1. Introduction

Studying texts is something which philologists as well as corpus linguists have done, do and will do. The question to be tackled in this paper is how corpus linguists work with texts that are compiled in corpora. In my understanding, both the philologist as well as the corpus linguist look for evidence in texts. The philologist has worked with texts for many centuries, whereas the corpus linguist (as he/she is understood today) only emerged in the 1960s with the advent of texts being available in digital form.

A definition of what philology is and how a philologist works is not easy to give. Crystal defines philology as a "traditional term for the study of LANGUAGE history, as carried on by 'COMPARATIVE philologists' since the late 18th century" (Crystal 1980, 264). Within a German tradition of philology, important results have been recorded in the 18th and 19th century with the comparison of languages by trying to develop a "Stammbaum" and additionally with phonological investigations leading to the description of sound laws and sound changes. Gneuss (1990, 55) mentions that the methods and principles employed by historical (comparative) philology had a major influence on the works focusing on contemporary English (1890-1935). Among others, he lists the following characteristics:

- "streng deskriptiver Charakter: der Grammatiker registriert den Sprachgebrauch, gibt keine Regeln, erläßt keine Verbote;" [a strictly descriptive character: the grammarians notices language use, does not give any rules, does not proscribe anything; my translation]
- "Streben nach Exaktheit;" [striving for precision/accuracy; my translation]
- "Streben nach Vollständigkeit;" [striving for being comprehensive; my translation] (1990, 55)

These three characteristics are also guiding principles within corpus linguistic research. With the advent of computer-readable texts, other possibilities of working with texts came into existence. Researchers could now easily calculate frequencies of linguistic items, consider their co-occurrence automatically and search and retrieve any unit they wanted to focus on. It was no longer necessary to scan whole texts for units that a researcher wanted to study. The software did this for the researcher. This meant that researchers did not necessarily need to have a profound knowledge of the texts which were in the corpora they used. This situation led Rissanen to describe "[t]he philologist's dilemma," which relates to the use of text corpora on the one hand and "getting acquainted with original texts" (Rissanen 1989, 16) on the other. The main concern voiced by Rissanen is that a meta-analysis of quantitative data will not take into account what is really in the text:

---


---


Anglistik, Volume 28 (2017), Issue 1
© 2017 Universitätsverlag WINTER GmbH Heidelberg
[T]here is a risk that corpus work and computer-supported quantitative research methods will discourage the student from getting acquainted with original texts, from being on really intimate terms with his material and thus acquiring a profound knowledge of the language form he is studying. (Rissanen 1989, 16)

Following Rissanen's argument, texts should be read. In other words: is a study fruitful which focuses on quantitative data only and as such gets away from the text? On the other hand, is a close analysis of a particular data sample also helpful?

In what follows, I will sketch two case studies. At the heart of both is that a study of the context is relevant to arrive at a detailed linguistic account. By presenting the case studies I would like to demonstrate possible ways of analyzing language samples by employing corpus-linguistic means that might be described as being rooted in a philological analysis, i.e. in a detailed consideration of the text surrounding the item under investigation (see Möhlig-Falke, this volume).

2. Case Study I

In the first case study I would like to demonstrate that a qualitative analysis of the data is necessary to discover functional differences in what from a (superficial) formal point of view is identical. The main aim will be to outline how information on the context is relevant in determining the function of a particular unit. Once a functional difference has been identified, a quantitative study of the data offers additional information.

The focus will be on the pattern "VERB TO VERB." Typical examples are given in (1) and (2).

(1) We knock at two doors and ask to come in (COCA fic 2007)
(2) They never seem to mind. (COCA fic 1990)

Two different linguistic explanations are given for the patterns found in (1) and (2), which are depicted in Figure 1 below. In (1), the main verb ask is followed by a to-infinitive clause functioning as the direct object of ask (see Quirk et al. 1985, 1187). In example (2), the verb seem is followed by to and the verb mind. The verb seem is not considered a main verb as ask is in (1). Instead, Quirk et al. (1985, 146) label SEEM to as a catenative verb. Huddleston and Pullum (2002, 1176ff.) analyze it as catenative which is followed by a non-finite clause as its catenative complement. Biber et al. (1999, 693ff) refer to it as a verb controlling a to-clause. Biber et al. mention that seem expresses probability. This brief overview makes two things clear:

1. There is no agreement on the terminology and thus the function of the verb before to (ask and seem in examples (1) and (2)) and the verb following to (come in and mind).

2. Quirk et al. and Huddleston and Pullum consider instances of ASK to V in (1) as different from examples such as SEEM to V in (2). Biber et al. mention them under the purely formal label "verbs controlling infinitive clauses in post-predicate position" (1999, 700). The difference advocated by Quirk et al. as well as by Huddleston and Pullum goes hand in hand with a different grammatical analysis, which is graphically

---

2 The reference to the source is given in its short form. The example is taken from the section on fiction (fic) from 2007 from the Corpus of Contemporary American English (COCA).

3 Verbs given in capitals refer to all forms of a verb. SEEM thus refers to seem, seems, seemed, and seeming.
realized in Figure 1. The verb phrases found in example (1) are depicted on the left side of Figure 1. The finite verb phrase *ask* is followed by a non-finite verb phrase *to come in*. The sentence has two clauses, each with its own verb phrase. The *to*-infinitive clause functions as the direct object of the matrix clause main verb *ask*. The verb phrase found in example (2), on the other hand, is depicted on the right side of Figure 1. It consists of just one verb phrase. The catenative verb *SEEM to* is followed by a non-finite verb *mind* realized as a bare infinitive.

**Figure 1:** Graphic realization of grammatical analysis of the verb phrases in (1) and (2)

The verb system of the English language consists of lexical verbs on the one hand and function verbs on the other. The function verbs are modal auxiliaries, the primary verbs *BE* and *HAVE* as well as the operator *DO*. Verbs such as *SEEM to* are not described thoroughly in the verb system. D. Mindt (2000) has analyzed the English verb system and has identified an additional category: catenative verbs. A catenative verb "introduces an additional meaning component into the verb phrase" (D. Mindt, 2000, 128). Catenative verbs are for example *SEEM to* and *LIKE to* – verbs that seem to be difficult to categorize for Quirk et al., Biber et al., and Huddleston and Pullum.

Taking for granted that catenative verbs constitute an additional element in the English verb phrase system, the question is how one can identify them. In a case study, where I will focus on *COME to see* only, I would like to describe how such an analysis may proceed. I will look at *COME to see* from a Modern English point of view and additionally consider its development from Old English to Modern English.

Six corpora were used for this study as outlined in Table 1: The Helsinki Corpus, the Corpus of Early English Correspondence – sampler version (CEECs), the Lampeter Corpus, the Old Bailey Corpus (OBC), the Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA). The search string was *COME to see*, i.e. any form of the verb *come* (*come, comes, came, ...*)
coming)\(^1\) followed by to and see. Intervening elements between COME, to and see such as punctuation marks or auxiliary verbs were also considered.\(^5\)

<table>
<thead>
<tr>
<th>Period</th>
<th>Corpus</th>
<th>Time span covered</th>
<th>British English/American English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old English (700-1150)</td>
<td>Helsinki</td>
<td>850-1150</td>
<td>BrE</td>
</tr>
<tr>
<td>Middle English (1150-1500)</td>
<td>Helsinki</td>
<td>1150-1500</td>
<td>BrE</td>
</tr>
<tr>
<td>Early Modern English (1500-1700)</td>
<td>Helsinki</td>
<td>1500-1710</td>
<td>BrE</td>
</tr>
<tr>
<td></td>
<td>Lampeter</td>
<td>1640-1740</td>
<td>BrE</td>
</tr>
<tr>
<td>Late Modern English (1700-1900)</td>
<td>Old Bailey</td>
<td>1720-1899</td>
<td>BrE</td>
</tr>
<tr>
<td></td>
<td>COHA</td>
<td>1810-1900</td>
<td>AmE</td>
</tr>
<tr>
<td>Modern English (1900-)</td>
<td>COHA</td>
<td>1900-2000</td>
<td>AmE</td>
</tr>
<tr>
<td></td>
<td>COCA</td>
<td>1990-</td>
<td>AmE</td>
</tr>
</tbody>
</table>

Table 1: Corpora consulted in this study

The corpora used for this study differ in size, register and variety. In order to have available as many cases as possible of COME to see, I have decided to use a wide range of corpora. American English is overrepresented in Modern English as is British English – naturally – for older stages of the English language. This might lead to a bias in the Modern English period towards American English. Since this study concentrates on COME to see which up to the present point is not known to represent a difference as to the varieties of the English language, the imbalance between British and American English might not have any effect on the findings and explanations outlined below.

<table>
<thead>
<tr>
<th>Period</th>
<th>absolute number of occurrences</th>
<th>occurrences per million words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old English (700-1150)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Middle English (1150-1500)</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Early Modern English (1500-1700)</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Late Modern English (1700-1900)</td>
<td>1,964</td>
<td>8</td>
</tr>
<tr>
<td>Modern English (1900-)</td>
<td>5,018</td>
<td>8</td>
</tr>
<tr>
<td>TOTAL</td>
<td>7,001</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Absolute frequencies of COME to see

The absolute frequencies of COME to see are outlined in Table 2. Three occurrences are found in Middle English. A massive increase in absolute numbers is observable from Early Modern English to Late Modern English and again from Late Modern English to Modern English. This increase is due to the fact that the corpora investigated for Late Modern English and Modern English contain more words than those for earlier periods of the English language. If the frequencies occurring per million words are compared, a slight increase from five to eight becomes visible be-

\(^4\) Spelling variants such as cumen, cume, komen, commen etc. for example are also included in the search. The same is true for the verb SEE, where spelling variants have also been considered. The source for the spelling variants is information given in the Oxford English Dictionary (OED) for the respective entries.

\(^5\) One example where quotation marks occur within COME to see is the following: The young woman going to Burlington to spend the week-end was surrounded with about fifteen other young women who had come to “see her off” (COHA fic 1915).
tween Middle English and Early Modern English. With regard to its frequency, COME to see does not increase or decrease in any significant way in the time period under observation. If a quantitative analysis does not reflect a trend, a qualitative analysis might do so. (3) and (4) are two examples from COCA.

(3) She came to see me again (COCA mag 2011)
(4) She came to see her panic attacks in exclusively psychoanalytic terms (COCA acad 1999)

As can be seen in Table 3, they both start with She, which is followed by came to see. Example (3) then has a pronoun as a noun phrase, whereas (4) has a noun phrase. Both are followed by an additional element. From a structural point of view there does not seem to be a difference between (3) and (4).

<table>
<thead>
<tr>
<th>Pronoun as NP</th>
<th>Verb CAME</th>
<th>to see</th>
<th>NP</th>
<th>additional element</th>
</tr>
</thead>
<tbody>
<tr>
<td>She</td>
<td>came</td>
<td>to see</td>
<td>me</td>
<td>again</td>
</tr>
<tr>
<td>She</td>
<td>came</td>
<td>to see</td>
<td>her panic attacks</td>
<td>in exclusively psychoanalytic terms</td>
</tr>
</tbody>
</table>

Table 3: Syntactic pattern of examples (3) and (4)

However, when the two examples are analyzed in more detail, a difference in meaning is detected. In example (3) the subject (she) came in order to see another person again. Thus, a movement to a particular place is expressed by came. To see refers to the meaning of visiting and also includes the more basic meaning of SEE as visually looking at someone. In (4), however, neither a movement towards a particular place nor a visual act is displayed. The unit came to see refers to the end of a process where a person, the subject she, noticed or realized something. Thus, a metaphorical movement in terms of a thought process is depicted which is strongly connected with the verb see. In cases such as (4), COME together with to see displays the meaning of realizing something. COME to see expresses the end of the development of a mental process. Biber et al. state that "come occurs in this construction only with a specialized sense related to come about or happen" (1999, 708).

Despite the fact that examples (3) and (4) have the same form, there is no clue that came to see expresses two totally different meanings. These two different meanings can only be detected by a close manual inspection of all 7,001 cases. The identification of these two meanings has indeed been the result of such a manual analysis. The meaning found in example (3) will be referred to as the purposive meaning. To is considered as an element that conveys both the function of a conjunction with the meaning 'in order to' and at the same time works as the to-infinitive marker. COME and see in example (3) are main verbs. Example (4) exemplifies a case of a catenative meaning, because the verb phrase came to see must be considered as a single finite verb phrase. COME is a catenative verb in example (4), which is followed by the to-infinitive marker and the main verb see.

By going even further back in time and considering cases from the Helsinki corpus, the CEEC-sampler, from the LAMPETER corpus and the Old Bailey Corpus, more cases can be added. Table 4 gives the absolute and the relative frequencies for each period in relation to the purposive and the catenative meaning.
Table 4: Absolute and relative frequencies of COME to see for periods of the English language

<table>
<thead>
<tr>
<th>Period</th>
<th>Purposive meaning</th>
<th>Catenative meaning</th>
<th>(row) TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>relative</td>
<td>absolute</td>
</tr>
<tr>
<td></td>
<td>frequency</td>
<td>frequency</td>
<td>frequency</td>
</tr>
<tr>
<td>Middle English</td>
<td>1100-1499</td>
<td>2</td>
<td>67%</td>
</tr>
<tr>
<td>Early Modern</td>
<td>1500-1699</td>
<td>14</td>
<td>88%</td>
</tr>
<tr>
<td>Late Modern</td>
<td>1700-1899</td>
<td>1,841</td>
<td>94%</td>
</tr>
<tr>
<td>Modern English</td>
<td>1900-</td>
<td>4,015</td>
<td>80%</td>
</tr>
</tbody>
</table>

The absolute frequencies show that COME to see does not occur at all in Old English. One instance is found in the Old English data where a form of COME is followed by a form of SEE. However, this neither represents the purposive nor the catenative meaning.

(5) þonne se mone bið anre nihte eald, swahwætswa þu gesiht, þate *kymð to gesean.*

(Helsinki OX/4 IS HANDA PROCC)

(5) is an example from an astronomical text taken from "Prognostications." In (5), the verb *kymð* is followed by *to gesean.* There is a debate as to whether *gesean* or *gefean* is the correct form in the manuscripts (Liuzza 2010, 174). In the entry for "gefean," the Dictionary of Old English (DOE) lists the spelling "gesean" standing for "gefean" in the text "Prognostics," where (5) is also taken from. The source in the DOE is the same as in the Helsinki corpus. Therefore, it is very likely that "gesean" is the noun with the meaning "joy, gladness" and thus does not represent an instance of the Old English verb *seon.*

Three cases are found in Middle English. Two represent the purposive meaning exemplified in (6) and one case has been observed for the catenative meaning given in (7). Note that *to be seene* in (7) is in the passive voice. A passive is only found in connection with the catenative meaning (175 cases out of 1,129 accounting for 16%).

(6) Moyses, er he migt *come to se* þis arke, & for to wite how it schuld be maad […] (Helsinki M3 IR RELT CLOUD)

(7) Sir George did promise securitye out of the land; and, when the deed *came to be seene,* yt is sayde that land was to passe in Sir George Bowes' sonne's name. Soe all is wronge there. (CEECS1)

Figure 2 shows the relative frequencies for each period of the English language in graphic form. All cases of each period are considered and are given as relative percentages for each single period to indicate the development over time. When only considering these relative frequencies, one interpretation could be that the catenative meaning has decreased from Middle English to Late Modern English and has slightly increased towards Modern English. However, the plot in Figure 2 must be related to the absolute frequencies. A total number of cases of three for Middle English and 16 for Early Modern English are not enough to allow for any clear conclusion with relation to their development. The last two periods under observation here indicate that
not only COME to see is used increasingly but that the distribution between the purposive and the catenative meaning changes slightly. Whereas in Late Modern English, 94% of cases were used in the purposive function, this figure drops to 80% in Modern English.

Figure 2: Relative frequencies as total of all occurrences in each period

From 1700 onwards a massive increase in the absolute frequency of COME to see is observable. Whether the dip in the relative frequency from Middle English to Early Modern English and Late Modern English is significant cannot be said based on the data analyzed here. What is clear is that the relative frequency of the catenative meaning rises slightly towards the Modern English period. With 80% of all cases, the purposive meaning is the most frequent one in Modern English as it is in all periods of the English language.

The manual inspection of the context has resulted in identifying two different meanings which can be related to two different grammatical structures. Neither the meanings nor the structures can be detected on quantitative grounds by looking at their frequency. Only a detailed, manual analysis makes it possible to detect the two meanings and analyze their distribution over time.

The two different meanings must be explained differently from a grammatical point of view as can be seen in Figure 3. The meaning found in example (3) represents the purposive meaning. COME, on the other hand, is a catenative verb in example (4), which is followed by the to-infinitive marker and the main verb see. The analysis in terms of the catenative meaning is identical to Analysis 2 presented in Figure 1.

Whereas SEEM to as in (2) can almost safely be considered to be a catenative verb in all of its occurrences, this is definitely not true for COME to see. COME to see occurs in about 80% of its cases in Modern English with the purposive meaning, where both COME and SEE are main verbs. However, the grammatical analysis of COME to see representing the purposive meaning does not follow Analysis 1 given in Figure 1. A different analysis for cases without a catenative verb must be given, because to does not only represent the to-infinitive marker but additionally the conjunction meaning “in order to.” Therefore, to see must be considered as introducing a non-finite adverbial clause as represented in Figure 3. 20% of the cases of COME to see in Modern English can be explained as conveying the catenative meaning. They can
safely be analyzed as comprising the catenative verb COME to which is followed by the main verb see. See can occur both in the active as well as in the passive voice. In my data, cases of the passive voice have only been found for the catenative meaning. Cases representing the catenative meaning must be analyzed as consisting of one verb phrase as can be seen in Figure 3.

Figure 3: Grammatical analysis of COME to V

This time-consuming analysis makes it possible to investigate one single characteristic of the English verb system in great detail. The results have an impact on the description of grammatical theory (see also Lange, this volume). The disagreement over the terms 'catenative verb,' 'catenative' and 'catenative complement' or verbs controlling a to-clause is a direct reflection of ongoing processes in the English verb system, where a clear-cut description at this point in time simply does not seem possible. Whereas WANT TO V and SEEM TO V are clear cases of catenative verbs, a language user has the option of employing COME as a catentative verb or as a main verb. In how far this distinction is possible for verbs other than COME to and whether there is a preference for certain verbs to occur in the catenative rather than the purposive meaning is a question for further research (see I. Mindt (forthcoming) for more details).

3. Case Study II

The second case study focuses on the unit I think and on linguistic factors that have an influence on the categorization of I think as a discourse marker. I would like to demonstrate some problems that occur when working with spoken data. The corpus which has been used in this case study is the International Corpus of English – the British component (ICE-GB). I will compare the results of two studies that have analyzed I think based on the same data set. The two studies differ in their application of criteria for determining the status of I think. My own study (I. Mindt, 2003) was based on written

---

6 The fact that one and the same verb can be used in two different functions is clearly a fascinating phenomenon of the English language.
transcriptions of the spoken part of ICE-GB. The main aim of the study was to find out under which linguistic circumstances *I think* can be considered a discourse marker. The other study has been conducted by Dehé and Wichmann (2010). They focus on sentence-initial *I think* (and on *I believe*) and adopt a prosodic approach in determining when *I think* is a discourse marker. They also use spoken data from ICE-GB.

Mindt (2003) based her definition of discourse marker on criteria discussed by Brinton (1996, 30-40) as well as on Jucker and Ziv (1998, 1-5). For her, discourse markers are said

\[\text{a) to have little or no propositional meaning on the semantic level,}\]
\[\text{b) to be either outside the syntactic structure or only loosely attached to it on}\]
\[\text{the syntactic level,}\]
\[\text{c) to occur in initial position (additionally also in medial or final position) of}\]
\[\text{the sentence, and}\]
\[\text{d) to convey an interpersonal and a textual function on the pragmatic level.}\]

At the point when the research was conducted, the spoken data files were not available. Therefore, criteria related to phonetics and phonology could not be taken into account. All instances of *I think* in written transcriptions of the spoken part of ICE-GB have been analyzed according to the criteria mentioned in a)-d) above. The result was that *I think* can only be considered a discourse marker if

\[\text{i. it is associated with a bleached meaning component,}\]
\[\text{ii. it is syntactically detached,}\]
\[\text{iii. it occurs in front position of the unit and}\]
\[\text{iv. it conveys a textual function only.}\]

Of all 445 cases analyzed, only 25.8% could be considered as discourse markers. One example of *I think* as a discourse marker is (8), where speaker A asks a question and speaker B tries to provide an answer which is introduced by the discourse marker *I think*.

\[(8)\] A: So what did you see as missing from other types of activities for the disabled which might have influenced you into starting this group

B: *I think* the main perception that I had was that the contributions from able-bodied people and disabled people were often unbalanced

*I think* in (8) does not carry the meaning of opinion or belief. Therefore, it is semantically bleached. A syntactically incomplete unit follows *I think* which led to analyzing *I think* as being syntactically detached. *I think* is found in front position and it conveys a textual rather than an interpersonal function because it is used as a textual device to indicate the start of the turn of speaker B.

Dehé and Wichmann (2010) take a prosodic approach. They differentiate between three functions of *I think*:

1. a main clause,
2. a comment clause and
3. a discourse marker.
They argue that the function of *I think* is reflected in the prosody, more precisely in the prosodic prominence. When *I think* occurs in a main clause, the accent placement is on the pronoun, when *I think* is considered as a comment clause, the accent placement is on the verb. When a speaker does not assign any stress to *I think*, it is considered to be a discourse marker (see Dehé and Wichmann 2010, 36). The two authors used PRAAT and based their prosodic analysis on acoustic parameters. Dehé and Wichmann give a shorthand prosodic transcription for most of the cases they discuss by indicating the pitch placement and the pitch contour. Additionally, they indicated the bracketing of the tone units.

In the research presented by Dehé and Wichmann (2010), six cases of *I think* are discussed in greater detail. These six cases have also been analyzed by Mindt (2003). The following discussion will thus focus on the analysis of these six cases.

Dehé and Wichmann have categorized three of the six cases as main clauses, one as a comment clause and two as a discourse marker. Mindt did not use the three-fold categorization between main clause, comment clause and discourse marker. Her categorization is based on a cluster of syntactic, semantic, positional and pragmatic factors, which points to a gradient analysis of the various cases of *I think*. The two extreme poles of this gradience are main clauses on one end of the gradience and discourse markers on the other. In Mindt's analysis, three clear instances of main clauses are present, which do not coincide with the three cases of main clauses as categorized by Dehé and Wichmann. According to Mindt's criteria, none of the six examples is a discourse marker.

(9) But I think uh uh regardless of his influences *I think* that he is the most neglected of that uh number of composers around the turn of the century (ICE-GB: S1B-032#103)

*I think* in (9) is considered to be a main clause by Dehé and Wichmann. The *that*-clause following *I think* is analyzed as a direct object in syntactic terms. The main clause is superordinate to a subordinate clause. A speaker expresses his/her attitude or opinion (Dehé and Wichmann 2010, 42).

(9a) \[ *I think* \[ *that* he is the most neglected] \]

H*L H% H* H* H*L L%

(prosodic analysis taken from Dehé and Wichmann 2010, 48)

The prosodic analysis as given by Dehé and Wichmann is depicted in (9a). *I think* occurs in a complete tonal contour. This is indicated by the brackets ([…]) surrounding *I think*. *I think* is realized by a fall-rise for which the shorthand is H*L+H%. "The nuclear peak is associated with the pronoun *I*" (2010, 48). They observe a "pitch discontinuity between *think* and *that*, followed by a step-up in pitch on *that*" (48).

Mindt's (2003) argumentation is in line with the one given by Dehé and Wichmann. For Mindt, *I think* expresses an opinion or belief on the semantic level. *I think* is part of the syntactic structure and functions as a main clause and additionally fulfills an interpersonal function in that it is used as a supporting device for the meaning to be conveyed.

(10) *I think* that Lord Scarman was right when he said that our accountability for the police in this country was muddled and incomplete (ICE-GB: S1B-033#37:1:E)
(10a) [I think that Lord Scarman was right]

\[ \text{H*} \rightarrow \text{H*L.H%} \]

(prosodic analysis taken from Dehé and Wichmann 2010, 50)

*I think* in (10) is considered both by Dehé and Wichmann as well as by Mindt as a main clause. Dehé and Wichmann's argumentation is based on the fact that a prenuclear pitch accent is found on the pronoun *I*. For them, "the whole sequence *I think that Lord Scarman was right* forms one intonation domain" (Dehé and Wichmann 2010, 50). In Mindt's analysis, *I think* conveys an opinion and belief. *I think* has an interpersonal function and introduces a main clause.

(11) I think it's all jolly good fun (ICE-GB: S1B-024#28:1:D)

\[ \text{H*} \rightarrow \text{L%} \]

(prosodic analysis taken from Dehé and Wichmann 2010, 56)

Yet another example of *I think* as a main clause for Dehé and Wichmann is (11). This categorization is based on the prosodic prominence found on the pronoun *I*. Dehé and Wichmann argue that the speaker "places an accent on the pronoun *I* rather than the verb *think*, expressing her own attitude" or "Real speaker attitude" as Dehé and Wichmann (2010, 55) point out. Mindt does not agree with this analysis. *I think* has a textual function and serves as a device indicating the speaker's wish to continue her turn. Mindt agrees with Dehé and Wichmann that *I think* still serves – in syntactic terms – as a main clause despite its pragmatic function. From a semantic point of view, *I think* does express an opinion or belief on the side of the speaker. Although the semantic and syntactic functions of *I think* are identical to the ones in (9) and (10), the pragmatic function is different. Thus, *I think* in (11) is not a typical example of a main clause because it depicts a textual rather than an interpersonal function. The discourse surrounding (11) is given in (12).

(12) A: You speak here in in as a double capacity because you're not only legal man and Rumpole man but you're Henley man <,>
C: Well actually I I haven't got the uh the uh joy and entertainment of being one of Mr Heseltine's constituents because I live just across the corner in Buckinghamshire
A: So whatever you're just that little bit of an outsider Mortimer
C: I'm just a bit of that <,> <laughter>
A: You get away with it yet again <,>
A: Heather Couper what's your view of the uh of the upfront fighting if we can use that clumsy awkward
A: Don't forget it is Monday morning
D: <laughter> If I can partly misquote John and say yes indeed it adds to the gaiety of the country
D: That's not the appropriate word *at this time* of the morning but still
D: *I think* it's all jolly good fun
D: In fact I went along to the House of Commons to watch an adjournment debate on Wednesday night last week and uh I loved the way everybody postures and poses

(ICE-GB: S1B-024#19-29)

Three speakers – labelled A, C and D – interact. "<,>" indicates a small pause. The label <laughter> as given in ICE-GB is depicted in (12). However, when listening to the
sound files, more laughter can be detected which – unfortunately – is not indicated in the transcription. Speaker D laughs when A utters can use that clumsy. Laughter also occurs when at this time is uttered. Mindt's interpretation of I think in (12) as a marker of continuation is supported by the analysis of the sound file, because speaker D wants to keep her turn after the laughter has taken place on at this time. Thus, I think is employed as a continuity device. This interpretation makes clear that there are subtle differences when analyzing I think. For Mindt, the function of I think in (11)/(12) is not identical compared to the one in (9) and (10). In cases where I think is considered a comment clause or a discourse marker, the main proposition is expressed in the subclause. I think "reflects the speaker's stance towards the proposition or functions as an epistemic marker" (Dehé and Wichmann 2010, 42). The main governing principle in Dehé and Wichmann's prosodic analysis is that a "word's prosodic salience is related to its meaning" (2010, 43). A comment clause is observed to occur in (13).

(13) Uh when at least among uh Christians uhm uh modern contemporary Christians I think that the problem of faith very often presents itself as an individual problem (ICE-GB: S1B-028#104:1:C)

[I think that the problem of faith] [...] 

H*  !H*  !H*L$L%

(prosodic analysis taken from Dehé and Wichmann 2010, 52)

In (13) one intonation unit is reported to exist from I think until faith. The nuclear pitch accent is on faith. Two prenuclear accents are observed on think and pro in problem. The verb think has received an accent by the speaker. This is the main reason for Dehé and Wichmann to analyze I think in (13) as a comment clause. For Mindt, I think in (13) clearly displays an opinion or belief, is an example of the interpersonal function and thus is considered as a main clause.

I think constitutes a discourse marker in (14) for Dehé and Wichmann.

(14) I think that these democratic ideals still have to be achieved in Britain and I hope to show in this programme that this can't be achieved until Britain becomes a republic (<ICE-GB: S2B-023#8:1:A)

[I think that these democratic ideals] [...] 

H*  !H*L$L%

(prosodic analysis taken from Dehé and Wichmann 2010, 52)

Dehé and Wichmann argue that I think in (14) is unstressed. One intonation unit is present which starts with I think and ends in ideals. "The first (prenuclear) prominence falls on the demonstrative these following that" (2010, 52). Mindt does not agree with this analysis. For her, yet again, an opinion or belief is expressed. I think serves an interpersonal function in that certainty on the proposition of the speaker is articulated. This becomes clear by looking at the context given in (15).

(15) A: But for me being patriotic has nothing to do with our monarchy

A: I think that the idea of hereditary privilege especially hereditary office is totally incompatible with the democratic ideals that British people have fought and died for <,>

A: I think that these democratic ideals still have to be achieved in Britain and I hope to show in this programme that this can't be achieved until Britain becomes a republic <,>

Anglistik, Volume 28 (2017), Issue 1
© 2017 Universitätsverlag WINTER GmbH Heidelberg
A: These ideas have been spreading and taken for granted everywhere else in the developed world for hundreds of years so it's absolutely amazing that in Britain republicanism is something that's hidden in the closet <,>

(ICE-GB: S2B-032#6-9)

The sentence preceding the one where I think occurs also starts with I think. In Mindt's interpretation, the speaker expresses his personal opinion. This is made clear by the repetition of I think occurring in sentence-initial position. In the sentence where I think is found, a similar structure (I hope) is used after the coordinator and. These three structures (I think – I think – I hope) are all deliberately employed to introduce propositions. Thus, I think clearly expresses an opinion and ought to be interpreted as a main clause rather than as a discourse marker – despite the fact that no accent placement has been detected.

The last example to be discussed is (16).

(16) B: do you no notice the sort of clinking of the glasses in incredibly particularly in the hospitality boxes at the the major test match grounds

B: I think there are many women who probably would take it far more seriously than some of the men who are only there for the <,,> <unclear-word>

(ICE-GB: S1B-021#55-56)

I think in (16) is analyzed by Déhé and Wichmann as a discourse marker, because it is realized as an unstressed prehead. They argue that I think introduces a contrast in the argumentation. "It can be interpreted to function as a politeness marker" (Déhé and Wichmann 2010, 60). For Mindt, I think in (16) is not a discourse marker although it conveys a textual function. It serves as a marker of continuity in the discourse, meaning that the speaker has not yet finished her turn. In Mindt's opinion, the meaning of opinion and belief is still present in this example. This is one of the main reasons why it is not a discourse marker for Mindt. In the audio file a sigh on the side of the speaker is audible on who after women. This supports the analysis of Déhé and Wichmann of referring to a contrast in the argumentation. Whether this contrast is introduced by I think is debatable. Based on the sound file, Mindt argues that the contrast is brought across by the prosodic prominence on women and the sigh on who. Whereas for Déhé and Wichmann it is a discourse marker, it is not one for Mindt because the semantic meaning of I think is still observable and not bleached.

The comparison between the prosodic approach by Déhé and Wichmann and Mindt makes clear that there is not one single approach that offers a sound and convincing categorization of I think. Without doubt, an analysis that is only based on written transcriptions of spoken recordings as has been done by Mindt (due to the fact that the recordings were not available at the time the research was conducted) misses out on important details. On the other hand, a focus on prosodic prominence only faces the same problem in that other linguistic factors such as syntax, semantics and pragmatics are not taken into account. It has been argued by Kaltenböck in various studies on I think (e.g. 2008) and Kaltenböck et al. (2015) that a distinction between reported clause, comment clause and discourse marker is useful. Additionally, a wide range of linguistic factors need to be taken into account to arrive at a convincing analysis. However, neither prosody nor semantic and pragmatic means on their own can be satisfactorily employed to classify I think unambiguously. Especially with regard to prosody, a clear-cut and straightforward analysis seems to be difficult to
come by. More research on the uses and functions of various prosodic parameters such as accent placement, pitch, vowel quality etc. is needed.

The comparison of the study by Mindt and the one by Dehé and Wichmann has demonstrated that a clear categorization is not always possible. The categorization of I think is probably best seen as one on a continuum between extreme poles. Clearly, the context – be it prosodic, syntactic, semantic, pragmatic or textual – is of great relevance in determining the function and use of this unit.

4. Discussion

Two case studies have been reported. In the first case study I focused on COME to see. COME to see has two different meanings which can be related to two different grammatical functions. A quantitative analysis can (at this point in time) not be applied to distinguish between the two meanings. A close manual inspection of all cases was necessary to analyze all cases and describe the two different meanings. COME to see is thus an example of how speakers of the English language make use of its inventory and develop new meanings that can then be described in new grammatical ways by linguists. COME is either used as a main verb or as a catenative verb. Catenative verbs are part of the verb phrase and give language users the opportunity to add new meanings to the verb phrase and thus offer new ways of expressing themselves.

The detailed study on COME to see has been strictly descriptive. It accounted for all cases of COME to see which have been found in the corpora investigated. The study takes into account all the characteristics mentioned by Gneuss above (1990, 55). Additionally, quantitative data have been used to describe the diachronic development. I outlined that the purposive meaning in COME to see accounts for the majority of cases over time. The catenative meaning seems to be used more frequently in Modern English than in previous periods of the English language. The option to additionally investigate frequency distributions is a major advantage which is only possible due to the advent of corpus linguistics and digital technologies.

In the second case study the focus was on the analysis of I think. By comparing two studies that work on the same data set with a different categorization, it was pointed out that neither study is sufficient in describing the status of I think. While Mindt focused on syntactic, semantic and pragmatic functions in determining where I think functions as a discourse marker, Dehé and Wichmann put their focus on analyzing prosodic prominence. From their point of view, an acoustic analysis accounts for determining the status of I think. The result of the comparison of the two studies is that a large range of functions must be taken into account when determining the discourse status of I think. The second case study makes clear that contextual information from a wide range of the descriptive branches of English linguistics is needed to account for a solid analysis of I think. Kaltenböck (2008) is an example of this as he has investigated the relation between prosody and pragmatic functions. Furthermore, the interpretation of the data is best considered as scalar or on a continuum rather than categorical. The analysis of spoken data poses new challenges for researchers. An acoustic analysis of the spoken language offers many opportunities. Prosodic prominence is clearly observable in an acoustic analysis of the sound files. However, in how far the acoustic measurements are also important stimuli in terms of auditive perception by a listener is something that needs to be researched thoroughly.
Much work needs to be done to relate the findings from acoustic research to linguistic theory.

5. Conclusion

If philological techniques are understood as analyzing items in a text as well as their surrounding context, then corpus linguistics does, can and will use these techniques. The two case studies outlined above have clearly demonstrated that a detailed analysis of every single case and its context is necessary to arrive at a comprehensive description which then allows for a linguistic categorization. A next step is to incorporate these findings into a linguistic theory. In a linguistic analysis, a focus on the context is important and necessary, especially when a detailed qualitative analysis is required. However, corpus linguistics has more to offer. The quantitative analysis of the data is one major asset. By combining quantitative with qualitative approaches, a thorough investigation is possible. Due to advances in digital philology, corpus linguistics offers new strands of analyzing data which were not possible in the 18th, 19th and the first half of the 20th centuries.

An important point to be made is that the research results should be directly derived from the language data. This allows researchers to work on the same set of data and ideally to reach the same conclusions after having analyzed the data. One obvious problem is that much of the results of an analysis are interpreted. Interpretations depend among other factors on the theoretical background applied in a study as well as on the focus and aim of the study. By nature, interpretations might deviate from each other. In an idealized world, the replicability of the studies with the same data set, same methods and identical analysis tools should yield identical results.

With regard to the investigation of spoken language, technical advances have made it possible to record, store and analyze spoken data. For an up-to-date account in this area see Durand et al. (2014). The acoustic analysis of speech is one area of research that offers unprecedented options to researchers. The possibility to directly access the sound files makes it possible to analyze not only phonetic detail but also suprasegmental features and other speech-related acoustically detectable aspects. An extension is – of course – a multi-modal study, where video data come in so that one can study not only the audio data but additionally gesture, mimic etc.

A combination of quantitative with qualitative methods of analysis seems to be the best possible way when conducting linguistic research. Thus, philological techniques of observing the context in detail by taking into account various linguistic features should be combined with the technical advances corpus linguistics has to offer. With regard to the study of spoken data this could also mean that the sound files are available and can be accessed easily. Aligning the sound with other tiers of analysis then indeed makes possible an analysis from within. The new options, which digital data provide, should be added to methods and approaches already in use. These options also include challenging ideas of statistical analyses as advocated by Gries (2015) or Hardie (2012). The target question when working with language data should be how to best describe them from a linguistic point of view. The analysis and categorization of the linguistic data must come from within.

7 Interpretations of linguistic data were also given in the second case study in this paper.
Works Cited

1. Corpora


Davies, Mark. The Corpus of Historical American English: 400 million words, 1810-2009. 2010-. Data used up to February 2015.


Huber, Magnus, Magnus Nissel, Patrick Maiwald, and Bianca Widlitzki. 2012. The Old Bailey Corpus. Spoken English in the 18th and 19th centuries. February 2015 <www1.uni-giessen.de/oldbaileycorpus>.


The Helsinki Corpus of English Texts. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English). Department of Modern Languages, University of Helsinki, 1991.


2. Secondary Literature


Dehé, Nicole, and Anne Wichmann. "Sentence-initial 'I think (that)' and 'I believe (that)'; prosodic evidence for use as main clause, comment clause and discourse marker." Studies in Language 34.1 (2010): 36-74.


—. "The diachronic development of 'COME TO V'." *Silesian Studies in English.* Forthcoming.

