding sub-categories of hedging strategies were used as a point of departure for the analysis of the Norwegian and English corpus data (described in section 4). The labels applied in this study are presented in the bottom row of Table 1.

*Table 5 Subtypes of hedging strategies in the literature.*

<table>
<thead>
<tr>
<th>Fretheim (1981)</th>
<th>Propositional Content</th>
<th>Speaker intention</th>
<th>Speaker commitment</th>
<th>Source (of the proposition)</th>
<th>Effect on the interlocutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hübler (1983)</td>
<td>Understatements</td>
<td>Hedges</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salager-Meyer (1994)</td>
<td>Approximators</td>
<td>Shields, Author’s personal doubt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caffi (1999)</td>
<td>Bushes</td>
<td>Hedges</td>
<td>Hedges</td>
<td>Shields</td>
<td></td>
</tr>
<tr>
<td>The present study</td>
<td>Rounders, Adaptors, Understaters</td>
<td>Plausibility shields, Speaker-oriented hedges</td>
<td>Attribution shields</td>
<td>Hearer-oriented hedges</td>
<td></td>
</tr>
</tbody>
</table>

The material in this study shared some features with the spoken data in Prince et al. (1982), therefore the labels rounders, adaptors and shields were used with the definitions proposed by Prince et al. (see above). For the types of hedging strategies which could not be labelled according to Prince et al. (1982), labels from the other frameworks were used. The choice of label was based on the degree to which the
hedging strategies found in the data fitted the description of the category. Each category is discussed and exemplified in sections 4.2.2–4.2.4.

4. Comparing the use of hedging strategies in informal Norwegian and English spoken conversations

The following sections presents the contrastive study. Section 4.1 describes the material and the methodological approach used to retrieve hedging strategies. Section 4.2 presents the classification of the hedging strategies and discusses differences and similarities between the languages.

4.1 Material and method

The material for this study consists of three corpora of spoken Norwegian conversations, The Norwegian Speech Corpus (NoTa), the BigBrother Corpus (BB), the Norwegian part of the Nordic Dialect Corpus (NDC) and the spoken British National Corpus 2014 (BNC2014). These corpora were chosen due to their degree of comparability and their availability. The corpora are not directly comparable with respect to manner of compilation and size, but they were all compiled between 2001 and 2016 and include the same types of texts, i.e. natural conversations between family members, friends, acquaintances and strangers. The sizes of the corpora are presented in Table 2. (See the websites listed after References for further information.)

Realisations of hedging strategies were retrieved through a bottom-up approach by using a probe. A probe can be defined as an element, e.g. a tag, a word or a string of words used to find other elements that cannot easily otherwise be found in a corpus (Hunston 2002: 62). The idea was thus to use a probe as a tool to retrieve a wide range of hedging strategies without searching for actual realisations of them. In this study, the contrastive conjunction but (English) and its Norwegian counterpart men were used as a probe (Johansen 2019). The choice of probe was based on the assumption that hedging strategies are often used for politeness purposes in spoken conversations to avoid threatening the hearer’s face or damaging the speaker’s face (Nikula 1997: 192), and expressing a contrastive view, for example introduced by men/but, could potentially be face-threatening. Therefore it was deemed likely that hedging strategies would occur in the proximity of a contrastive marker such as men/but. This approach does not capture all potential realisations of hedging strategies in a corpus, but retrieves a variety of different strategies and thus gives an insight into how hedging is expressed in both languages. Furthermore, it ensures that similar speech situations are compared across the corpora.

---

29 See (Johansen 2019) for an extensive overview of the identification of probes.
Table 6 Overview of corpora and sample.

<table>
<thead>
<tr>
<th></th>
<th>Total no. of words</th>
<th>Total no. of ‘but/men’ concordance lines in the sample</th>
<th>Total no. of randomly selected ‘but/men’ concordance lines</th>
<th>Total no. of contrastive ‘but/men’ concordance lines</th>
<th>Total no. of hedging strategies co-occurring with ‘but/men’</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoTa</td>
<td>674,596</td>
<td>8,210</td>
<td>266</td>
<td>186</td>
<td>253</td>
</tr>
<tr>
<td>BB</td>
<td>440,354</td>
<td>5,854</td>
<td>267</td>
<td>193</td>
<td>195</td>
</tr>
<tr>
<td>NDC</td>
<td>1,199,651</td>
<td>14,280</td>
<td>267</td>
<td>203</td>
<td>229</td>
</tr>
<tr>
<td>Total Norw. corpora</td>
<td>2,314,601</td>
<td>28,344</td>
<td>800</td>
<td>582</td>
<td>677</td>
</tr>
<tr>
<td>BNC2014</td>
<td>11,422,617</td>
<td>103,439</td>
<td>800</td>
<td>659</td>
<td>762</td>
</tr>
</tbody>
</table>

In total, 800 occurrences of *but* and 800 occurrences of *men* were retrieved randomly from the corpora, by using the random selection settings in the corpus interfaces.\(^{30}\) Table 2 shows the total number of hits for the probe as well as the randomised samples. Subsequently, all 1600 instances of the probe were examined manually and the contrastive uses were identified. The material yielded 582 contrastive uses of *men* and 659 contrastive uses of *but*. *But* and *men* were mainly considered contrastive when they denoted denial of expectation, opposition of two elements (antithetic use), modification or explanation of a previous statement, restriction of a previous statement, or when a speaker objects to something said by another speaker. Any instances of *but/men* not used in their contrastive sense, but, for example, used as a discourse marker to structure units of talk or as a topic changer were disregarded. Following the identification of the contrastive uses of the probe, potential hedging strategies were identified to the right of the probe, as illustrated in Figure 1.

![Utterance](image)

*Figure 1 Illustration of the probe but/men in context.*

\(^{30}\) These were Glossa for the Norwegian corpora (https://www.hf.uio.no/lin/english/about/organization/text-laboratory/services/glossa/) and cqpweb for BNC2014 (https://cqpweb.lancs.ac.uk/).
The example in (1) illustrates the use of the probe, but, and a co-occurring hedging strategy, kind of, to the right of the probe.

(1)

[…] in my mind I was thinking he's he's most likely dead but I kind of hoped that he wasn't

(BNC2014 text S7BR no. 748)

As shown in Table 2, 60% of the 582 contrastive uses of men co-occurred with hedging strategies. Of the 659 contrastive uses of but, 53% co-occurred with hedging strategies, which in turn gave a total of 1,439 hedging strategies to be categorised.

4.2 Results and discussion

This section presents the overall results of the classification giving some quantitative data in 4.2.1 and a more detailed qualitative discussion of the subcategories in sections 4.2.2-4.2.4.

4.2.1 Overall classification and comparison

Overall, the material could be grouped into three main types of hedging: within the proposition, between the speaker and the proposition, and between the speaker and the hearer. Subsequently, an attempt was made to sub-classify the strategies using terminology from the existing classificatory frameworks as outlined in section 3. Compared to the results of a pilot study using Prince et al.’s (1982) classification system (Johansen 2018), the present framework performed very well in being able to account for the whole dataset and not leaving a substantial part of it to a ‘miscellaneous’ category. The outcome of the classification is presented in Table 3 and illustrated in Figure 2. For each subcategory the two languages have been compared using a log-likelihood test.31 The p-values are shown in the rightmost column.

31 http://ucrel.lancs.ac.uk/llwizard.html
Table 7 Types of hedging strategies in Norwegian and English.

<table>
<thead>
<tr>
<th>Types of hedging strategies</th>
<th>Norwegian</th>
<th></th>
<th>English</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw frequencies</td>
<td>%</td>
<td>Raw frequencies</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>Within the proposition</td>
<td>Rounders</td>
<td>67</td>
<td>9.9</td>
<td>115</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>Adaptors</td>
<td>63</td>
<td>9.3</td>
<td>94</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>Understater</td>
<td>92</td>
<td>13.6</td>
<td>68</td>
<td>8.9</td>
</tr>
<tr>
<td>Between the speaker and the proposition</td>
<td>Plausibility shields</td>
<td>150</td>
<td>22.2</td>
<td>190</td>
<td>24.9</td>
</tr>
<tr>
<td></td>
<td>Attributions shields</td>
<td>4</td>
<td>0.6</td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Speaker-oriented hedges</td>
<td>106</td>
<td>15.4</td>
<td>197</td>
<td>25.9</td>
</tr>
<tr>
<td>Between the speaker and the hearer</td>
<td>Hearer-oriented hedges</td>
<td>195</td>
<td>28.8</td>
<td>96</td>
<td>12.6</td>
</tr>
</tbody>
</table>

677 ≈ 100 762 ≈ 100

32 This category has not been tested for significance due to the low number of frequencies in both languages.
4.2.2 Propositional hedging strategies

Propositional hedging strategies could be divided into three sub-categories: rounders, adaptors and understaters. An example of a rounder is given in (2). Like has an approximative function indicating that an hour is an approximate estimate of time. Similarly, in (3), en plass (‘a place’, ‘somewhere’) signals an approximate location. Kind of in (4) is an example of an adaptor which modifies freaked (me) out. An interpretation of this example may be that the speaker was shocked by what she read, but perhaps not to the extent that she completely lost control or panicked, as the choice of verb, freak out, could indicate. The use of kind of signals that the actual situation is close to, but not identical with the prototypical situation of freaking out (Prince et al. 1982: 11). Sånne perioder in (5) is another example of an adaptor. The speaker here may be uncertain about whether the term perioder (‘periods’) is the appropriate term to use or he decides that the exact term is not relevant and chooses the closest relevant term (Prince et al. 1982: 9). This adaption is marked by sånne (‘like’).

(2)
Schools will close the kids will be out at half past one and then they have to be back for half past three but it’s crazy cos then they go back for like an hour and a half
BNC2014 text: S2XJ no. 361

(3)
[...] men det viste seg det at han gikk inn der fikk ut der og der lå det et lite fiskebein # som hadde satt seg fast i spiserøret en plass
‘[...] but it turned out that he went in there and got out there and there there was a small fishbone # which had gotten stuck in the esophagus somewhere’

NDC no. 553 (nes_03gm><who_avfile nes_03gm-04gk)

(4)

I read I read it a bit wrong (.) I do n't know what I read but it kind of freaked me out and made me stop

BNC2014 text: SFP5 no. 364

(5)

[…] så bodde jeg da disse sju årene i Afrika da men det var da m med såanne # perioder hjemme imellom

‘[…] so then I lived then these seven years in Africa but it was then m with like # periods home in between’

NoTa, no. 825 (070><who_avfile 069-070)

Rounders and adaptors in Prince et al. (1982) are related to (non-)prototypicality. In the material, however, there was a type of hedging strategies with a propositional scope which did not indicate (non-)prototypicality and could not be labelled rounders or adaptors. These expressions seemed to downplay the importance of the whole or parts of the proposition and could be said to have a dual function in that they both modify a concept locally, but at the same time mitigate the force of the utterance to make it less of a burden on the hearer. Because of this potentially dual function, such expressions have been categorised separately and labelled understaters, in line with Hübler (1983: 70) who states that understaters are expressions which have a detensifying effect on the phrastic part of a sentence or utterance. Understaters are typically used to underrepresent the state of affairs (Markkanen and Schröder 1997: 7) and can have both a propositional and a relational scope.

In (6) speaker B interrupts speaker A by objecting to what speaker A is saying, which may be perceived as face-threatening. However, by suggesting that the river is only a bit (‘litt’) wider, the force of the objection is reduced and may be easier to accept. Litt says something about the size of the river and thus has a propositional scope.

(6)

A: ja ja ja veldig gøy det passa fint i i de strykene der
B: *men det var
B: litt bredere bredde på den elva da

‘A: yes yes yes a lot of fun it fitted nicely in those rapids there
B: *but it was
Comparing the two languages, the results show that English speakers used significantly more rounders than Norwegian speakers, whereas Norwegian speakers used significantly more understaters than English speakers (Table 3). The expressions classified as rounders were typically vague quantifiers, such as *a lot of* and *a number of*, and vague nouns, such as *thing* and *stuff*. The reasons for the high number of such expressions are unclear. In a study of vague expressions, Jucker, Smith, and Lüdge (2003: 1755) state that “it is a challenge to explain why so many vague expressions for describing frequency exist in English” but suggest that it may be to indicate that the exact number is not relevant. The high number of understaters in Norwegian may be related to what has been referred to as the “distance rule of politeness” in Norwegian (Rygg 2017: 2). Rygg (2017) argues that Norwegians show consideration by not imposing on other people, and this may also be reflected in the use of understaters to tone down the importance of the message.

4.2.3 Hedging between the speaker and the proposition

Hedging strategies indicating uncertainty on the part of the speaker were frequent in both languages. Expressions indicating uncertainty have been labelled plausibility shields, and examples (7), (8) and (9) show some of the forms in which these plausibility shields occurred. There are several types of hedging strategies occurring together in (7). *Kanskje* (‘maybe’), which occurs twice, is an example of a plausibility shield. As described in Prince et al. (1982: 11), shields imply that the speaker is not fully committed to the truth of the proposition. *Liksom* (‘like’), which has a hedging function (Beeching 2016: 128; Hasund 2003), and the general extender *og sånn* (‘and stuff’) will be discussed in more detail below. *Litt* (‘a bit’) functions as a propositional understater here and modifies the degree to which women decide matters. Examples (8) and (9) show the expression *I think* and the superficially similar *jeg tror* as plausibility shields. In (8), speaker A is objecting to what speaker B is saying. *No*, here, signals agreement, but then the agreement is modified by *but*. Speaker A modifies his argument by not fully committing to the proposition. In (9) speaker A is turning down her own suggestion, after speaker B has indirectly turned it down, by suggesting something else.
(7) 

før i tida # så var det kvinnfolka som var hjemme og sånn og sånn så _uforståelig_ men nå er liksom # nå er blitt mer sånn # nå # det kanskje kvinner som bestemmer # kanskje til og med litt mer # hun kjefter og liksom sånn og sånn og sånn

‘back in the day # it was the women who were home and stuff and like _incomprehensible_ but now is like # now is has become more like # now # it maybe women who decide # maybe even a bit more # she yells and _like stuff and stuff and stuff_’

BB no. 222 (Ramsy><who_avfile 5)

(8) 

A: >> yeah yeah yeah but the other thing is they do n't know how to write CVs and they actually do n't know know

B: mm I 'm not I 've seen millions of CVs and I 'm going to be honest I 'm not sure if I know the right CV

C: >> --UNCLEARWORD well I do think I think that's that's a problem everybody has but erm

A: >> no _but I think_ it's easier when you 're at an --UNCLEARWORD

BNC text: SA6K, no. 431

(9) 

A: kanskje vi skal lage et lite prosjekt rundt det?
B: det er så lite kaffeplanter i Norge vet du det er det _uforståelig_
A: * ja * ja men ja ja m # nei _men vi jeg tror_ vi skal prøve å eksperimentere med _litt andre ting jeg_

‘A: maybe we should make a small project around it?
B: there are so few coffee plants in Norway you know it is it _incomprehensible_
A: *yes *yes but *yes yes m # no _but we I think_ we should try to experiment with _some other things I_’

NoTa no. 1043 (102><who_avfile 101-102)

A type of strategy which rarely occurred in the material was the type referred to as attribution shields, shields and writer-oriented hedging strategies, i.e. attributing the belief to someone else or hiding the source of the proposition. This type of strategy is illustrated in (10). The lack of such examples in the material may be due to the type of data investigated. Hyland (1996) states that writer-oriented hedges are used when a writer wishes to avoid personal responsibility for the proposition, which is a common characteristic of scientific texts where one’s work is supposed to build on the previous work of others. Prince et al. (1982) state that attribution shields were mainly used to relate background information to the present situation, typically the cause of hospitalisation which is natural in a hospital setting.
Although Prince et al. (1982: 19) maintain that attribution shields also occur in other domains, they do not substantiate this claim.

(10)

A: er Yuri Gagarin and people like that
B: yeah that's right (.) and was that proved? Proven to have been faked?
A: no there's no question marks about that because they didn't actually make it get it to the moon (.) but
B: right
A: but it according to this book if I remember correctly they they were genuine

BNC2014 text: SVBJ no. 543

Strategies saying something about the speaker’s commitment to the force or the content of the proposition were labelled speaker-oriented hedging strategies partly corresponding to writer-oriented hedges in Hyland (1996). This type of strategies does not indicate uncertainty, but a common denominator of most of these strategies was that they seem to create distance between the speaker and the proposition. In (11), speaker A downplays his/her opinion about the food by stating that he/she is not an expert on Chinese restaurants. In (11), the whole clause functions as a hedge. In (12), speaker A reduces the commitment to the proposition by using liksom (‘like’) in clause-final position. Liksom in clause-final position can have a hedging effect on the preceding expression (Hasund 2003: 199).

(11)

I mean it really is uh (.) the quality of the food I thought was really good (.) but (.) but I'm not um I'm not an expert on Chinatowns of course we don't have a Chinatown here

BNC text: SWMV no.582

(12)

Nei men du kommer deg mye mer fram med trikken liksom spesielt nå som den går hver femte minutt
‘No but you get there more easily by tram like especially now when it departs every five minutes’

NoTa no. 964 (030><who_avfile 029-030)

Speaker-oriented hedging strategies were significantly more frequent in the English data than in the Norwegian data. A contributing factor could be the frequent use of like to signal lack of commitment. Like is very common in informal speech and is often used to downplay a potential dogmatic delivery
and hedge a potentially critical stance made by the speaker (Beeching 2016: 132, 34). The *like* in (13) is an example of a softening *like*.

(13)

A: I could take you out for a special breakfast at some point
[...]
A: the --ANONplace?
[...]
B: yeah **but** we won't want to get *a little* closer to it *er* like so we don't have to walk ages cos it makes me really hungry

BNC text: S839 no.346

4.2.4 Hedging between the speaker and the hearer

Hedging on the relationship between the speaker and the audience is only included in the framework of Hyland (1996). Still this type of hedging occurred frequently in the material. It seems as if speakers in conversations tend to be very much aware of their interlocutor and that maintaining a good relationship between the interlocutors is a priority. This was typically done by trying to establish common ground between the speaker and the hearer. In (14) speaker A uses *you know* in clause-final, but utterance-medial position. In medial position *you know* can be used to invite co-construction of the content or the formulation of the message (Beeching 2016: 97). The general function of the pragmatic marker *you know* is to appeal to actual or fictive common knowledge between the interlocutors (Beeching 2016: 98). By combining the second person pronoun, *you*, with the cognitive verb, *know*, it has a stronger appeal to the addressee (Beeching 2016: 98).

(14)

A: I'd like to be lecturer
B: yeah er
A: >> and I don't mind the first years **but** I'd like to mix them with (.). *kind of* fourth years and masters students *you know* so you get *a bit* more of a spread cos you get intolerant of (.). I shouldn't *you know* but

BNC text: SEPP no. 324

Hedging strategies on the relationship between the speaker and the hearer occurred significantly more in the Norwegian material than in the English material (Table 3). A reason for this may be the Norwegian particle *jo* (‘after all’, ‘of course’) which shares some of the pragmatic functions of *you know*. Jo occurs frequently in spoken informal language (Berthelin and Borthen 2019: 1) and has traditionally been considered to indicate some sort of ‘givenness’ and it can be interpreted as a way of
establishing a mutual manifest between the speaker and the hearer (Berthelin and Borthen 2019). The *jo* in (15) may be interpreted as an assertion of common ground between the speakers.

(15)
A: det er i Oslo # så er det sånn t- sånn noe ruiner av et sånt kloster der_inne
B: ja **men** det i Trondheim har *jo* vært der lenge *da*

‘A: it is in Oslo # so it is like t- like some ruins from a like convent in there
B: yes **but** the one in Trondheim has [pragmatic particle] been there for a long time [pragmatic particle]’

BB no. 23 (Lars_Joakim><who_avfile 12)

A challenge with the Norwegian particles is that they may serve more than one function simultaneously, and it can be difficult to determine their exact pragmatic function and meaning. A particularly challenging particle is *da* (‘then’, ‘consequently’, ‘right?’) also occurring in (15). The effect of the use of *da* is to a large degree dependant on the context in which it occurs and its effect ranges from modifying and questioning to assertive and having a strengthening effect on the proposition (Borthen 2014: 257).

Another example of how a speaker may seek agreement or solidarity with the hearer can be found in (7). The general extender *og sånn* (‘and stuff’) can serve a variety of functions, not all of which will be discussed here. In this example, a likely interpretation could be that it indicates a wider semantic scope (Andersen 2010), i.e. that the speaker is only mentioning one of several activities, *kjefter* (‘yells’), or that it serves as a marker of positive politeness inviting solidarity between the speaker and the hearer (Overstreet 1999: 146).

Tag questions may also serve a hedging function. Particularly asymmetrical tag questions (with opposite polarity) with rising intonation can indicate uncertainty or doubt on behalf of the speaker (Hübler 1983: 108-09) or be a way of inviting the hearer into the conversation and thus creating common ground (Brown and Levinson 1987: 101). Tag questions with falling intonation may also assert common ground, e.g. by presupposing that the speaker and the hearer share certain knowledge (Brown and Levinson 1987: 103). (16) is an example of an asymmetrical tag question, which could be interpreted as the speaker seeking confirmation from the interlocutor. In (16) the speakers are talking about the Brexit vote and the speaker uses both the understater *a bit* and the tag-question *wasn’t I* to modify his/her statement.
(16)

A: what did you vote then?
B: I voted in but I’m saying I would be more ha- I was saying I was a bit on the fence wasn’t I? I’d be more happy about it if if we would have more control over what we do as a country

BNC2014 text: S8CV no. 59

5. Concluding remarks

The purpose of this study was twofold. The primary aim was to discover similarities and differences in the use of hedging strategies in Norwegian and English informal spoken conversations. Second, in order to compare strategies across the two languages, existing classificatory frameworks for hedging strategies needed to be revised and adapted to fit informal spoken conversations. The majority of these frameworks were based on formal spoken or written language mainly in English, thus they were not directly applicable to informal spoken language.

In summary, the results showed that existing classificatory frameworks can be used to account for hedging strategies in Norwegian and English spoken conversations, but that more than one framework is necessary and that challenges arise when frameworks that are based on other types of discourse are applied to contrastive conversational data. For example, categories, such as attribution shields (Prince et al. 1982), shields (Caffi 1999) and a sub-type of speaker-oriented hedges (Hyland 1996), which are thoroughly described and exemplified in the descriptions of the respective frameworks, were scarce in the conversational data. Conversely, examples of hearer-oriented hedges (Hyland 1996) were frequent in the data, but are only included in one of the classificatory frameworks. Another challenge with classifying hedging strategies was to distinguish between the different sub-categories. Many of the forms used to express hedging, such as several of the Norwegian particles, may serve different functions simultaneously and thus may be challenging to classify.

The threefold distinction of hedging strategies, i.e. within the proposition, between the speaker and the proposition and between the speaker and hearer, challenges the binary distinction on which most of the classificatory frameworks are built. However, the inclusion of the speaker-hearer dimension highlighted some interesting cross-linguistic differences in the data. The results showed that English speakers used significantly more speaker-oriented hedging strategies than Norwegian speakers. This may be due to the frequent use of like to signal lack of commitment in informal conversations.

33 Referred to as writer-oriented hedges in Hyland (1996).
34 Referred to as reader-oriented hedges in Hyland (1996).
Norwegian speakers used significantly more hearer-oriented strategies. This difference could be related to the appellative aspect of many Norwegian modal particles often used for hedging purposes, such as *jo*, for which there is no literal English translation and which is very common in informal spoken language. Within the propositional hedging category, the results also showed that English speakers used significantly more rounders than the Norwegian speakers and that Norwegian speakers used more understaters than English speakers. Future studies – preferably including hedging strategies in different types of context, and in different discourse types and languages – will test the more general applicability of the classificatory framework proposed here.
References


Diewald, Gabriele. 2006. Discourse particles and modal particles as grammatical elements. in Kerstin Fischer (ed.), *Approaches to discourse particles*. Amsterdam: Elsevier.


Hasund, Ingrid Kristine. 2003. The discourse markers *like* in English and *liksom* in Norwegian teenage language : a corpus-based, cross-linguistic study, University of Bergen, Agder University College.


**Corpora**

"BigBrother-korpuset." In.: Tekstlaboratoriet, ILN, Universitetet i Oslo. [https://www.hf.uio.no/iln/english/about/organization/text-laboratory/services/index.html#speech](https://www.hf.uio.no/iln/english/about/organization/text-laboratory/services/index.html#speech)


"Norsk talespråkskorpus -oslodelen." In.: Tekstlaboratoriet, ILN, Universitetet i Oslo. [https://www.hf.uio.no/iln/english/about/organization/text-laboratory/services/index.html#speech](https://www.hf.uio.no/iln/english/about/organization/text-laboratory/services/index.html#speech)
Title: “I guess anyone would do that wouldn't they?” How do native speakers of Norwegian and English hedge in informal conversations?

Author: Stine Hulleberg Johansen

Status: Under review. Revised according to reviews and re-submitted to Languages in Contrast (8 April 2020)
“I guess anyone would do that wouldn't they?”

How do native speakers of Norwegian and English hedge in informal conversations?

Abstract
Hedging is a complex phenomenon with an indefinite number of potential realisations. The complexity and versatility of hedging strategies make them particularly interesting to study across languages. This contrastive study compares the realisations of the pragmatic function of hedging in Norwegian and English everyday conversations by using data from four corpora of Norwegian and English informal spoken conversations (the Norwegian Speech Corpus, the Nordic Dialect Corpus, the BigBrother corpus, and the BNC2014). The results show that speakers of both languages mainly use pragmatic particles, adverbs, and first/second person pronoun + cognitive verb (1/2 pers. + Cog. V) to express hedging. Furthermore, English speakers use significantly more 1/2 pers. + Cog. V and modal verbs than Norwegian speakers, whereas Norwegian speakers use significantly more adjectives, prepositional phrases and clauses to hedge their utterances.

Keywords: hedging, spoken conversations, corpus pragmatics

1. Introduction

Hedging strategies are rhetorical or discourse strategies that reduce the force or truth of the whole or parts of an utterance (Kaltenböck, Mihatsch, and Schneider 2010: 1) and may take almost any linguistic form. Hedging strategies may function within the proposition (He is sort of nice), between the speaker and the proposition (He is probably at the office), or between the speaker and the hearer (I didn’t like the way he said it, you know) and are used for various purposes in both formal and informal spoken and written text. But although hedging strategies are both important and frequent, they are also complex linguistic tools which require a high level of competence and sophistication, even in one’s own mother tongue (Markkanen and Schröder 1997: 13). If speakers fail to hedge appropriately, they risk being perceived as impolite, arrogant, or offensive (Fraser 2010: 1).

This complexity and versatility make hedging strategies particularly interesting to study from a contrastive perspective, and several studies have compared the use of hedging strategies across
languages or varieties (e.g. Kozubíková Šandová (2017), and Kranich (2011)). However, the number of studies on written language significantly outnumber those on spoken language, mainly due to the relatively few spoken corpora available (Adolphs and Carter 2013: 1). Furthermore, the studies of hedging in spoken language have typically been restricted to certain lexical items used for hedging purposes (e.g. Farr and O'Keeffe (2002)). Thus, the use of hedging in spoken conversations needs further exploration.

The present study aims to identify and compare the forms used to express hedging in informal spoken conversations in Norwegian and English. *Form* is used here to signify a word, a phrase, or a clause realising the pragmatic function of hedging in an utterance. The study aims to discover potential similarities and differences in the forms used to express hedging by native speakers of Norwegian and English. Although hedging in English has been studied from various perspectives, hedging in Norwegian remains relatively underexplored. This study aims to serve as a first exploration into the differences and similarities between these languages strengthening the empirical basis for further research. Specifically, this study aims to answer the following research question:

What are the differences and similarities in the form of hedging strategies used by native speakers of English and Norwegian in informal spoken conversations?

Norwegian and English seem to have somewhat different repertoires of expressions that can function as hedges. In Norwegian, hedging in spoken conversations is frequently expressed using modal particles, such as *jo*, *da*, and *nok*, of which there often are no literal translations into English (e.g. Lind (1994), Opsahl (2002), and Berthelin and Borthen (2019)). Similar meanings must therefore be expressed through other pragmatic devices in English (Andvik 1992: 1). In English, hedging may be realised through the pragmatic use of expressions such as *you know* and *I mean* (Beeching 2016), tag questions, such as *isn’t it* (Hübler 1983; Kimps, Davidse, and Cornillie 2014), or modal verbs, such as *would* or *may* (Farr and O'Keeffe 2002; Hardjanto 2016).

In this study, hedging strategies will be retrieved through a bottom-up approach from corpora of spoken Norwegian and English. To circumvent the need to pre-define hedging strategies, a probe (defined in Section 3) will be used to identify speech situations in which hedging strategies are likely to occur. Hedging strategies will be retrieved from the context of the probe and grouped according to their linguistic form, and their use will be compared across the two languages. The research question will be addressed in the light of previous research on hedging outlined in Section 2. The Norwegian and English corpora and the methodological approach are presented in Section 3, and the categorisation
of the findings is explained in Section 4. The results are presented and discussed in Section 5 and some concluding comments are offered in Section 6.

2. Previous research on hedging

Hedges have been studied since the late 1960s. The term *hedge* was introduced by Lakoff (1972) to describe words that make things more or less fuzzy. The concept of fuzziness was rooted in logic and formal semantics and was related to the idea that there are degrees of membership of objects in a set (Zadeh 1965: 338-39). Lakoff looked at the properties of certain words and how they expressed non-prototypical characteristics of objects, e.g. *A penguin is sort of a bird.* This type of hedging is known as ‘propositional hedging’ (Fraser 2010: 22). Inspired by Lakoff (1972), Fretheim (1979) introduced the Norwegian term *demper* (‘attenuator’) corresponding to the English term ‘hedge’. Attenuators were described as words without semantic content and which served a pragmatic function (Fretheim 1979: 216). Moreover, attenuators were related to both hedging in the Lakoffian sense and ‘downtoners’, i.e. adverbs that decrease the force of an utterance, introduced by Quirk et al. (1972: 439 ff). The concept of hedging was later broadened to include expressions reducing the force of a speech act, i.e. ‘speech act hedging’ (Fraser 2010: 22). Today, hedging is understood as any linguistic strategy which attenuates the truth or force of an utterance and is no longer seen as a property of certain words or phrases but as a pragmatic function with an indefinite number of potential realisations, i.e. “no linguistic items are inherently hedgy, but can acquire this quality depending on the communicative context or the co-text” (Markkanen and Schröder 1997: 6).

Hedging has been studied from various perspectives. From a politeness perspective, hedging is mainly regarded as a positive or negative politeness strategy to be applied in face-threatening situations. For example, to avoid indicating disagreement when expressing a contrasting opinion, hedging expressions can make it less clear that one disagrees (Brown and Levinson 1987: 116). There have also been attempts to develop and test classificatory frameworks for hedging strategies (e.g. (Prince, Frader, and Bosk 1982; Hübler 1983; Salager-Meyer 1994; Hyland 1996, 1998)). Other studies have investigated properties and content of individual hedging expressions (e.g. Aijmer (1984), Farr and O’Keeffe (2002) and Mauranen (2004). Studies on the properties of individual hedges have largely overlapped with studies of pragmatic markers. Pragmatic markers constitute a fuzzy category and have been defined in various ways (see for example Jucker and Ziv (1998) and Aijmer and Simon-Vandenbergen (2006)). The term ‘pragmatic marker’ is often used to signify any linguistic sign, e.g.
word, word order, or intonation pattern, used to communicate something beyond the propositional content (Fraser 1996: 167). From the perspective of pragmatic markers, hedging is often regarded as one of several potential pragmatic functions a marker may perform.

Hedging has also been studied across languages, genres, and in learner language (e.g. Kranich (2011), Buysse (2017), and Magliacane and Howard (2019)). Many of these studies have looked at hedging in written academic discourse (e.g. Markkanen and Schröder (1997) and Hyland (1996, 1998)). Studies on hedging in spoken text have typically been concerned with individual hedging expressions: for example, Buysse (2018), who investigates the use of actually and in fact in spoken learner English, and Tchizmarova (2005), who studies the Bulgarian discourse marker xajde in spoken and spoken-like texts.

Hedging in Norwegian remains underresearched (Vold 2006: 63). Whereas English hedging has been studied from various perspectives, Norwegian hedging research has become closely intertwined with research on pragmatic particles, i.e. a subtype of pragmatic markers. Borthen (2018) offers an extensive overview of research on Norwegian pragmatic particles. In addition, a few unpublished theses could be mentioned: Solberg (1990), Lind (1994) and Opsahl (2002). All of these studies start with predefined lexical items and investigate their functions in different contexts. For example, Hasund, Opsahl, and Svennevig (2012) studied the use of bare ‘just’, ‘(be) like’ and suggested that the quotative use of bare could indicate an approximate rendering of what has been said and thus express hedging. Hasund (2003) compared the textual, subjective and interpersonal uses of the Norwegian discourse marker liksom and the English like. Fretheim (1981) studied various uses and meanings of vel ‘well’ and nok ‘probably’.

To the author’s knowledge, the only cross-linguistic study of hedging strategies in Norwegian and English is that of Vold (2006). Vold compares the use of epistemic modality markers, such as perhaps and probably, in Norwegian, English, and French academic texts and investigates whether language and academic discipline affect the use of hedging strategies. Her results show that language influences the frequency of hedging strategies more than academic discipline, and that epistemic modality markers indicating uncertainty are more frequent in English research articles than in Norwegian and French (Vold 2006: 78).
3. Material and approach

For the Norwegian analysis, the conversational parts of the Norwegian Speech Corpus (NoTa), and the Norwegian Dialect Corpus (NDC) (Johannessen et al. 2009) were used, as well as the entire BigBrother Corpus (BB). The English data comes from the Spoken British National Corpus (BNC2014) (Love et al. 2017). Table 1 outlines the properties of the corpora and the number of words in the sample used for this study. The corpora are not directly comparable with respect to size, or time and method of compilation but they all represent informal spoken conversations between family members, friends, acquaintances, and strangers recorded in more or less natural settings.

Table 1. Corpus overview.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Year of compilation</th>
<th>No. of words in the sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norwegian Speech Corpus</td>
<td>2005-2006</td>
<td>674,596</td>
</tr>
<tr>
<td>Nordic Dialect Corpus</td>
<td>2006-2012</td>
<td>1,199,651</td>
</tr>
<tr>
<td>BigBrother Corpus</td>
<td>2001</td>
<td>440,354</td>
</tr>
<tr>
<td>Spoken British National Corpus 2014</td>
<td>2012-2016</td>
<td>11,422,617</td>
</tr>
</tbody>
</table>

To access potential hedging strategies in the data, a probe was used. A probe may be explained as an element – a word, a string of words, or a tag – used to find other elements which cannot easily be found in a corpus (Hunston 2002: 66). A probe can be used to retrieve hedging strategies without searching for actual forms of such strategies (Johansen 2019). Thus, it does not limit the study to a pre-defined set of linguistic items but rather retrieves them through a bottom-up approach. Using a probe may fall within the function-to-form category of methodological approaches, where you start from a function and investigate realisations of that function (Aijmer and Rühlemann 2015: 9). It is not function-to-form in its purest sense, i.e. “the holy grail” of corpus pragmatics (O’Keeffe 2018: 599), but rather form1-to-function-to-form2 (Johansen 2019: 127), where the probe (form1) is used as a tool to retrieve a pragmatic function in the corpora and subsequently realisations (form2) of that function, thus the probe itself is not the subject of study. The approach can be described as a combination of

---

35 There are few available corpora of spoken Norwegian and these three are the only ones representing spoken informal conversations.
form-to-function and function-to-form approaches. As stated by O'Keeffe, Clancy, and Adolphs (In press), there is a need to establish “workarounds” to get to pragmatic functions in a corpus.

The choice of using a probe to access hedging strategies is also related to the degree of comparability of the corpora. As the corpora are only partly comparable, what is studied within the corpora must be as similar as possible. In this study, the probe functions as a marker of a certain communication situation where hedging strategies are likely to occur, and potential hedging strategies are retrieved from the immediate co-text of the probe (Johansen 2019: 122). Thus, the probe is advantageous in identifying similar speech situations across the corpora ensuring that the hedging strategies compared are produced in similar settings. Furthermore, a study comparing the number of hedging strategies retrieved using a probe to those retrieved by looking at random pieces of text revealed that using a probe gave more hedging strategies than the random pieces of text, although the difference was only significant for the Norwegian data (Johansen 2019).

The coordinator but and the corresponding Norwegian men were chosen as probes based on previous research, frequency in spoken everyday language in both languages (Biber et al. 1999: 81), and similar semantic and syntactic properties. But and men typically denote a contrastive relationship of some sort, either explicitly by connecting two elements which are in direct semantic opposition (Blakemore 1989: 15), implicitly in the content of the elements they connect (Faarlund, Lie, and Vannebo 1997: 1138) or inferable from the context if the proposition violates the speaker’s expectations (Schiffrin 1987: 156). Expressing contrasts may be a way of disagreeing either with an interlocutor or with something the speaker him-/herself said earlier in the conversation. Expressing disagreement could potentially be face-threatening to both interlocutors (Brown and Levinson 1987: 66, 68). Hedging is likely to appear in potentially face-threatening situations to reduce the threat or shield the speaker from a possible objection from the addressee (Clemen 1997: 244). Figure 1 and example (1) illustrate how could, probably and a bit follow the probe but within an utterance.

---

36 See Johansen (2019) for an extensive overview of the selection of probes.
From the corpora, 800 random instances of *but* and 800 random instances of *men* (267 from BB, 267 from NDC and 266 from NoTa) were extracted, and the contrastive uses of *but/men* were identified manually. *But/men* were considered contrastive when they expressed denial of expectation; opposition of two elements; modification, explanation or restriction of a previous statement; or when a speaker objected to a previous statement (Johansen 2019: 132). Non-contrastive uses for example *but/men* used to change the topic of conversation or return to a previous line of thought were disregarded. Cases where limited context made it impossible to say whether the probe was used contrastively were excluded from the analysis.

Since the probe was considered a potential trigger for the hedging strategies, only those occurring to the right of the probe were investigated. Overall, there were 659 (82%) and 582 (72%) contrastive uses of *but* and *men*, respectively. Of the 659 contrastive *buts*, 335 (51%) occurred with one or more hedging strategies. Of the 582 contrastive *mens*, 328 (56%) occurred with one or more hedging strategies. The total number of hedging strategies were 715 in the English data and 610 in the Norwegian data, produced by 157 and 158 speakers, respectively.

4. **Classification of occurrences**

Each of the 1,325 hedging strategies was classified based on its form, resulting in 13 categories. Because the study surveys the correspondence between linguistic form and pragmatic function, the strategies were grouped according to formal rather than functional criteria. Furthermore, the use of function-based categories, such as pragmatic marker, might leave some differences between the two languages undiscovered. As mentioned in Section 3, pragmatic markers constitute a broad category encompassing “everything” beyond the proposition, and forms belonging to this category may be
drawn from nearly all segments of the lexicogrammar (Fraser 1996: 167, 70). Therefore, form-based categories were used to make the repertoire of hedging strategies in the two languages more observable. The categories are presented in Table 2.

*Table 9. Overview of the categories of forms.*

<table>
<thead>
<tr>
<th>Groups of forms</th>
<th>English examples</th>
<th>Norwegian examples</th>
<th>English gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pragmatic particles</td>
<td>just, like</td>
<td>jo, da, vel, nok,</td>
<td>‘after all’, ‘then’, ‘well’,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bare</td>
<td>‘probably’, ‘just’</td>
</tr>
<tr>
<td>Adverbs</td>
<td>almost, a bit,</td>
<td>antagelig, litt,</td>
<td>‘probably’, ‘a bit’, ‘such’,</td>
</tr>
<tr>
<td></td>
<td>sort of, kind of</td>
<td>sånn, nesten</td>
<td>‘almost’</td>
</tr>
<tr>
<td>Adjectives</td>
<td>sort of, kind of</td>
<td>sånn, lille</td>
<td>‘such’, ‘little’</td>
</tr>
<tr>
<td>1/2 person pronoun +</td>
<td>I think, I suppose,</td>
<td>tror jeg, jeg føler,</td>
<td>‘think I’, ‘I feel’, ‘know you’</td>
</tr>
<tr>
<td>cognitive verb</td>
<td>you know</td>
<td>vet du</td>
<td></td>
</tr>
<tr>
<td>Tag questions</td>
<td>don’t you, isn’t</td>
<td>ikke sant, var det</td>
<td>‘not true’, ‘was it’</td>
</tr>
<tr>
<td>General extenders</td>
<td>and stuff, or</td>
<td>og alt mulig, eller</td>
<td>‘and everything possible’,</td>
</tr>
<tr>
<td></td>
<td>whatever</td>
<td>sånt</td>
<td>‘or like that’</td>
</tr>
<tr>
<td>Vague nouns</td>
<td>thing, stuff,</td>
<td>ting, greier, folk</td>
<td>‘thing’, ‘stuff’, ‘people’</td>
</tr>
<tr>
<td>Vague pronouns</td>
<td>anything,</td>
<td>man, noe, annet</td>
<td>‘one’, ‘some’, ‘other’</td>
</tr>
<tr>
<td></td>
<td>somewhere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vague determiners</td>
<td>some</td>
<td>noe</td>
<td>‘some’</td>
</tr>
<tr>
<td>Modal / semi-modal verbs</td>
<td>might, seem to</td>
<td>måtte, det virker</td>
<td>‘must’, ‘it seems like’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>som</td>
<td></td>
</tr>
<tr>
<td>Prepositional phrases</td>
<td>for me, in our</td>
<td>på en måte, for</td>
<td>‘in a way’, ‘for me’</td>
</tr>
<tr>
<td></td>
<td>eyes</td>
<td>meg</td>
<td></td>
</tr>
<tr>
<td>Clauses</td>
<td>I am no expert</td>
<td>det viste seg</td>
<td>‘it turned out’</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>yeah</td>
<td>ja</td>
<td>‘yes’</td>
</tr>
</tbody>
</table>

The pragmatic particles comprise one-word expressions which are uninflectable, often semantically and phonologically reduced, occur postverbally, cannot take stress, and have no inherent lexical-semantic meaning (Andvik 1994: iv). In previous research, the term *pragmatic particle* has been assigned to more forms than those which have been included here. Östman (1982), for example, regards all elements which connect an utterance to the speaker’s attitude as particles and includes, for example, *you know, I mean, and you see* in this category. Similarly, expressions such as *ikke sant* ‘not true’, ‘innit?’ have been labelled pragmatic particles in several studies, e.g. Svennevig (2007, 2008). However, in such studies, ‘pragmatic particle’ is often understood as a pragmatic function, i.e.
expressing the speaker’s attitude towards the content of the proposition, rather than a grammatical form. In the present study, the term is restricted to one-word expressions. In his analysis of Norwegian particles, Andvik (1992: 15) argues for not including ‘borderline cases’ such as ‛liksom’ (‘like’) in this category because they can occur with stress without losing their modal meaning. In the present study, only modal uses of ‛liksom’ etc. are studied, so ‛liksom’ and ‛like’ are regarded as pragmatic particles.

The category of adverbs includes mainly three types of adverbs: adverbs expressing uncertainty such as ‛probably’, ‛possibly’, and ‛kanskje’ (‘maybe’); downtoners such as ‛a bit’ and ‛litt’ (‘a bit’); and adverbs expressing vagueness such as ‛rundt omkring’ (‘round about’), ‛about (three weeks)’, and ‛kind of’ and ‛sort of’, when ‛kind of’ and ‛sort of’ are used to modify adjectives, verbs or clauses. The category of adjectives includes words such as ‛lille’, ‛liten’ (‘little’), (‘small’), and ‛sort of’ and ‛kind of’ when they are used as premodifiers of nouns (Quirk et al. 1985: 451).

The category first/second person pronoun + cognitive verb (1/2 pers. + Cog.V) includes expressions with the first person pronoun ‛I’/ ‛jeg’ such as ‛I think’, ‛I suppose’, ‛tror jeg’ (‘think I’) and second person pronoun ‛you’/ ‛du’ such as ‛you know’, ‛you see’, ‛du vet’ (‘you know’). Structures such as ‛I think’ are often used parenthetically as a comment modifying the proposition as a whole and also as the subject and main verb of a clause. Verbs such as ‛think’, ‛suppose’, and ‛guess’ are typically used to weaken the claim of truth to the content of the proposition (Fraser 1980: 348).

Tag questions may serve many functions, two of which can be associated with hedging: 1, expressing uncertainty or reservation; 2, seeking involvement or agreement with the interlocutor. Particularly, asymmetrical tags with (weak) rising intonation indicate uncertainty or the need for confirmation (Hübler 1983: 110). In the present study, intonation is not considered; however, all tag questions in the English material are asymmetrical and seem to express one of the two above-mentioned functions. Most of the tag questions in the Norwegian material are in the form of ‛ikke sant’ (‘not true’, ‘innit?’) which serves a similar purpose (Estling Vannestål 2015: 173).

General extenders include expressions which typically occur in clause-final position and consist of ‛and/og’ or ‛or/eller’, followed by a generic expression such as ‛things’ or ‛stuff’. They are non-specific and extend grammatically complete utterances (Overstreet 1999: 3). This structure has also been referred to as ‛set-marking tags’ (Andersen 2010) and ‛vague tags’ (Channell 1994).

The categories vague nouns, vague pronouns, and vague determiners include expressions such as ‛thing’, ‛greie’ (‘stuff’), ‛everything’, ‛anything’, ‛man’ (‘one’), and ‛noe’ (‘some’). Such expressions may be
used if precise information is not available or relevant or to express uncertainty (Clemen 1997: 240-41). Vague language is not necessarily synonymous with hedging, but, like pragmatic particles and modal verbs, it can be used for hedging. Using a vague expression could protect the speaker from making a potentially false statement. It can also be a cooperative strategy to avoid providing more information than required and thus appealing to the interlocutor by indirectly stating that you share information not explicitly expressed (Andersen 2010: 35).

Modality is another area which only partly overlaps with hedging. The group of modal or semi-modal verbs mainly includes verbs expressing epistemic modality indicating “the speaker’s confidence or lack of confidence in the truth of the proposition expressed” such as might and could (Coates 1987: 112).

The category of prepositional phrases includes expressions such as på en måte (‘in a way’), from my point of view and according to this book. The category of clauses includes structures such as if I remember correctly and det lønner seg ikke (‘it doesn’t pay off’). For both of these categories the phrase or clause as a whole creates the hedging effect, and it would not make sense to classify each word individually.

A small number of the hedging strategies found in the material did not fit into any of the categories. Expressions such as så å si ‘more or less’ and yeah were grouped as ‘miscellaneous’.

5. Results and discussion

Table 3 presents the classification of the 1,325 hedging strategies. Numbers in bold indicate significantly higher frequencies (see further below). The proportion of each category is visualised in Figure 2.
Table 10. Frequencies of each hedging category.

<table>
<thead>
<tr>
<th>Forms of hedging</th>
<th>Norwegian</th>
<th>English</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw frequencies</td>
<td>%</td>
<td>Raw frequencies</td>
</tr>
<tr>
<td>Pragmatic particles</td>
<td>251</td>
<td>41.1</td>
<td>257</td>
</tr>
<tr>
<td>Adverbs</td>
<td>121</td>
<td>19.8</td>
<td>113</td>
</tr>
<tr>
<td>Adjectives</td>
<td>25</td>
<td>4.1</td>
<td>11</td>
</tr>
<tr>
<td>1/2 pers. + Cog.V</td>
<td>74</td>
<td>12.1</td>
<td>163</td>
</tr>
<tr>
<td>Tag questions</td>
<td>18</td>
<td>3.0</td>
<td>29</td>
</tr>
<tr>
<td>General extenders</td>
<td>24</td>
<td>3.9</td>
<td>22</td>
</tr>
<tr>
<td>Vague nouns</td>
<td>29</td>
<td>4.8</td>
<td>48</td>
</tr>
<tr>
<td>Vague pronouns</td>
<td>25</td>
<td>4.1</td>
<td>20</td>
</tr>
<tr>
<td>Vague determiners</td>
<td>14</td>
<td>2.3</td>
<td>16</td>
</tr>
<tr>
<td>Modal / semi-modal verbs</td>
<td>4</td>
<td>0.7</td>
<td>27</td>
</tr>
<tr>
<td>Prepositional phrases</td>
<td>11</td>
<td>1.8</td>
<td>4</td>
</tr>
<tr>
<td>Clauses</td>
<td>12</td>
<td>2.0</td>
<td>4</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>2</td>
<td>0.3</td>
<td>1</td>
</tr>
</tbody>
</table>

|   | 610  | ≈ 100 | 715  | ≈ 100 |

^37 This category has not been tested for significance due to the low number of frequencies in both languages.
Figure 3. Proportions of the types of hedging strategies in Norwegian and English.

The results showed that various types of forms were used to express hedging in both languages, and a Fisher's exact test showed a significant association between language and form of hedging strategy \( (p = 0.0004998) \). However, closer examination of each individual category revealed that the difference was significant in only five of the 13 categories. As shown in Table 3, English speakers used significantly more 1/2 pers. + Cog.V and semi-/ modal verbs than Norwegian speakers, whereas Norwegian speakers used significantly more adjectives, prepositional phrases and clauses than English speakers. Within the 1/2 pers. + Cog. V category, most of the English realisations were either *I think* or *you know*. (See further section 5.3.) Within the semi-/ modal verb category, English speakers used a broader repertoire of modal verbs, and each verb had a higher frequency than the modal verbs used by Norwegian speakers. The most frequent verb among English speakers was *might* followed by *could* and *must*. In the Norwegian material, all the modal verbs in the category, *må* ‘must’, *måtte* ‘must’, *kunne*, ‘could’ and *det virker* ‘it seems’, occurred only once. Within the adjective category, *sånn* (‘kind of’, ‘sort of’, ‘like’) used before a noun was the most frequent, for example, *såne ting* (‘(that) kind of thing’) or *sånn klatring* (‘kind of climbing’). *Kind of* and *sort of* were the only two types of adjectives in the English data. Prepositional phrases and clauses were rare in the material, and it was thus difficult
to see any trends. An example of a recurrent prepositional phrase in the Norwegian data is på en måte (‘in a way’, ‘sort of’). An example of a clause is la oss si (‘let us say’).

The groups with the highest raw frequencies in both languages were pragmatic particles, adverbs and 1/2 pers. + Cog. V. These three categories made up 73% of the total strategies in Norwegian and 75% in English. Although, the difference between the languages was significant only in the latter category, there were several interesting qualitative similarities and differences in all three categories which will be addressed in more detail in Sections 5.1–5.3.

5.1 Pragmatic particles

Pragmatic particles were the most frequent type of hedging strategy in both languages. However, the two languages exhibited some variation within this category (Table 4). Norwegian speakers used a broader repertoire of particles, whereas English speakers mainly used like.

Table 11. Types and token within the pragmatic particle category.

<table>
<thead>
<tr>
<th>NOR</th>
<th>Gloss</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
<th>ENG</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
</tr>
</thead>
<tbody>
<tr>
<td>bare</td>
<td>‘just’</td>
<td>31</td>
<td>12.3</td>
<td>just</td>
<td>78</td>
<td>30.4</td>
</tr>
<tr>
<td>liksom</td>
<td>‘like’</td>
<td>53</td>
<td>21.1</td>
<td>like</td>
<td>167</td>
<td>65.0</td>
</tr>
<tr>
<td>jo</td>
<td>‘after all’, ‘you know’, ‘of course’</td>
<td>135</td>
<td>53.8</td>
<td>well</td>
<td>12</td>
<td>4.7</td>
</tr>
<tr>
<td>Other (e.g. da, vel)</td>
<td>‘surely’, ‘well’</td>
<td>32</td>
<td>12.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>251</td>
<td>≈ 100</td>
<td></td>
<td>257</td>
<td>≈ 100</td>
</tr>
</tbody>
</table>

The form like accounted for 65% of the total number of pragmatic particles in the English material. Beeching (2016: 128) distinguishes five functions of like, one of which is the ‘hedging like’. This like represents hedging in a narrow sense. Some of the other types of like could also have a hedging function if hedging is defined in a wide sense. Moreover, like may serve several functions at the same time. The first like following but in (2) occurs as part of the structure BE like and is an example of quotative like. The second like is classified as a focusing like (Beeching 2016: 127) and thus used not to hedge but to draw attention to a certain piece of information.
(2) **but** it’s *pretty much* me who’s impacted then I’m always gonna be *like* oh I’ll do intellectual job because that’s what I wanna do [...] and it’s *like* I’m not ready to make those kind of choices yet

(BNC2014 SHHG no. 99)

*Liksom*, which often corresponds to *like* (Hasund 2003: 2-8), was the second most used particle in the Norwegian material. *Liksom* in initial or medial position is often used to modify any following linguistic elements (Hasund 2003: 124, 70). In final position, it is typically used to signal epistemic stance in relation to the preceding expression (Hasund 2003: 199), as in (3). However, although *liksom* and *like* seem to overlap considerably in use and pragmatic function, there are some important differences. For example, *liksom* occurs more frequently in final position than *like*, and final *liksom*, unlike final *like*, may be used as a hedging device to avoid sounding too categorical or to soften a potential face-threat (Hasund 2003: 207).

(3) **men** det er *jo* dårfen e # hårfint skille mellom drittsekk og # bevege seg på den grensa *liksom*

“*but there is* [pragmatic particle] a fine e # fine line between being a jerk and # balancing on that line *like*”

(BB Rodney> <who_avfile 12 no. 20)

The most frequently used particle in the Norwegian material was *jo*. *Jo* serves many functions: for example, it may indicate that the propositional information is given, uncontroversial, or shared by the interlocutors or that what is stated is to be interpreted as “a premise for and deriving and supporting an available conclusion” (Berthelin and Borthen 2019: 22). In (3), two speakers are talking about another person who they both think is behaving badly. The pragmatic particle *jo*, occurring sentence-internally, implies that there is consensus about the proposition (Andvik 1992: 85; Berthelin and Borthen 2019: 8). Thus, *jo* may serve some of the same pragmatic functions as *you know*. For example, sentence-internal *jo* and *you know* may both be used to create common ground between speakers (Berthelin and Borthen 2019: 1; Beeching 2016: 98). Furthermore, *you know* has a literal counterpart in Norwegian *du vet*, which is also used for various pragmatic purposes but is less common as *you know* (cf. 5.3).

5.2 Adverbs

The choice of adverb types also varied between the languages as shown in Table 5. Overall, English speakers made use of a broader variety of hedging adverbs than Norwegian speakers.
Table 12. Types and tokens within the adverb category.

<table>
<thead>
<tr>
<th>NOR</th>
<th>Gloss</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
<th>ENG</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adverbs expressing uncertainty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>kanskje</code></td>
<td>'maybe'</td>
<td>11</td>
<td>9.1</td>
<td><code>probably</code></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Other (e.g. sikkert, antagelig)</td>
<td>'surely', 'probably'</td>
<td>11</td>
<td>9.1</td>
<td><code>maybe</code></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Other (e.g. perhaps)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Downtoning adverbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>litt / lite grann</code></td>
<td>'a bit'</td>
<td>42</td>
<td>34.7</td>
<td><code>a bit</code></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td><code>ikke så / ikke noe særlig, etc.</code></td>
<td>'not so much'</td>
<td>13</td>
<td>10.7</td>
<td><code>not so / not that / not very, etc.</code></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Other (e.g. nesten)</td>
<td>'almost'</td>
<td>5</td>
<td>3.7</td>
<td>Other (e.g. reasonably)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Adverbs expressing vagueness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><code>sånn / sånt</code></td>
<td>'about', 'like'</td>
<td>30</td>
<td>24.8</td>
<td><code>kind of</code></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Other (e.g. her og der)</td>
<td>'here and there'</td>
<td>4</td>
<td>3.3</td>
<td><code>about</code></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Sort of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Miscellaneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e.g. ganske, gjerne, egentlig</td>
<td>'reasonably', 'pretty', 'really'</td>
<td>1</td>
<td>0.8</td>
<td>e.g. <code>basically, technically, really</code></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>SUM</td>
<td></td>
<td>121</td>
<td>≈ 100</td>
<td>113</td>
<td>≈ 100</td>
</tr>
</tbody>
</table>

*Kanskje* (`maybe`), *probably*, and *maybe* are the most frequent adverbs expressing uncertainty; they may serve the same function as parenthetical constructions such as *I think* and *jeg tror*, i.e. to reduce commitment on behalf of the speaker (Hübler 1983: 119). They express the speaker’s attitude towards the propositional content. *Probably* is often used to express strong doubt about the truth of the proposition (Hübler 1983: 124), which is exemplified in (4), where the speaker speculates about someone else’s budget. Similarly, in (5), the speaker is uncertain whether it is better to have a cat or a dog.

(4) [...] whatever your budget you start with you can double it and that’s closer on what it would be *but* theirs *probably* was more [...]  

(BNC2014 SDJ9 no. 46)
Among the downtoning adverbs, *litt* (‘a bit’, ‘a little’), and *a bit* were the most frequent. In (6), the speaker modifies the statement by using *litt* repeatedly. The speaker partially agrees with the interlocutor, but is more positive towards the topic they are talking about. By using *litt* to tone down the meaning, the speaker reduces the risk of disagreeing and thereby creating a face-threatening situation. Similarly, in (7), the speaker uses *a bit* to modify the description of another person’s reaction.

(6)  A: […] # det blir nå godt når begynner å lysne litt snart nå  
B: ja *men* det var jo *litt* skjønt i dag da når sola var framme *litt* # det hjelper *litt*  
“A: […] # it will be good when it starts getting a bit brighter soon  
B: yes *but* it was a *bit* lovely today when the sun was out *a bit* # it helps *a bit*”

(7)  A: we I mean you've got an amazingly efficient rubbish collection system here  
[…]  
B: >> do you use the food waste one thing?  
A: yeah because well we did set up a compost thing *but* then --ANONnameF and --ANONnameM got *a bit*  
B: >> you couldn't do it  
A: >> flustered about it they said it attracts the rats

(8) Among the vague adverbs, *sånn* (‘such’, ‘about’, ‘sort of/kind of’, ‘like’) was the most frequent in Norwegian, and *kind of* in English. Vague expressions may be used to withhold information to avoid committing oneself, express the lack of the necessary words or phrases for the concepts the speaker wishes to convey, express uncertainty, protect oneself from being proven wrong, and avoid face-threatening situations (Channell 1994: 173 ff). Norwegian *sånn* is a very versatile form. As a hedge, it can be used to indicate reservations on behalf of the speaker with respect to the appropriateness of the lexical expression or “to signal that the referent or the lexical item may not be completely familiar to the speaker or the hearer” (Svennevig 2010: 178). In (8) the speaker is arguing that the place Bufjell is nicer than Holmetjønn because it is more remote but the speaker is using the word *inneklemt*
‘sandwiched’ which may have negative connotations and thus not be the appropriate word: sånn signals this inappropriateness.

(8)    A: for det at eller det er fint på Holmetjønn # misforstå meg riktig liksom # men det er et eller annet med Bufjell for det at det er sånn

B: * ja ja * nei nei men det er fint

A: litt sånn inneklemt på en måte

“A: for it that or it is nice at Holmetjønn # don’t get me wrong like # but there is something about Bufjell because it that it is like

B: * yes yes* no no but it is nice

A: a bit like sandwiched kind of”

(NDC rollag_02uk><who_avfile rollag_01um-02uk no. 666)

Kind of and sort of overlap in both meaning and use (Aijmer 2002: 207). Kind of thus has both an epistemic and affective function. For example, it can indicate that the speaker does not know the appropriate or correct word to use or be used to establish common ground to avoid disagreement (Aijmer 2002: 191). In (9), kind of may serve both of these purposes.

(9)    A: and they’ve both kind of forgotten about it erm but obviously I think it's been kind of

B: yeah it’s been in the background

A: yeah in the back of his mind […]

(BNC2014 SVBH no. 65)

Overall, Norwegian speakers used significantly more downtoning adverbs than English speakers (p < 0.05), whereas English speakers used significantly more adverbs of uncertainty than Norwegian speakers (p < 0.01). The wish to tone down and create as little imposition as possible may be related to Norwegian politeness. There is limited research on how politeness is expressed in Norwegian, but Rygg (2017: 1) describes Norwegian politeness as showing consideration by not imposing on people. Correspondingly, native speakers of English use more expressions of uncertainty when a situation is perceived as face-threatening (Holtgraves and Perdew 2016: 8).
5.3 1/2 pers. + Cog.V

English speakers used 1/2 pers. + Cog.V significantly more than Norwegian speakers. This structure was also the second most frequent strategy in the English data and the third most frequent strategy in the Norwegian data. Table 6 presents the forms used in both languages.

Table 13. Types and token within the 1/2 pers. + Cog.V category.

<table>
<thead>
<tr>
<th>NOR</th>
<th>Gloss</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
<th>ENG</th>
<th>Raw frequencies</th>
<th>% of total in category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person pronoun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jeg tror / tror jeg</td>
<td>‘I think’, ‘think I’</td>
<td>37</td>
<td>50.0</td>
<td>I think</td>
<td>48</td>
<td>29.4</td>
</tr>
<tr>
<td>Other (e.g. jeg vet ikke, jeg føler)</td>
<td>‘I don’t know’, ‘I feel’</td>
<td>17</td>
<td>23.2</td>
<td>Other (e.g. I suppose, I feel)</td>
<td>25</td>
<td>15.3</td>
</tr>
<tr>
<td>2nd person pronoun</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>du vet / vet du</td>
<td>‘you know’, ‘know you’</td>
<td>18</td>
<td>24.3</td>
<td>you know</td>
<td>62</td>
<td>38.0</td>
</tr>
<tr>
<td>kan du si</td>
<td>‘you might say’</td>
<td>2</td>
<td>2.7</td>
<td>Other (e.g. you see)</td>
<td>4</td>
<td>2.4</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>74</td>
<td>≈ 100</td>
<td>163</td>
<td>≈ 100</td>
<td></td>
</tr>
</tbody>
</table>

*Jeg tror/ tror jeg* (‘I think’/ ‘think I’) was the most frequent form within this category in Norwegian, *I think* was the second most frequent in English. These structures were often used parenthetically modifying the proposition as a whole as in (10) and (11).

(10) I don’t know what happened there **but** I paid well over the odds for it *I think*
(BNC2014 S4YQ no. 290)

(11) A: *m jeg syns ikke de skulle fornekte at det # har vært en krigshistorie der
B: nei *men* dette ha- dette har vært en trend ei stund nå *tror jeg* fordi at em vi ser det samme oppi # oppi Sjåk

“A: *m I don’t think they should deny that there # has been a war story there
B: no *but* this ha- this has been a trend for a while now *I think* because em we see the same up in # up in Sjåk”

(NoTa 067><who_avfile 067-068 no. 822)
You know was the most frequent form in the English material whereas du vet/vet du ‘you know’ / ‘know you’ was the second most frequent in the Norwegian material. The core function of the pragmatic you know is to create common ground between the speakers (Beeching 2016: 98). You know in final position is commonly used to seek agreement from the interlocutor (Beeching 2016: 103). In (12), you know is used to seek some kind of confirmation from the interlocutor, which the speaker also gets in the affirmative yeah.

(12) A: (...) she raised me (.) albeit in a you know
     A: in a bad way but
     B: >> a cack handed way yeah
     A: she still did it and and there was a lot of love there for me you know
     B: >> yeah

(BNC2014 SEPP no. 613)

Although the Norwegian du vet/vet du can establish common ground in the same way as you know, e.g. as in (13), it is not as widely used.

(13) A: hæ # ordne håret? jeg skal jo spise middag
     B: [...]* ja men
     B: _latter_ må style håret ditt vet du # s- kom da # ja jeg syns det var morsomt jeg _uforståelig_ kom da
     “A: what # do my hair? I am having dinner
     B: [...]* yes but
     B:_laughter_ have to style your hair you know # s- come then # yes I think it was funny I_incomprehensible_come then”

(BB Anette><who_avfile 19 no. 42)

Jeg vet ikke (‘I don’t know’) was the third most frequent form in the Norwegian material, although it only occurred five times. I mean, the third most frequent form in the English material, occurred 24 times. Similar to you know, I mean often serves multiple pragmatic functions at the same time, but its core pragmatic function is to establish and negotiate meaning with the interlocutor (Beeching 2016: 185). In (14), I mean could serve the following functions: hedging, i.e. softening the strength of the utterance, clarification, exemplification, elaboration, reformulation, or justification for the attitude expressed in the first part of the utterance or a preceding utterance.

(14) A: well --ANONplace people are sort of thick --ANONplace dumplings
B: no necessarily though
A: no I know **but** that's the impression everybody get (..) *I mean*
B: >> yeah
C: >> yeah
A: it's the same with Dorset people *isn't it*

(BNC2014 S99N no. 322)

6. Concluding remarks

This study compared the forms of hedging strategies in English and Norwegian informal conversations aiming to strengthen the empirical basis for the study of hedging in these languages. By using the probe *but/men*, 1,325 hedging strategies were retrieved from four spoken corpora and subsequently grouped together into 13 categories based on their form. Although the probe did not give access to all hedging strategies in the corpora, it facilitated a study of a broader range of strategies than the more traditional form-to-function approaches and simultaneously ensured that the strategies occurred in the same type of context in both languages (Johansen, 2019).

The analysis and comparison of the results revealed both quantitative and qualitative differences between the languages. Overall, 1/2 pers. + Cog. V and semi-/modal verbs were used significantly more by English speakers than by Norwegian speakers, whereas adjectives, prepositional phrases, and clauses were used significantly more by Norwegian speakers. However, the majority of hedging strategies in the material was classified as either pragmatic particles, adverbs, or 1/2 pers. + Cog. V. Norwegian speakers used a wider range of pragmatic particles than English speakers, whereas English speakers had a broader repertoire of adverbs. Norwegian speakers used significantly more downtoning adverbs than English speakers, whereas English speakers used more adverbs expressing uncertainty than Norwegian speakers. English speakers used significantly more 1/2 pers. + Cog. V than Norwegian speakers and whereas *jeg tror* (‘I think’) was the most frequently used form in Norwegian, *you know* was the most frequently used form in English.

Building on the results of this study, some interesting topics for further research may be suggested. For example, the most frequently used Norwegian pragmatic particle was *jo*, which does not have a literal translation into English, but which serves many of the same pragmatic functions as *you know*, which in turn was the proportionally most frequent form in the 1/2 pers. + Cog. V category.
It would be interesting to explore the degree of overlap in meaning and use between these two forms. Furthermore, Norwegians seem to show politeness through minimising the degree of imposition, and this may be reflected in the significant use of downtoning adverbs. Such cultural implications of hedging could be explored further.
References

——. 1994. Recourse to Consensus in Norwegian Modal Particles, University Microfilms. The University of Texas, Arlington, Texas.
"BigBrother-korpuset." Tekstlaboratoriet, ILN, Universitetet i Oslo.
——. 2018. ‘It was a bit stressy as well actually’. The Pragmatic Markers actually and in fact in Spoken Learner English, Journal of Pragmatics.
Kozubíková Šandová, J. 2017. “...and our study might therefore have been slightly underpowered”: A Cross-Linguistic Analysis of Hedging in English and Czech Medical Research Articles, Acta Universitatis Carolinae Philologica: 115-30.


"Norsk talespråkskorpus - Oslodelen." Tekslaboratoriet, ILN, Universitetet i Oslo.


Rygg, K. 2017. ‘Typically Norwegian to be Impolite.’ Impoliteness according to whom?, 2017, 4.


Article 4

Title: Hedging in spoken conversations by Norwegian learners of English
Author: Stine Hulleberg Johansen
Status: Submitted to Nordic Journal of Language Teaching and Learning (26 May 2020)
Hedging in spoken conversations by Norwegian learners of English

Abstract
Hedging is an important aspect of pragmatic competence, but it is also a complex phenomenon that is difficult to master. This has resulted in underuse of hedging strategies by language learners, and many learners limit their hedging repertoire to a few strategies. This study compares the use of 10 hedging expressions commonly used in informal spoken English, *a bit, I mean, I think, just, kind of/kinda, like, might, probably, thing(s), and you know,* in data from LINDSEI-no, a corpus of Norwegian advanced learners, and LOCNEC, a comparable corpus of native English speakers. Norwegian learners typically show a high level of grammatical competence, but research on their pragmatic competence is limited. This study adds to the empirical research on Norwegian advanced learners by comparing the use of these expressions in native and learner language. The results indicate that as a group, Norwegian learners underuse hedging strategies, but when each expression and individual variation are considered, the picture is more nuanced. In fact, several of the Norwegian learners’ hedging practices partially overlap with several of the native speakers’ practices concerning hedging frequency and types of hedging strategies used.

Keywords: Hedging, Learner Corpus Research, Norwegian Learners, Corpus Pragmatics

1. Introduction
Norwegian learners generally perform well on European language tests measuring various aspects of grammatical competence; however, less is known about their pragmatic competence. Learners’ grammatical and pragmatic competence tends to develop unevenly (Romero-Trillo, 2002, 70). Whereas acquiring pragmatic competence in one’s first language is “an unconscious process which co-occurs with socialisation and acculturation into the community into which one is born” (Holmes et al., 2020, 1), learners in a non-target environment acquire the target language asynchronously through formal instruction. However, as pragmatic competence is difficult to integrate in educational syllabi, it is not always given the attention it deserves in second- and foreign-language teaching. Language learners may therefore produce “grammatically flawless speech that nonetheless fails to achieve its

---

communicative aims” (Fraser, 2010, 15). The higher the level of fluency, the greater the risk of communicative failure due to the underlying expectation that non-native speakers who master the target language on the surface also master that language’s underlying social conventions (Nikula, 1997, 188–189).

This study investigates hedging, a complex aspect of pragmatic competence that may cause problems for language learners. Hedging may be defined as discourse strategies that attenuate the force, truth, or effects on the hearer (Hyland, 1996, 1998; Kaltenböck et al., 2010) and may be realised through nearly any linguistic form. As with many pragmatic features, these forms often do not correspond across languages (Romero-Trillo, 2018), and such cross-linguistic differences may lead learners to favour certain pragmatic expressions. Several studies of learners with various first languages (L1s) have demonstrated the over- and underuse of various hedging expressions (e.g. Nikula, 1997). Moreover, previous research on hedging by lower secondary school Norwegian learners showed that learners use fewer hedging strategies than their native English counterparts and exhibit less variation in their hedging repertoires (Hasselgreen, 2004). Still, Norwegian learners’ hedging strategies remain under-researched (Vold, 2006). To date, there are no corpus-based studies devoted primarily to hedging in spoken English by Norwegian advanced learners.

Previous comparative studies of learner and native varieties have been criticised for focusing on frequencies as the only diagnostic of importance (Paquot & Plonsky, 2017) and for assuming the homogeneity of non-native and native speaker groups (Gablasova et al., 2017). Thus, studies are needed that nuance observed differences at the group level by looking at, for example, dispersion (Gries, 2008).

This study aims to strengthen the empirical research on hedging by studying 10 hedging expressions, some of which have never been investigated before, in Norwegian advanced learners’ spoken English. The investigation concerns the use of *a bit, I mean, I think, just, kind of/kinda, like, might, probably, thing(s), and you know* in data from the Norwegian Louvain International Database of Spoken English Interlanguage (LINDSEI-no)\textsuperscript{39} and the comparable Louvain Corpus of Native English Conversation (LOCNEC) (De Cock, 2004).\textsuperscript{40} For the rationale behind these 10 words, see section 3. Inter- and intraspeaker variation are considered to see whether Norwegian advanced learners show signs of underuse, as observed in studies of lower-proficiency learners, or whether their use of hedging

\textsuperscript{40} https://uclouvain.be/fr/node/11973
strategies is similar to that of native speakers. Specifically, the study tackles the following research questions:

RQ1: Do Norwegian advanced learners use fewer instances of hedging strategies (tokens) than native English speakers?

RQ2: Do Norwegian advanced learners use a more limited range of expressions (types) than native English speakers?

Section 2 of this paper outlines developments in hedging research and presents studies of hedging in learner language. Section 3 describes the materials and methods. Section 4 presents and discusses the results, and section 5 offers concluding remarks.

2. Hedging as an aspect of pragmatic competence

Pragmatic competence has been described as “the linguistic ability to communicate your intended message with all its nuances in any socio-cultural context and to interpret the message of your interlocutor as it was intended” (Fraser, 2010, 15). Mastering hedging is merely one aspect of this competence. Nevertheless, hedging is critical for communication success (Fraser, 2010, 15). Hedging strategies allow the speaker to tone down a potentially face-threatening proposition through expressions of tentativeness or possibility and to avoid sounding too categorical by minimising commitment to a proposition. Hedging may also mitigate or reduce negative effects on the hearer (Hyland, 1998, 177). Hedging strategies may have a propositional scope, affecting the content of the proposition, as exemplified in (1). They may also say something about the relationship between the speaker and the proposition (speaker-oriented), as in (2), or affect the relationship between the speaker and the hearer (hearer-oriented), as in (3).

(1) I kind of changed my mind (LOCNEC EN0029)
(2) I think it was in Berlin or somewhere (LOCNEC EN009)
(3) […] it depends how good it is you know (LOCNEC EN0022)

Research on hedging in English interlanguage is relatively limited compared to that in native English (Neary-Sundquist, 2013, 153), and as with most corpus linguistic studies, studies on written language outnumber those on spoken language (Adolphs & Carter, 2013, 1). Overall, studies on hedging in written learner language show that learners tend to use a limited range of hedging strategies and fewer strategies than native speakers. For example, Ventola and Mauranen (1990) found that Finnish learners used a limited range of hedging expressions in academic writing. In a study of indirectness in L1 and
L2 academic writing, Hinkel (1997) found that Chinese, Korean, Japanese, and Indonesian learners used certain hedges more frequently than native speakers of English but did not use other indirectness devices. A study of English papers by Lithuanian undergraduate students showed that learners preferred the modal verb *can* over other hedging strategies (Šeškauskienė, 2008).

Studies of hedging in spoken learner English generally support the tendencies indicated by studies of hedging in learner writing. For example, Kasper (1979) studied German learners’ use of hedging strategies and found that modality is present in speaker planning but not necessarily in speaker production. Kärkkäinen (1990) compared the use of epistemic modality expressions among Finnish learners and native speakers of English and discovered that the learners used fewer hedging strategies and had less variation in their repertoire than native English speakers. The results also indicated that the limited use of strategies was not necessarily a result of socio-pragmatic transfer, as the Finnish learners frequently used hedging strategies when speaking Finnish. Nikula (1997) also studied hedging in spoken English by Finnish learners and found that they used fewer strategies than native English speakers. In a recent study of hedging strategies in learner and teacher talk in a classroom setting, Friginal, Lee, Polat, and Roberson (2017) found that L2 speakers of English only used 20 out of 102 potential hedging expressions examined, whereas the teachers used about half.

Studies of hedging by Norwegian learners are relatively scarce, particularly of spoken interlanguage. However, the existing studies of spoken discourse indicate that Norwegian learners generally underuse hedging expressions in English. The perhaps most elaborate study of Norwegian learners, Hasselgreen’s (2004), compared the use of eight expressions (*I think, like, sort of/kind of, a bit, just, or something, not really*, and *and everything/that/stuff/things*) in Norwegian lower secondary school learners and native English speakers (Hasselgreen, 2004, 205). The comparison showed that the number of tokens was generally lower among the Norwegian learners (Hasselgreen, 2004, 208), and *I think* emerged as the learners’ preferred hedging strategy. In an unpublished master’s thesis, Sandal (2016) reported a general underrepresentation of the discourse marker use of *like, well*, and *you know* among Norwegian advanced learners of English (Sandal, 2016, 116). However, Sandal looked at pragmatic functions beyond those of hedging.

Most studies of hedging in written discourse by Norwegian learners only discuss hedging as one of several pragmatic functions of particular expressions, e.g. Hasselgård & Johansson’s (2011) discussion of *seem*. Thomson (2016) studied hedging strategies in written English by Norwegian lower secondary school pupils and found that the pupils had a good understanding of hedging devices but used a limited
range of hedging expressions (Thomson, 2016, 5). The results also showed significant individual variation among pupils in the use of hedging strategies.

Despite these studies, hedging by Norwegian learners remains under-researched, and more research is needed to better understand the hedging competence of Norwegian advanced learners of English. The study proposed here contributes to this research by investigating 10 hedging expressions that a recent corpus-based study of hedging strategies in spoken native English conversations identified as important (Johansen, under review). The study aims to investigate the extent to which more advanced learners still underuse hedging expressions, as pupils in lower secondary school have been found to do, the extent to which this is true for individual expressions, and whether the learners use a narrower range of expressions than native speakers.

3. Materials and methods

3.1 Materials

The 10 forms investigated here come from a previous study of hedging strategies in informal conversations using data from the spoken British National Corpus 2014 (Johansen, under review; Love et al., 2017). Through a partly bottom-up approach, the previous study retrieved 715 hedging strategies, and the 10 proportionally most frequent forms among these strategies serve as the starting point for this study. The forms investigated are a bit, I mean, I think, just, kind of/kinda, like, might, probably, thing(s), and you know.

To investigate Norwegian learners’ use of these hedging strategies in spoken English, data from the Norwegian LINDSEI corpus, LINDSEI-no, was used. LINDSEI, one of the most recognised learner corpora (Romero-Trillo, 2018, 115), was created for systematically studying the spoken English of learners with different L1s. The LINDSEI corpus consists of spoken discourse produced by university undergraduates and learners of English with an advanced proficiency level (Gilquin et al., 2010, 7). LINDSEI-no was not part of the original corpus launched in 1995 but was compiled following the same format by Inland Norway University of Applied Sciences between 2010 and 2015. The LINDSEI-no data is compared to data from the comparable LOCNEC corpus of native-speaker English.

The LINDSEI and LOCNEC corpora consist of informal interviews. Each interview includes three speech tasks, of which a free informal discussion is considered the main part (Gilquin et al., 2010, 3). This study uses only this free discussion part because the free discussion part was considered most
similar to authentic conversation, the “genre” from which the expressions investigated were drawn, and only the learners’ utterances are included. Table 1 presents the number of words in the sample.

Table 1 No. of words in the corpora and sample

<table>
<thead>
<tr>
<th></th>
<th>Total no. of words of learner language</th>
<th>Total no. of words in free discussion (tokens)</th>
<th>Total no. of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINDSEI-NO</td>
<td>122,956</td>
<td>43,431</td>
<td>50</td>
</tr>
<tr>
<td>LOCNEC</td>
<td>126,666</td>
<td>70,957</td>
<td>50</td>
</tr>
</tbody>
</table>

3.2 Methods

The corpus data was analysed using WordSmith 7 following the form-to-function approach commonly applied in corpus pragmatic studies. This approach uses previously identified lexical constructions as the starting point for the study and examines their potential pragmatic functions in natural text (Aijmer & Rühlemann, 2015). All instances of the chosen words were read vertically using the keyword in context function, and all canonical uses were removed. The pragmatic uses were investigated more carefully to determine whether they were used as hedges or for other purposes. When it was difficult to determine whether a word was used as a hedge, more context was consulted.

An inclusive approach was chosen with respect to what was considered a hedge. Many of these words serve multiple pragmatic functions simultaneously, and it is often difficult to determine a word’s exact function(s), even in context. When a word’s function was ambiguous, the word was included as a hedge in the analysis of both the learner and the native speaker data to ensure the validity of the comparison.

Example (4) presents an ambiguous use of just. Pragmatic uses of just include both intensifying and toning-down uses, and its purpose is often difficult to determine. The pragmatic use of just is also closely related to its canonical use in the sense of merely, and clearly distinguishing these two uses may be difficult.

(4) (erm) .. I don't find it very easy but I love the language […] I loved it at school (erm) and I I get a real buzz out of speaking the language […] or I did when I was well I did when I was in France but not when I'm in England it doesn't get you anywhere […] I just enjoyed it (LOCNEC EN0032)

Following the identification of the hedging uses, relative frequencies were compared across the two varieties. In line with previous studies on learner corpora (e.g. Aijmer, 2011), a log-likelihood test compared the raw frequencies of the forms.41 However, corpus studies of learner language have been criticised for being too focused on frequency, which may be misleading when used as the only measure

41 http://ucrel.lancs.ac.uk/llwizard.html
to describe the data (Gries, 2008, 404). Paquot and Plonsky (2017) identify the focus on frequencies as a shortcoming of many learner corpus research studies. Similarly, Gablasova et al. (2017, 133–135, 137) argued that to make sound comparisons, interspeaker variation should also be considered.

Considering interspeaker variation in addition to frequency is important not only for the study of learner language. It is equally important to look at interspeaker variation in L1 speakers, as they are not necessarily homogenous (Gablasova et al., 2017, 141). This study examined intra-speaker variation in both varieties.

4. Results and discussion

Of the 1,259 instances of the 10 expressions retrieved from LINDSEI-no, 853 were hedging uses. Of the 3,037 instances retrieved from LOCNEC, 2,357 were hedging uses. Table 2 presents the total number of hedging strategies per language variety and the frequencies for each form in each variety.

Table 2 Frequencies for all forms in both varieties

<table>
<thead>
<tr>
<th></th>
<th>LINDSEI-NO</th>
<th>LOCNEC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw freq.</td>
<td>Freq. per 1,000 words</td>
</tr>
<tr>
<td>a bit</td>
<td>50</td>
<td>1.2</td>
</tr>
<tr>
<td>I mean</td>
<td>30</td>
<td>0.7</td>
</tr>
<tr>
<td>I think</td>
<td>148</td>
<td>3.4</td>
</tr>
<tr>
<td>just</td>
<td>211</td>
<td>4.9</td>
</tr>
<tr>
<td>kind of kinda</td>
<td>65</td>
<td>1.5</td>
</tr>
<tr>
<td>like</td>
<td>142</td>
<td>3.3</td>
</tr>
<tr>
<td>might</td>
<td>17</td>
<td>0.4</td>
</tr>
<tr>
<td>probably</td>
<td>38</td>
<td>0.9</td>
</tr>
<tr>
<td>thing(s)</td>
<td>89</td>
<td>2.1</td>
</tr>
<tr>
<td>you know</td>
<td>53</td>
<td>1.2</td>
</tr>
<tr>
<td>Total</td>
<td>853</td>
<td>≈ 100</td>
</tr>
</tbody>
</table>

4.1 Do Norwegian advanced learners use fewer instances of hedging strategies (tokens) than native English speakers?

Overall, at a group level, Norwegian learners hedged less often than native English speakers did. As can be seen from Table 2, the total number of hedging strategies was significantly higher for native speakers than for learners (p < 0.0001), indicating an underuse of hedging strategies by Norwegian learners. This result is in line with the expectations outlined in section 1 and corroborates previous research on learners in general and Norwegian learners in particular. For example, in her study of
Finnish learners of English, Nikula (1997) found that learners hedged less often than native English speakers. Similarly, previous research on Norwegian learners has shown that they hedge less often than English native speakers. Hasselgreen (2004) found that Norwegian learners generally used fewer instances of hedging strategies than their native English counterparts. However, Hasselgreen (2004) studied Norwegian learners at the lower secondary school level; thus, the data is not directly comparable to that of the current study. The results are also in line with those of Sandal (2016), who discovered an underrepresentation of all pragmatic uses of like and you know by Norwegian advanced learners. As shown below, cross-linguistic influence from the Norwegian learners’ L1 may help explain this pattern.

However, if we look at individual expressions presented in Table 2, we have to nuance our claim about underuse. Although five out of 10 expressions (a bit, I mean, like, thing(s), and you know) are underused, there is no statistically significant difference for four of the expressions (I think, just, might, and probably), and in one case (kind of), the data actually show overuse by the Norwegians.

Among the hedging strategies where the non-native and native use diverge, the biggest difference is found in the use of you know and I mean. Previous research has indicated that learners prefer explicit hedging strategies, those that refer explicitly to the speaker (or hearer) (Nikula, 1997) and automatically introduce “an explicit argumentative perspective to the discourse” (Baumgarten & House, 2010, 1185). Therefore, we might have expected Norwegian learners to prefer such expressions.

A potential explanation for the low number of you know and I mean in the Norwegian learner data could be that the direct translations of these two expressions into Norwegian du vet/vet du (‘you know’, ‘know you’) and jeg mener (‘I mean’), although they can be used to hedge, have not undergone the same pragmaticalisation process as you know and I mean in English (Aijmer, 1997, 2) and thus are not used to the same extent in Norwegian (Johansen, under review). Norwegian learners may transfer the conventions of use for these strategies into English, which could explain the underrepresentation of you know and I mean. Furthermore, the modest number of occurrences of you know among the Norwegian learners may also relate to the mismatch in form between you know and the Norwegian pragmatic particle jo (‘of course’, ‘after all’), although jo shares many pragmatic functions with you know, such as signalling that the information in the proposition is given or shared by the interlocutors. Jo does not have a direct translation into English, and advanced learners might be unaware of this overlap in functions between jo and you know and therefore not use you know in some situations where they would have used jo. Jo is frequent in Norwegian informal speech (Berthelin & Borthen, 2019, 1), but you know is underrepresented in English by Norwegian learners.
Among the hedging strategies for which there was no statistical difference between L1 and L2, *I think* and *just* were the most frequently used. Previous research on *I think* in native English has shown that it is “almost ubiquitous” as a stance-marker in spoken L1 English (Baumgarten & House, 2010, 1186). Influence from English may be a reason for the high *I think* use in the Norwegian learner data; however, frequency in input does not seem to be the only factor influencing Norwegian learners’ performance. Otherwise, we might have expected a higher number of other frequently used expressions, such as *you know* and *I mean*, which are underrepresented. The relatively high occurrence of *I think* may be due to its close resemblance to the Norwegian expression *jeg tror* (‘I think’), which is frequently used to express uncertainty in Norwegian. Thus, *I think* differs from *I mean* and *you know*. In addition to resembling the Norwegian uncertainty marker *jeg tror*, *I think*, like *I mean* and *you know*, is an explicit way of expressing uncertainty or lack of commitment to the proposition. By using *I think*, the speaker expresses his/her feelings, attitudes, value judgements, and assessments towards the proposition in addition to the propositional content (Baumgarten & House, 2010). Hasselgreen (2004, 208) suggests that the explicit nature of *I think* could be a reason for its frequent use by Norwegian learners.

The frequent use of *just* by Norwegian learners is in line with previous research on Norwegian learners. Hasselgreen (2004) reported that *just* is frequent among both Norwegian lower secondary school learners and English native speakers; *just* and *I think* are the two most favoured hedges among the learners in her data. One possible explanation for the high use of *just* in the Norwegian learner data could be that *just* is one of several translations of the Norwegian pragmatic marker *bare* (‘just’, ‘merely’), which is frequently used in informal spoken discourse (Hasund et al., 2012; Opsahl, 2002; Opsahl & Svennevig, 2007). *Bare* is used to express several pragmatic functions in Norwegian. When used as a hedge, *bare* reduces the speaker’s commitment to the proposition and has a toning-down function (Hasund et al., 2012, 45).

Transfer from Norwegian cannot easily explain the overuse found with *kind of*, the only form occurring more frequently among the Norwegian learners than among native speakers. (There were no instances of the contracted form *kinda* in either of the corpora.) Previous research on *kind of* has indicated that although *kind of* and *sort of* seem to have the same meaning (Kay, 1997, 151), *kind of* is not used to the same extent as *sort of* in British English (Aijmer, 2002, 207), which is the variety of the LOCNEC speakers. However, both *kind of* and *sort of* are more frequent in spoken discourse than in written discourse, and *kind of* is the more common of the two in spoken discourse (Gries & David, 2007). Influence from spoken American English may thus be a reason for the frequent use of *kind of* by Norwegian learners. The English spoken by Norwegian learners is heavily influenced by American
television and is often humorously characterised as being somewhere in the mid-Atlantic, meaning that Norwegian learners incorporate features from both American and British English.

The underuse by Norwegian learners illustrated in Table 2 becomes even more nuanced when we consider interspeaker variation. Figure 1 shows the total number of hedges (tokens) per speaker in both varieties.

![Figure 4 Total number of hedges (tokens) per speaker.](image)

Although most Norwegian learners are outside the interquartile range of the native speaker data, 16% of the Norwegian learners (eight out of 50 speakers) are within the 75th percentile of the native speakers. Furthermore, 18% of the native speakers (nine out of 50 speakers) use 22 or fewer strategies and thus resemble many of the Norwegian learners, indicating that a number of the Norwegian learners display a native-like frequency of hedging.

If we look at the dispersion of each hedging strategy (Figures 2 and 3), native speakers generally exhibit more variation than Norwegian learners. For the learners, most of the observations are concentrated at the lower end of the scale, and the medians are generally lower than for the native speakers. However, several native speakers are quite low on the scale, which means that the Norwegians are, to a large extent, within the range of native speakers.
For eight out 10 expressions (a bit, I think, just, like, might, probably, thing(s), and you know) there is partial overlap between the first quartile in the native speaker data and the third quartile in the learner data, which means that some Norwegian learners are in fact within the native speaker range.
In summary, the total numbers at the group level indicate that Norwegian learners use fewer hedges than their native English counterparts. However, the total number per speaker and the total number of uses per word indicate that this statement needs to be modified. As illustrated in Figure 1, some native speakers use the same number of hedges as the Norwegians learners, although the majority tend to use more. Also, when we look at each individual expression, many native speakers use a similar number of hedges as the Norwegian learners.

4.2 Do Norwegian advanced learners use a more limited range of expressions (types) than native English speakers?

At the group level, both the native speakers and the learners used all 10 expressions. Therefore, we cannot claim that the learners exhibit a more restricted range than the native speakers. This contrasts with previous research on learners with different L1s. For example, Kärkkäinen’s (1990) study of Finnish learners showed that the learners exhibited limited variation when speaking English. Similarly, Friginal et al. (2017) found that language learners in a native environment used fewer hedging expressions than their teachers. However, the current study investigates only 10 expressions, so we must be careful when generalising from these results. If the full range of possible hedging expressions had been examined, the learners might have turned out to have a smaller repertoire.

Just like with instances of hedging, it is possible to look at the range of expressions at an individual level. Figure 4 shows the total number of hedging strategies (types) per speaker in LINDSEI-no and LOCNEC.
On average, the native speakers used approximately eight of the 10 expressions; the Norwegian learners used approximately five. However, as the boxplots in Figure 4 show, there is some overlap between the native speaker range and the learner range. Norwegian learners in the outskirts of the third quartile overlap with native speakers in the first quartile. If the whiskers are taken into account, the native speaker plot covers an even larger part of the learner plot. This raises the question of whether a broader range of the Norwegian learners could be described as native-like, if native-like is understood as ‘what (any) native speakers do’.

Figure 5 compares the hedging profiles of a learner and a native speaker, drawn from the upper part of the third quartile of the learner data and the lower part of the first quartile of the native speaker data.
Figure 5 is one example of similar native speaker and learner profiles. There were, however, several similar speaker profiles in the data. Figure 6 is another example of native and learner profiles that correspond to some degree.

Figure 4 also shows that there is great intra-speaker variation in both groups. The importance of intra-speaker variety has also been underlined by Gablasova et al. (2017). Neither learners nor native speakers are homogenous groups. Learners are often characterised according to external factors, such as numbers of years of study, but such factors do not necessarily say anything about (variation in) their
proficiency level. Furthermore, a low number of hedging strategies or a limited range of types does not necessarily indicate lack of mastery or activation; it could also be a matter of choice (Gablasova et al., 2017, 149). Some speakers in the native English material deviate from the majority, as illustrated in Figure 4. Although statistically, speakers such as the one indicated by the dot in Figure 4 are outliers, their use of hedging strategies will not necessarily be perceived as less native-like. Figure 7 illustrates the profiles of a native speaker outside the interquartile range and a learner with the same pattern. This example raises the question of whether learners should be perceived as non-native-like because they do not use all the hedging expressions or the same number of hedging expressions as the majority of native speakers.

Figure 10 Third example of a learner and native speaker profile.

Section 4.2 shows that the Norwegian learners as a group used all 10 expressions and thus on the surface did not seem to use a more limited range of expressions. However, an overview of the total number of different types per speaker showed that Norwegian learners on average used fewer types of hedges than native English speakers. Still, the performance of the Norwegian learners overlapped with several of the native speakers and thus could perhaps be described as native-like.

5. Concluding remarks

This paper aimed to discover whether there were any differences in the use of 10 hedging expressions by Norwegian advanced learners of English and native English speakers. Previous research has shown that learners tend to use both fewer and a more limited range of strategies. This has been proven for
Norwegian learners at the lower secondary school level, but little was known about hedging in the spoken English of advanced learners.

The results of this study indicate that although Norwegian advanced learners used fewer hedging strategies overall, this result was nuanced when the individual expressions and the individual speakers were considered. Only five of the 10 expressions were used significantly more by native speakers, and 16% of the Norwegian learners were within the 75th percentile of the English speakers when it came to the total number of hedging strategies used per speaker. The results also showed a partial overlap of the number of types of strategies used, further nuancing the results. As a group, Norwegian advanced learners used fewer hedges (tokens) and hedging strategies (types), but individual variation within both language varieties complicates the picture.

These results contribute to our understanding of Norwegian advanced learners’ pragmatic competence; however, they also raise some questions and red flags for future research. First, the study underlines the importance of looking at individual variation in the study of over- and underuse. Drawing conclusions based on group-level data alone may give an incomplete picture of intervarietal differences. Second, it raises the question of what native-like actually is. If the use of hedging strategies by a native speaker deviates from the majority of native speakers, will the speaker be perceived as non-native-like?
References


