

Exploring the Use of Probes in a Corpus Pragmatic Study of Hedging Strategies

Stine Hulleberg Johansen, University of Oslo

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.

Keywords: corpus pragmatics; form₁-to-function-to-form₂; hedging; probes

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-

Hulleberg Johansen, Stine. 2019. “Exploring the Use of Probes in a Corpus Pragmatic Study of Hedging Strategies.” *Nordic Journal of English Studies* 18(1):121-148.

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, Fraser 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)¹ (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.

Table 1 Form-to-function and function-to-form matrix

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

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

¹ 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.

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 (2006). 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. The use of 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 in that it searches for forms that are not part of the hedging strategies studied, 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	A: [...] no now it has been a good period [pragmatic particle]
B: * ja det har _uninterpretable_ nå	B: * yes it has _uninterpretable_ now
B: men denne steinen som har drevet og rasa ned uti e Nautvika _uninterpretable_	B: but this rock which have been collapsing down in e Nautvika _uninterpretable_
A: * ja	A: * yes
A: men det er nå <i>litt e</i> # overdrevet <i>trur jeg</i> [...]	A: but that is <i>a bit e</i> # overrated <i>I think</i>

NDC voldda_04gk<>who_avfile voldda_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—Oslo delen,”) 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 Övdalian 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

	BNC2014	NDC	NoTa	BB
Age distribution	9-91	15-88	15-85	22-36
Gender distribution	61 % female 39 % male	57 % female 43 % male	59 % female 41 % male	42 % female 58 % male
Years of recording	2012-2016	2006-2012	2005-2006	2001
Relationship between the interlocutors	71 % close family and friends 24 % friends and wider family 13 % acquaintances 3 % colleagues 1 % strangers	9 % family 26 % friends 59 % acquaintances 6 % strangers	12 % family 64 % friends 19 % acquaintances 5 % strangers	The participants are strangers who get to know each other and establish various kinds of relationships throughout the recording period

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 interface² whereas the English data was extracted through the CQPWeb interface³. 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.

Table 3 Raw frequencies of clauses with and without hedging in Norwegian and English

	<i>men</i>	random Norwegian	<i>but</i>	random English
with hedging	85	66	73	67
without hedging	65	84	77	83
sum	150	150	150	150

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

² <http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/glossa/index.html>

³ <https://cqpweb.lancs.ac.uk/usr/index.php?thisQ=login&uT=y>

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 give a better understanding of the dialogue.

(4)⁴

A: [...] # det lønner seg egentlig å være litt streng på IT-en for da # lærer du deg du har jo uttelling for det # nå	A: [...] # it pays off to be a bit strict in IT-class because then # you learn you benefit from that # now
---	--

B: * ja

B: *yes

B: ja # men a- de var jo d- m de var jo litt strenge det var bare at # gikk ikke så bra liksom _latter_ sånn karaktermessig	B: yes # but a – it was [<i>pragmatic particle</i>] d - m they were [<i>pragmatic particle</i>] 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 or the objection as a whole. *Liksom* (*like*) reduces the commitment of the speaker to the proposition, whereas *sånn* (*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.

⁴ 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.

(5)⁵

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?

A: that he is not moody?

B: mm

B: mm

C: ja at han ikke er *sånn e* sær # som han pleier å væreC: yes that he is not *like e* weird # like he usually isD: **sånn* grinete og sær

D: *like cranky and weird

BB Anette<<who_avfile 72

Sånn (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 *sånn (like)* and the hesitation marker *e*.

⁵ The >> symbol indicates overlapping speech. The symbols used in the original transcription are kept in these examples.

(7)⁶

A: oh I know (...) I do need to finish it (.) I need to get into it

B: *I just think er* if you if you did give it *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 *I think*, which says something about the speaker's commitment to the truth of the proposition and *just* which has a downtoning effect. *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 *men/but* contained hedging more often than random clauses did. First the total of *men* and *but* co-occurring with hedging strategies were compared to the total of random Norwegian and English clauses with hedging.

Table 4 Total of *men* and *but* compared to the total random units

	<i>Men + but</i>	random (Norwegian + English)
with hedging	158	133
without hedging	142	167
sum	300	300

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.

⁶ The (...) symbol indicates a short pause.

However, the significance is marginal and the two languages need to be considered separately before anything can be concluded.

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.

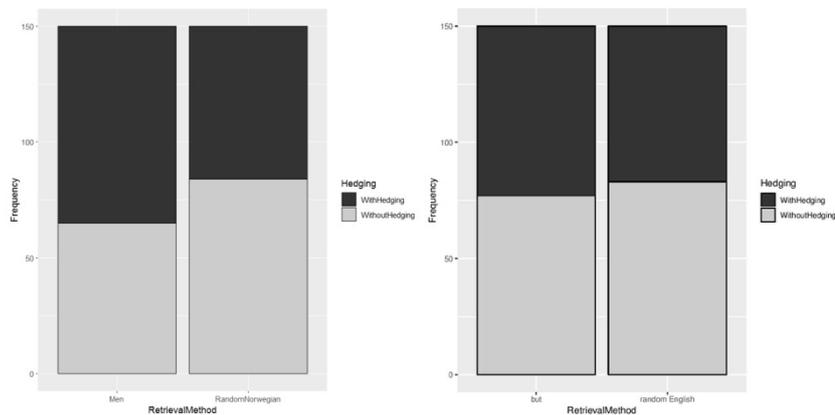


Figure 1 Frequencies of clauses with and without probes containing hedging in Norwegian and English

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 monolingual 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 ?

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

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

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

NDC karlsoey_02uk

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

A: saklig hele tiden ## det er greit å kunne lære seg det **men** når når *liksom* det er det eneste vi skriver om hver gang

A: to the point all the time ## it Is ok to have to learn it **but** when when *like* it is the only thing we write about every time

B: * ja * alt må *liksom* være så veldig ...

B: *yes* everything must *like* be so very ...

A: så er det de får ressursheftet så skal vi skrive om *et eller annet* som har med det å gjøre og det er # artikkel artikkel artikkel # hver gang *liksom* # det

A: so is it they get the support texts then we are supposed to write about *something or other* which has to do with that and it is # article article article # everytime *like* # it

B: mm * mm

B: mm * mm

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 “form₁-to-function-to-form₂” 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

- Adolphs, Svenja. 2008. *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam: John Benjamins.
- Aijmer, Karin. 1984. “‘Sort of’ and ‘Kind of’ in English Conversation.” *Studia Linguistica* 38: 118-28.
- and Christoph Rühlemann. 2015. *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press.
- and Anna-Brita Stenström. 2005. “Approaches to Spoken Interaction.” *Journal of Pragmatics* 37: 1743-51.
- Becher, Viktor. 2011. “When and why do Translators add Connectives? A Corpus-Based Study.” *Target* 23: 26-47.
- Beeching, Kate. 2016. *Pragmatic Markers in British English: Meaning in Social Interaction*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

- Blakemore, Diane. 1989. "Denial and Contrast: A Relevance Theoretic Analysis of *but*." *Linguistics and Philosophy* 12: 15-37.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Clemen, Gudrun. 1997. "The Concept of Hedging: Origins, Approaches and Definitions." In Raija Markkanen and Hartmut Schröder (eds.), *Hedging and Discourse Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Berlin: Walter de Gruyter. 235-48.
- Deutschmann, Mats. 2003. *Apologising in British English*. Umeå: Institutionen för Moderna Språk, Umeå University.
- Faarlund, Jan Terje, Svein Lie, and Kjell Ivar Vannebo. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Farr, Fiona, and Anne O'Keeffe. 2002. "Would as a Hedging Device in an Irish Context: An Intra-varietal Comparison of Institutionalised Spoken Interaction." In Randi Reppen, Susan M. Fitzmaurice and Douglas Biber (eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins. 25-48.
- Flöck, Ilka, and Ronald Geluykens. 2015. "Speech Acts in Corpus Pragmatics: A Quantitative Contrastive Study of Directives in Spontaneous and Elicited Discourse." In Jesús Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015: Current Approaches to Discourse and Translation Studies*. Cham: Springer International. 7-37.
- Fraser, Bruce. 1975. "Hedged Performatives." In Peter Cole and Jerry L. Morgan (eds.), *Syntax and Semantics*. New York: Academic Press. 187-210.
- . 1999. "What are Discourse Markers?" *Journal of Pragmatics* 31: 931-52.
- . 2010. "Pragmatic Competence: The Case of Hedging." In Gunther Kaltenböck, Wiltrud Mihatsch and Stefan Schneider (eds.), *New Approaches to Hedging*. UK: Emerald Bingley. 15-34.
- . 2013. "Combinations of Contrastive Discourse Markers in English." *International Review of Pragmatics* 5: 318-40.
- Goffman, Erving. 1955. "On Face-Work: An Analysis of Ritual Elements in Social Interaction", *Psychiatry* 18: 213-31.
- 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 Tyrkkö (eds.), *Towards Multimedia in Corpus Linguistics*. Helsinki: University of Helsinki. http://www.helsinki.fi/varieng/journal/volumes/02/gries_david/
- Halliday, Michael A. K., and Christian M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar* London: Arnold.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hübler, Axel. 1983. *Understatements and Hedges in English*. Amsterdam/Philadelphia: Benjamins.
- Johannessen, Janne B., Joel Priestley, Kristin Hagen, Tor A. Åfarli, and Øystein A. Vangsnes. 2009. "The Nordic Dialect Corpus – an Advanced Research Tool". In Kristina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. <http://tekstlab.uio.no/nota/scandiasyn/>
- Jucker, Andreas H. 2009. "Speech Act Research between Armchair, Field and Laboratory: The Case of Compliments." *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies* 41: 1611-35.
- Jucker, Andreas H., and Irma Taavitsainen. 2014. "Complimenting in the History of American English: A Metacommunicative Expression Analysis." In Irma Taavitsainen, Andreas H. Jucker and Jukka Tuominen (eds.), *Diachronic Corpus Pragmatics*. Amsterdam/Philadelphia: John Benjamins. 257-76.
- Kaltenböck, Gunther, Wiltrud Mihatsch, and Stefan Schneider. 2010. *New Approaches to Hedging*. UK: Emerald Bingley.
- Lakoff, George. 1972. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts." In Paul M. Peranteau, Judith N. Levi and Gloria C. Phares (eds.), *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society. 183-228.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina and Tony McEnergy. 2017. "The Spoken BNC2014". *International Journal of Corpus Linguistics*, 22(3), 319-44. doi:10.1075/ijcl.22.3.02lov <https://cqpweb.lancs.ac.uk/>
- McCarthy, Michael J. and Anne O'Keeffe. 2003. "'What's in a Name?' – Vocatives in Casual Conversations and Radio Phone in Calls." In Pepi Leistyna and Charles F. Meyer (eds.), *Corpus Analysis:*

- Language Structure and Language Use*. Amsterdam: Rodopi. 153-185.
- McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice* Cambridge: Cambridge University Press.
- Mulder, Jean, and Sandra A. Thompson. 2006. "The Grammaticization of *but* as a Final Particle in English Conversation." In Keith Allan (ed.), *Selected Papers from the 2005 Conference of the Australian Linguistic Society*. <http://www.als.asn.au>.
- NAOB – Det Norske Akademis ordbok. Det Norske Akademi for Språk og Litteratur. Kunnskapsforlaget <https://www.naob.no/>.
- Nikula, Tarja. 1997. "Interlanguage View on Hedging." In Raija Markkanen and Hartmut Schröder (eds.), *Hedging and Discourse Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Berlin / New York: Walter de Gruyter. 188-207.
- O’Keeffe, Anne. 2018. "Corpus-based Function-to-form Approaches." In Andreas. H. Jucker, Klaus. P. Schneider and Wolfram. Bublitz (eds.), *Methods in Pragmatics*. Berlin: De Gruyter Mouton. 587-618.
- Prince, Ellen, Joel Frader and Charles Bosk. 1982. "On Hedging in Physician-Physician Discourse." In Robert J. di Pietro (ed.), *Linguistics and the Professions*. Norwood/New Jersey: Ablex. 83-97.
- Prokofieva, Anna and Julia Hirschberg. 2014. "Hedging and Speaker Commitment." New York: University of Colombia.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rudolph, Elisabeth. 1996. *Contrast: Adversative and Concessive Relations and their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Berlin: de Gruyter.
- Salager-Meyer, Françoise. 1994. "Hedges and Textual Communicative Function in Medical English Written Discourse." *English for Specific Purposes* 13: 149-70.
- Sanders, Robert E. 2015. "Discourse Strategy." In Karen Tracy, Todd Sandel and Cornelia Ilie (eds.), *The International Encyclopedia of Language and Social Interaction*.
- Schauer, Gila A. and Svenja Adolphs. 2006. "Expressions of Gratitude in Corpus and DCT Data: Vocabulary, Formulaic Sequences, and Pedagogy." *System* 34: 119-34.

- Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Stenström, Anna-Brita. 1994. *An Introduction to Spoken Interaction*. London: Longman.
- Taavitsainen, Irma, and Andreas H. Jucker. 2015. "Twenty Years of Historical Pragmatics: Origins, Developments and Changing Thought Styles." *Journal of Historical Pragmatics* 16: 1-24.
- Tagliamonte, Sali and Rachel Hudson. 1999. "Be like et al. beyond America: The Quotative System in British and Canadian Youth." *Journal of Sociolinguistics* 3: 147-72.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.

Corpora

- BigBrother-korpuset. <http://www.tekstlab.uio.no/nota/bigbrother/>
- The English-Norwegian Parallel Corpus.
<http://www.hf.uio.no/ilos/english/services/omc/enpc/>
- Norsk talespråkskorpus - Oslodelen.
<http://www.tekstlab.uio.no/nota/oslo/index.html>