



BRILL

# In the Case of Theme: Topic Identifiers in English and Norwegian Academic Texts

*Hilde Hasselgård*

University of Oslo, Oslo, Norway

*hilde.hasselgard@ilos.uio.no*

## Abstract

This paper investigates the use of clause-initial constituents prefaced by topic-identifying expressions such as *in terms of*, *in the case of* and their Norwegian counterparts. The focus is on the nature, frequency and discourse functions of these in a corpus of published academic writing in English and Norwegian and across three disciplines. Such expressions are rather infrequent overall, but medicine uses them the least and linguistics the most in both languages. The functions of the construction can be compared either to those of left dislocation or to other types of clause-initial adverbials depending on the degree of coreference between the theme and some element in the rheme. The pattern with coreference is more common in Norwegian than in English. Generally, topic identifiers are used for announcing explicitly a theme that represents a topic shift or a contrast with the preceding discourse. The study contributes to contrastive pragmatics through its focus on the discourse-pragmatic functions of the expressions under study and the cross-linguistic comparison of this type of information structuring device across different disciplines of academic writing.

## Keywords

theme – topic identifier – left dislocation – discourse structure – academic writing – English/Norwegian

## 1 Introduction

Contrastive pragmatics deals with how two (or more) different languages are used in context to create meaning (Kranich 2016: 4, Verschueren 2016). This

study aims to contribute to the field by examining a particular device for signalling a discourse topic at sentence level, namely clause-initial constituents, or clause themes, introduced by expressions such as *in the case of*, *as regards*, *with respect to* and their Norwegian counterparts such as *når det gjelder*, *i forhold til*, *med hensyn til*, as illustrated in (1) and (2).<sup>1</sup> I will refer to such expressions as ‘topic identifiers’.

- (1) *In the case of writing*, the English orthographic system bears traces of Old Norse ... (engling<sub>30</sub>)
- (2) *Når det gjelder ordfølgen i Holbergs prosa*, finner vi både direkte og invertert ordstilling godt belagt, ... (noling<sub>31</sub>)  
[When it concerns the word order in Holberg’s prose, find we both direct and inverted word order well documented.]

In a section on how to identify a clause theme, Halliday writes that “Sometimes in English the Theme is announced explicitly, by means of some expression like *as for ...*, *with regard to ...*, *about ...*” (1994: 39).<sup>2</sup> Phrases such as *in the case of writing* consist of a complex preposition (Biber et al. 1999: 75; Teleman et al. 1999: 718) and a complement. They function as adjuncts and thereby marked themes (Halliday 1994: 44). The markedness is both syntactic and discursive: the topic identifier draws attention to the thematic element and presents it as new or contrastive information in relation to the preceding context.

The current investigation was motivated by an observation (Hasselgård 2018; 2019) that the expression *when it comes to* is frequent in academic texts by Norwegian users of English but absent from texts by native speakers. Suspecting that the overuse was L<sub>1</sub>-induced, I became interested in comparing topic identifiers in English and Norwegian. Using a comparable corpus of published academic articles (Section 4) this study tackles the following research questions:

- What topic identifiers are found in English and Norwegian academic texts, and how frequent are they?
- What are the similarities and differences in the preference of topic identifiers across languages and disciplines?
- What are the discourse functions of topic identifiers in both languages?

The expectations are that Norwegian topic identifiers are more frequent than English ones (due to the overuse in Norwegian-produced L<sub>2</sub> English) and that

1 The examples are from the KIAP corpus (Section 4). Norwegian examples are followed by a literal translation given in square brackets. The relevant expressions are italicised. Three dots at the end of an example indicate that the sentence continues beyond the quoted part.

2 See Section 2 for a more detailed definition of ‘theme’.

expressions that are similar in form may have different conditions of use. This may in turn lead to different preferences among the topic identifiers that exist in both languages. Previous studies of related phenomena (Section 2) have discussed the discourse functions of thematising constructions in general, hence providing a point of departure for investigating the pragmatic functions of topic identifiers in English and Norwegian academic texts. Linguistic differences between academic disciplines have been explored for English (e.g. Hyland 2012), but cross-disciplinary studies of Norwegian and cross-linguistic studies of academic writing are still rare (but see Fløttum et al. 2006). The present study contributes to filling this research gap.

The expressions functioning as topic identifiers can also introduce phrases in non-initial position, as in (3). Non-initial uses are not regarded as topic-identifying, but rather as expressions of a restriction on the validity of the proposition in the matrix clause (Hasselgård 2018: 109). Adjuncts such as the one in (3) have scope only over the predicate (Hasselgård 2010: 48). In initial position, however, they function as topicalisation constructions (Lambrecht 1994: 147–149) assigning the pragmatic notion of ‘aboutness’ to a constituent (ibid.: 150).

- (3) We find that our results are robust *with respect to choice of measure*.  
(engecon04)

Although clause-initial uses of topic identifiers are the main concern of this study, non-initial occurrences will receive some attention in Section 5. They are relevant insofar as otherwise similar-looking expressions in Norwegian and English have different potentials for being used initially with a discourse-pragmatic function.

In the remainder of this article, Section 2 reviews the theoretical framework of the study and relevant previous research. Section 3 summarises some similarities and differences between English and Norwegian word order and thematic structure. The material and method for the study are discussed in Section 4 before the study proper is presented in Section 5, which presents similarities and differences between the languages and between the academic disciplines. Sections 6 and 7 offer further discussion and concluding remarks.

## 2 Theoretical Framework and Previous Research

The study uses a predominantly systemic-functional linguistic (SFL) framework for analysing theme. SFL identifies theme in English clauses by its initial

position. The theme must contain an experiential (referential) element, but can in addition comprise conjunctive and modal adjuncts such as *however*, *of course* (Halliday and Matthiessen 2004: 79). The experiential part of the theme, referred to as *topical* (*ibid.*), is characterised functionally as “the element which serves as the point of departure of the message; it is that which locates and orients the clause within its context” (2004: 64). In declarative main clauses, the subject normally functions as theme; other types of clause-initial constituents are *marked themes* (2004: 73). The marked themes introduced by a topic identifier are ‘matter adjuncts’ probed by the question “What about?” (*ibid.*: 263).

The definition of ‘theme’ has been much debated, as has its delimitation (Thompson and Thompson 2009). This study follows Halliday and Matthiessen’s principle that the theme “ends with the first constituent that is either participant, circumstance or process” (2004: 79). This definition may not apply as generally to Norwegian as to English (*ibid.*: 81; Hasselgård 2005: 40 ff), but for the purposes of this study, it is sufficient. Unlike Halliday and Matthiessen I apply the thematic analysis not to the clause but to the T-unit, i.e. “a clause complex centred around an independent clause” (Thompson and Thompson 2009: 46), which entails that a dependent clause can function as theme. As noted by Thompson and Thompson (2009: 46), the T-unit approach is common in discourse-oriented studies of thematic structure.

In systemic-functional analysis, a clause is textually divided into theme and rheme. While theme is explicitly identified, the rheme is more loosely defined as e.g. “the remainder of the message” (Halliday and Matthiessen 2004: 64), or the part of the clause that follows the theme. Note that theme and rheme are identified purely sequentially in SFL and do not equal given and new information. The theme is what the speaker chooses to mention first, and given information is what the hearer is assumed to have access to (*ibid.*: 93). Hence, the thematic or rhematic placement of constituents such as *with respect to choice of measure* (example 3) reveals whether or not the constituent functions as the point of departure of the message but says nothing about information status. However, thematic structure (theme-rheme) and information structure (given-new) often correlate (*ibid.*), see also Davies (2017: 308).

Fontaine (2013) uses Thompson’s (2004) term ‘preposed Theme’ for constituents such as those italicised in (4) and (5). The construction – regardless of its realisation as PP or NP – is described as “a way of highlighting the Theme by announcing it before the clause. It also allows the speaker to direct the addressee’s focus” (Fontaine 2013: 154).

- (4) *As for the little old woman*, she took off her cap. (Fontaine 2013: 154)  
 (5) *Happiness*, that’s what life is about. (Thompson 2004: 153)

Thompson's (2004: 153) examples of preposed theme are all nominal constituents, including (5), and thus represent what is probably better known as 'left dislocation' (Geluykens 1992) or 'left detachment' (Lambrecht 1994). Left dislocation seems to share many of its syntactic and functional characteristics with thematic matter adjuncts. Interestingly, both Prince (1998) and Geluykens (1992) consider such examples as (4) a variant of left dislocation, and Lambrecht argues that "constructions of the *as-for* type [...] are detachment constructions in disguise" (1994: 182).

As noted above, the themes investigated here can be analysed as a type of adjunct adverbial, termed 'matter adjunct' in Halliday and Matthiessen (2004: 263) and 'respect adjunct' in Hasselgård (2010: 244 ff). In the corpus studied by Hasselgård (2010), respect adjuncts were infrequent overall, but a "striking finding [was] that academic writing contains more than half of the respect adjuncts in the entire material" (2010: 247).<sup>3</sup> However, only ten of these (c. 10%) were clause-initial (2010: 68), thus giving no basis for conclusions on thematic respect adjuncts. Gosden (1992) includes *in the case of* among marked themes with the function of 'real condition' in his study of research articles (1992: 213). Apart from remarking on the dominance of this expression within its (infrequent) category (1992: 221), however, he offers no detail on its use. Possibly due to their generally low frequency, thematic respect/matter adjuncts in English or Norwegian have been little studied, but some relevant work on the related phenomenon of left dislocation will be discussed below.

Studies of left dislocation (of nominal constituents) agree that the phenomenon belongs chiefly in informal speech (Prince 1998, Aijmer 1989, Geluykens 1992, Gómez-González 2001). For example, Geluykens (1992: 33) reports that left dislocation is vastly more frequent in conversation than in other spoken and written text types, which leads him to describe it as a "typically conversational phenomenon" (1992: 34). 'True dislocation' is characterised by coreference between the dislocated element and a constituent in the matrix clause (1992: 19). Similarly, Gómez-González considers it a key feature that the construction contains "a COREF clause-internal pro-form that is *anaphoric*, i.e. that refers back to, the external topical Theme" (2001: 287). Examples (4) and (5) thus qualify as true left dislocation due to the coreference between *woman – she* and *happiness – that*. Similar views are found in Prince (1998) and Fontaine (2013), but Aijmer (1989) seems to relax the requirement of coreference in

3 Hasselgård (2010) examined a subset of the British component of the International Corpus of English, ICE-GB (2010: 6–10).

observing that “there are sentences with left-dislocated structures in which the following discourse unit contains no source for the element in dislocated position” (1989: 138).

Geluykens (1992: 22) uses the term ‘quasi-LD’ for constructions that are superficially similar to true dislocations, but lack the element of coreference. An example is given in (6).

(6) *As for London, Trafalgar Square is nice.* (Geluykens 1992: 21)

Most of Geluykens’s examples of quasi-dislocation are actually introduced by a topic identifier. However, the same topic identifiers may appear in examples classified as both true and quasi-LDs, e.g. *as for* (1992: 20–21). Interestingly, “the type of constructions which we have labelled quasi-LDs [...] are relatively frequent in non-conversational discourse” (1992: 115).

A major function of left dislocation, according to Geluykens, is to (re-)introduce a referent into the discourse (1992: 49). The dislocated element may also bear a contrastive relation to a preceding topic (1992: 86) or represent an item on a list (1992: 89). Quasi-dislocations are similarly considered to be primarily referent-introducing: “they explicitly mark a referent as becoming topical in the subsequent discourse” (1992: 133). Aijmer (1989) argues that the dislocated element (‘Theme’ in her terminology) can be used to establish a discourse topic for one or more utterances (1989: 143). Sometimes “a referent can be placed as Theme if it has been previously mentioned but no longer remains relevant or accessible for processing by the hearer” (1989: 145–146). The latter point is echoed by Gómez-González, who finds that a majority of left-dislocated elements “re-introduce a participant or circumstance in discourse” which is either directly recoverable or inferable from preceding discourse or the situational context (2001: 193).

The Norwegian reference grammar describes left dislocation in terms of a ‘detached front-field’ (‘løst forfelt’) in which the topic of a sentence occurs outside the core of the clause, as in *Ibsen, han var ein stor dramatikar* [“Ibsen, he was a great dramatist”] (Faarlund et al. 1997: 904). The detached phrase is normally repeated in the matrix clause by means of a pro-form (1997: 905), but there are also cases without a pro-form as in *Jula ja, du har vel huska å kjøpe julegaver* [“Oh Christmas, you have remembered to buy presents, have you?”] (1997: 906). However, in such cases there is usually an indirect connection between the detached element and the content of the clause (ibid.). As in English, a detached front-field is more typical of speech than of writing (Faarlund et al. 1997: 907 f).

### 3 Word Order and Thematic Structure in English and Norwegian

English and Norwegian share many syntactic features such as the lack of grammatical case and the use of word order to indicate syntactic function. An important difference is that Norwegian is a verb-second (V<sub>2</sub>) language, i.e. the finite verb must be the second constituent in a declarative main clause (Faarlund et al. 1997: 859; Holmes and Enger 2018: 415), while English is more consistently SV (Biber et al. 1999: 153). For clauses with topic identifiers, this means that the subject follows the marked theme directly in English, as in (1) above, but is preceded by the finite verb in Norwegian, as in (2). Incidentally, a left-dislocated NP does not trigger subject-verb inversion in Norwegian (Holmes and Enger 2018: 448), indicating that a thematic PP is better integrated in the clause structure.

The choice of thematic constituent is rather similar in English and Norwegian, with the subject being the preferred option. In fact, the syntactic potential of the two languages is similar enough that the topical theme is almost always retained in translation between them (Hasselgård 2004: 194 f). However, Norwegian thematises non-subject constituents, particularly adverbials, more often than English (Hasselgård 2004: 190; 2005: 36). Furthermore, Norwegian seems to have a stronger preference than English for syntactically and informationally light themes, which led Hasselgård (2005: 45) to conclude that Norwegian themes are normally associated with less prominence than English ones.

## 4 Material and Method

### 4.1 Corpora

This study uses two corpora for two different purposes. The main investigation is based on the KIAP corpus,<sup>4</sup> developed as part of the project *Cultural Identity in Academic Prose* (Fløttum et al. 2006). The corpus contains published academic articles in Norwegian, English and French from the three academic disciplines economics, medicine and linguistics (only English and Norwegian are studied here). Each discipline is represented by 50 journal articles in each language. Fløttum et al. (2006: 7) observe that the English articles are on average longer than the Norwegian ones, hence the different word counts shown

4 KIAP is an acronym for the Norwegian project title, *Kulturell Identitet i Akademisk Prosa*, which is a word-by-word counterpart of the English one (Fløttum et al. 2006: 1X).

TABLE 1 Number of body words in the English and Norwegian parts of the KIAP corpus (Fløttum et al. 2006: 7)

	Economics	Linguistics	Medicine	Total
English	298,319	437,798	163,663	899,780
Norwegian	312,850	269,913	102,660	685,423
Total	611,169	707,711	266,323	1,585,203

in Table 1. The numbers include only the so-called ‘body words’ of the articles, thus excluding e.g. titles, abstracts, references, and acknowledgements. The corpus searches were performed using the online search engine Corpuscle.<sup>5</sup>

KIAP is a comparable corpus, meaning that there is no translation relation between the texts in the different languages (Hasselgård, *forthc.*). Therefore, to ensure that the selected expressions can be regarded as semantically and functionally equivalent across the languages, a bidirectional translation corpus was used in a preparatory investigation. This corpus, the English-Norwegian Parallel Corpus (ENPC; Johansson et al. 1999/2002), contains comparable fictional and non-fictional text extracts in English and Norwegian, each translated into the other language. Only the non-fiction part of the corpus was used in this study. ENPC non-fiction contains about 252,000 words of original English and 220,100 words of original Norwegian while the translations comprise 252,700 words in English and 244,000 words in Norwegian.

#### 4.2 *Establishing a Tertium Comparationis and Identifying Search Terms*

‘Topic identifiers’ are not an established linguistic category, so a set of expressions had to be identified for each language. To this end, and to ensure that the two sets of topic identifiers were as equivalent as possible, a preparatory study was undertaken using the ENPC. The idea was to apply the technique of *translation paradigms* (Johansson 2007: 23), or *semantic mirrors* (Dyvik 2004: 315) in order to gain an overview of the expressions used for topic marking in both languages. The translation relation between the expressions gives the study a sound *tertium comparationis*, i.e. a background of sameness that enhances the reliability of the study (Johansson 2007: 39; Hasselgård *forthc.*). The other elements of *the tertium comparationis* are the common functional descriptive

<sup>5</sup> <http://clarino.uib.no/korpuskel/>.



framework and the similarity between the KIAP subcorpora with regard to register and academic discipline.

The expression that triggered the study, *when it comes to*, was not found in English originals in the ENPC, but occurred as a translation of *når det gjelder*. Hence, the investigation started from the Norwegian expression, which is quite common in Norwegian original non-fiction (38 hits, or 17.3 per 100,000 words). An example is given in (7).

- (7) *Når det gjelder* anleggsmidler for øvrig benyttes følgende prinsipper:  
(ENPC: NFRA1)  
[When it concerns fixed assets beyond this ...]  
*With regard to* all other fixed assets, the following principles have been utilised: (NFRA1T)

The following translations of *når det gjelder*, given in descending order of frequency, were found more than once and constitute the translation paradigm of the expression: *with regard to*, *in terms of*, *when it comes to*, *in*, *where X is concerned*, *for*. A search in Norwegian translations of *når det gjelder* yielded much the same expressions as English sources, plus *as regards*, *in the case of* and *concerning*. These expressions, minus the simple prepositions,<sup>6</sup> were then searched for in English originals to identify their Norwegian translation paradigms. The procedure was repeated for recurrent members of the new paradigms, resulting in the set of expressions detailed in Table 2, which shows correspondences, i.e. both translations and sources (Johansson 2007: 23), of the expressions investigated. Each expression included in Table 2 accounts for at least 1% of the total occurrences of the expressions searched for, thus the cut-off is 4 for the English correspondences and 3 for the Norwegian ones.<sup>7</sup> Note that Table 2 comprises both thematic and rhematic occurrences because the purpose of the ENPC study was to elicit as many expressions as possible. Simple prepositions constitute a large proportion of the correspondences in both languages. The most frequent ones are the English *in*, *for*, *about*, *to*, *of* and *on*, and the Norwegian *om* ('about'), *for*, *overfor* ('facing, in relation to') and *i* ('in'). *Om*, with 61 occurrences, mostly corresponds to *concerning*.

6 The reason for excluding simple prepositions was that the number of irrelevant hits was too high.

7 This excluded *when it comes to*, which occurred only three times (as a translation of *når det gjelder*).

TABLE 2 Recurrent correspondences of the Norwegian expressions *når det gjelder, med hensyn til / mht, i forhold til, i forbindelse med* and the English expressions *in terms of, with regard to, concerning, in the case of, in relation to, as regards*

Correspondences of Norwegian expressions (N = 308)	Freq.		Correspondences of English expressions (N = 279)	Gloss	Freq.	
	N	%			N	%
<i>in relation to</i>	26	8.4	<i>når det gjelder</i>	when it concerns	33	11.8
<i>as regards</i>	16	5.2	<i>i forhold til</i>	in relation to	24	8.6
<i>with regard to</i>	16	5.2	<i>med hensyn til / mht</i>	with regard to / wrt	21	7.5
<i>in connection with</i>	12	3.9	<i>i forbindelse med</i>	in connection with	9	3.2
<i>compared to/with</i>	9	2.9	<i>som gjelder</i>	which concerns	7	2.5
<i>in terms of regarding</i>	9	2.9	<i>hva angår X, hva X angår</i>	what concerns X, what X concerns	6	2.2
<i>concerning</i>	4	1.3	<i>vedrørende</i>	concerning	3	1.1
<i>in the case of</i>	4	1.3	simple prepositions		109	39.1
<i>with respect to</i>	4	1.3				
<i>in regard to</i>	4	1.3				
simple prepositions	78	25.3				

## 5 Corpus Analysis

### 5.1 *Thematic and Rhematic Uses of Potential Topic Identifiers in the KIAP Corpus*

The recurrent expressions identified via the ENPC (cf. Table 2) were retrieved from KIAP. Some of the expressions from the ENPC never occurred thematically as topic identifiers in KIAP (e.g. *som gjelder, vedrørende, in connection with*), and *compared to/with* marked contrast rather than topic. These were therefore excluded from further investigation. Also excluded were simple prepositions due to the huge amounts of manual analysis required to distinguish

TABLE 3 Potential topic identifiers and their frequencies in the KIAP corpus (raw and normalised per 100,000 words)

English			Norwegian		
	#	norm		#	norm
<i>in terms of</i>	339	37.7	<i>i forhold til</i>	349	50.9
<i>with respect to</i>	253	28.1	<i>når det gjelder</i>	224	32.7
<i>in the case of</i>	168	18.7	<i>med hensyn til</i>	111	16.2
<i>concerning</i>	96	10.7	<i>mht</i>	29	4.2
<i>with regard to</i>	58	6.4	<i>i forbindelse med</i>	80	11.7
<i>as regards</i>	5	0.6	<i>hva angår X, hva X angår</i>	13	1.9
<i>when it comes to</i>	5	0.6			
Total	924	103.0		806	117.6

topic identifier uses from all other occurrences.<sup>8</sup> Table 3 shows the expressions that were included in the investigation of the KIAP corpus and their total frequencies in both thematic and rhematic position.<sup>9</sup> The selected expressions occur in 122 out of the 150 English KIAP texts and in 126 out of the 150 Norwegian ones.

The frequencies reported in Table 3 are higher relative to corpus size in Norwegian than in English (LL = 7.86,  $p < 0.01$ ). However, this difference should be interpreted with caution, since the frequencies may not be directly comparable considering that different numbers of expressions have been included and excluded in the two languages.

A manual examination of concordance lines in order to separate thematic and rhematic uses revealed that the expressions have very different likelihoods of being used thematically, i.e. as topic identifiers, as visualised in Figure 1. Most of the expressions are only sporadically thematic in the KIAP texts. *In the*

8 *As for*, mentioned e.g. by Geluykens (1992) and Lambrecht (1994), is absent from this investigation because it was infrequent in the ENPC and because searches for it have low precision, particularly in rhematic position. A case-sensitive search retrieves sentence-initial instances, but misses any that are preceded by e.g. a conjunction.

9 Numbers for *når det gjelder* include the 'Nynorsk' variant *når det gjeld*. The 'Nynorsk' counterparts of other expressions were not found in thematic position and were therefore not counted.

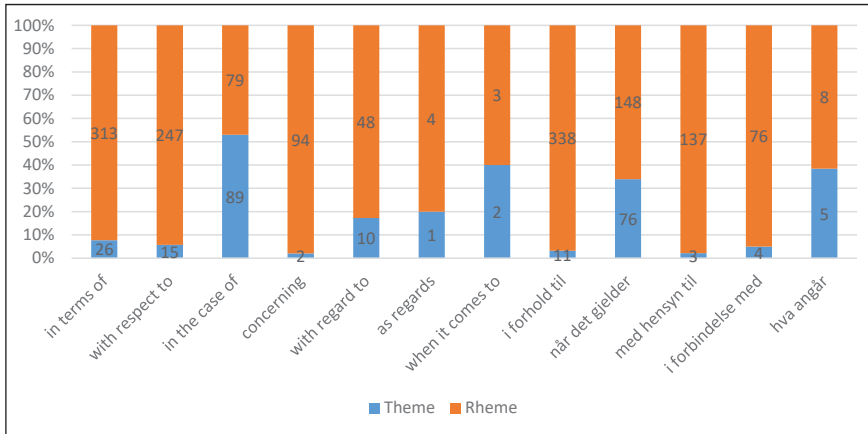


FIGURE 1 Thematic and rhematic uses of potential topic identifiers in KIAP

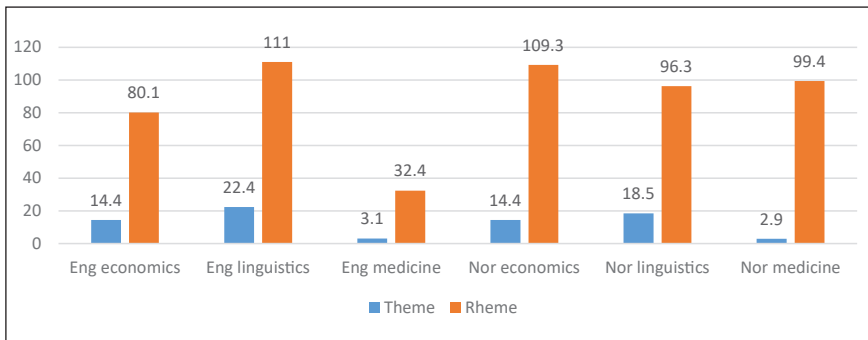


FIGURE 2 Thematic and rhematic uses across languages and disciplines. Frequencies per 100,000 words

*case of* and *når det gjelder* stand out among the frequent expressions by being thematic in over half and a third of the cases, respectively.

The disciplines in the KIAP corpus use topic identifiers to different extents. Figure 2 shows normalised frequencies of thematic and rhematic uses across languages and disciplines. While rhematic position is most common across the board, thematic use is marginal in medical articles in both English and Norwegian. The highest frequencies of thematic use occur in linguistics in both languages. In economics, thematic use in economics is equally frequent between the languages. Rhematic uses are more frequent in Norwegian than in English except in linguistics. The following sections leave rhematic occurrences aside and instead explore topic-identifying uses since only these can be defined as topic identifiers.

TABLE 4 The frequencies and dispersion of topic identifiers in English and Norwegian economics and linguistics

Language/ discipline	Topic identifier		Texts (of 50)		Frequency per text (N)		Dispersion measures	
	N	/100k	N	%	Min-max	Mean	SD	Juillard's D
Eng economics	43	14.4	24	48	1–5	1.8	1.17	0.81
Eng linguistics	98	22.4	31	62	1–8	3.2	2.33	0.83
Nor economics	45	14.4	26	52	1–5	1.7	1.12	0.82
Nor linguistics	50	18.6	22	44	1–7	2.4	1.64	0.77

### 5.2 *English and Norwegian Topic Identifiers in Economics and Linguistics Texts*

In addition to narrowing its focus to (thematic) topic identifiers, the remainder of this study will concern only economics and linguistics, due to the near-absence of topic identifiers in medicine (Figure 2). Table 4 shows the frequency and dispersion of topic identifiers in English and Norwegian economics and linguistics articles in KIAP. As emphasised by e.g. Brezina (2018: 47) an analysis of dispersion is essential for making sure that the high frequency of a linguistic item is not linked to particular contexts or the preferences of individual writers.

Table 4 shows raw and normalised frequencies of topic identifiers per discipline, the number and percentage of texts that contain at least one topic identifier, the number of topic identifiers and the mean frequency per text in those texts that contain topic identifiers at all, and finally the values for standard deviation and Juillard's D across the 50 texts in each discipline.<sup>10</sup> The figures for English and Norwegian economics are practically identical. While the frequencies of topic identifiers are higher in linguistics than in economics in both languages, English linguistics has the highest frequencies and the widest dispersion of them. Moreover, more texts contain many topic identifiers in English linguistics (six papers have 6–8 instances), but recall that this discipline also has the longest texts (compare Table 1 and Figure 2). The standard deviation shows that there is less variation within economics than linguistics in both languages. Juillard's D shows that the degree of homogeneity of the

<sup>10</sup> The dispersion measures were calculated with Lancaster Stats Tools online (<http://corpora.lancs.ac.uk/stats/toolbox.php>).

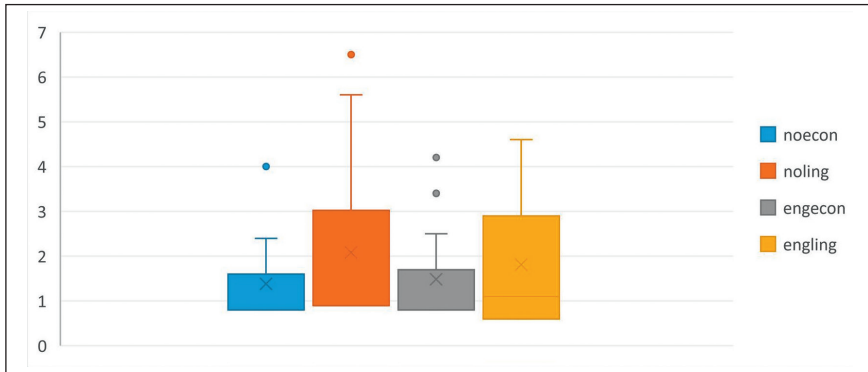


FIGURE 3 Dispersion of topic identifiers across texts in English and Norwegian economics and linguistics (normalised frequencies)

distribution is relatively similar across the subcorpora, though the distribution is slightly less even in Norwegian linguistics (cf. Brezina 2018: 51).

A different view of the dispersion is given in Figure 3, which ignores those corpus texts that do not contain topic identifiers and takes account of the different sizes of the subcorpora. Unfortunately, the sizes of individual corpus texts were not available, so instead I calculated the mean text length per subcorpus and normalised the occurrences in each text per 5,000 words. The figure shows again the greater variation in linguistics than in economics in both languages: Norwegian and English economics have very similar frequencies and dispersions. The highest frequency of topic identifiers occurs in a Norwegian linguistics text, but the greatest interquartile range is found in English linguistics, which is also the only discipline where the median frequency is noticeably different from the minimum frequency. Recall also the higher proportion of texts without topic identifiers in Norwegian linguistics (Table 4). To sum up, the frequency and dispersion measures show more differences across disciplines than across languages. To the extent that there is a cross-linguistic difference, it occurs in linguistics, not in economics.

### 5.3 *Coreference in Sentences with Topic Identifiers*

As detailed in Section 2, accounts of left dislocation usually require coreference between the dislocated element and an element elsewhere in the sentence. This section examines coreference between the NP following the topic identifier (henceforth referred to as the ‘identified topic’) and a rhematic constituent. My analysis of coreference takes a scalar view inspired by Hasan’s (1985) work on cohesive chains, which, in addition to (full) coreference, includes the concepts of co-classification and co-extension (1985: 73–74).

The latter two imply that elements can be related on account of similarity of reference through e.g. hyponymy or synonymy, or by belonging to the “same general field of meaning” (1985: 74). In contrast to e.g. Gelyukens (1992), the present analysis, like Hasan’s, allows non-pronominal coreference, i.e. lexical repetition (see also Aijmer 1989: 145). There are three types of coreference with varying strengths:

- Full coreference: identify of reference through pronominal or lexical repetition as in (8).
- Partial coreference (co-classification): lexical relation between overlapping concepts: near-synonymy, hyponymy, meronymy, as in (9).
- Indirect coreference (co-extension): a relation that requires inference / world knowledge, as in (10).

- (8) In the case of *the economy as a whole*, it is sometimes conceptualised as a living organism ... (engling27)
- (9) Når det gjelder *klassifisering av verb* er altså, som allerede antydnet, forholdene mer kompliserte enn det som fremgår av *den tradisjonelle klasseinndelingen* ... (noling27)  
 [When it concerns *classification of verbs* is thus, as already indicated, the situation more complicated than what appears from *the traditional class-division*]
- (10) In the case of *the P<sub>1</sub> test* then, it seems there is a decisive *outcome*: ... (engecon33)

In (8) the complement of *in the case of* is resumed in the rheme by the coreferential pronoun *it*. In (9) the ‘classification of verbs’ is echoed by the near-synonymous ‘class-division’, while in (10) the concepts ‘test’ and ‘outcome’ are related by the fact that tests are known to have outcomes. As the examples show, the coreferential element in the rheme need not be a pronominal copy of the identified topic: (partial) repetition by means of a full noun phrase is also rather common, as in (9) and (10). Finally, there may be no coreference between the theme and the rheme, as illustrated by (11).

- (11) Også når det gjeld *kasus og forholdet mellom aktiv og passiv*, er det stort samsvar mellom norrønt og nyislandsk. (noling40)  
 [Also when it concerns *case and the relationship between active and passive*, is there large similarity between Norse and Modern Icelandic.]

Table 5 shows the different degrees of coreference between theme and rheme distributed across languages and disciplines in KIAP. The Norwegian material

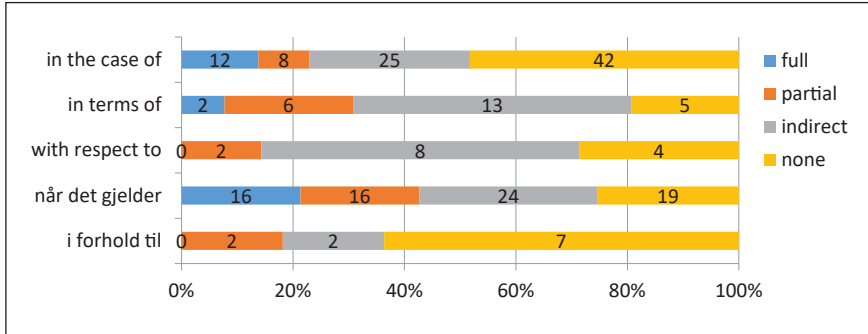


FIGURE 4 Coreference and lexical variation

has a larger share of full and partial coreference with a corresponding higher percentage of no coreference in the English material. Hence, the Norwegian construction may seem to be more closely related to left dislocation, while the English construction is more likely to convey a marked theme not resumed in the rheme.

Digging deeper into the issue of coreference, I examined individual topic identifiers. Figure 4 shows that there is considerable lexical variation on this point. Of the English expressions, *in terms of* and *with respect to* are more likely to occur with some degree of coreference than *in the case of*, which lacks coreference 50% of the time. By contrast, the most frequent Norwegian topic identifier, *når det gjelder*, is associated with coreference close to 80% of the time, while *i forhold til* is more likely to occur without coreference.<sup>11</sup> As shown in Figure 4, full coreference is rare (or non-occurring) with most of the expressions. It is more common with *når det gjelder* than with *in the case of*. Moreover, conflating the two strongest types of coreference, full and partial, we see a clear difference between the two most frequent English and Norwegian topic identifiers. This suggests that marked themes with *når det gjelder* may be closer to left dislocation than themes with other topic identifiers.

As a final point of studying topic identifiers and coreference, I considered the syntactic function of the coreferential element in the rheme. Some identified topics refer to the subject of the matrix clause, as in (12), where the pronoun *de* “they” has the same reference as the identified topic. In such cases, what might have been an unmarked subject theme, as in (12a), becomes a marked adjunct theme instead, thereby attracting more attention to itself.

11 Other topic identifiers are too rare (1–4 occurrences) to show any patterns and are not included in Figure 4.



TABLE 5 Coreference in sentences with topic identifiers

Language/discipline	full		partial		indirect		none		tot
	N	%	N	%	N	%	N	%	
Eng economics	1	2.3	2	4.7	21	48.8	19	44.2	43
Eng linguistics	16	16.3	14	14.3	32	32.7	36	36.7	98
Nor economics	8	17.8	10	22.2	15	33.3	12	26.7	45
Nor linguistics	10	20	9	18	14	28	17	34	50

- (12) Når det gjelder *de høye fremre vokalene*, så er *de* palatale i utgangspunktet ... (noling34)  
 [When it concerns the high front vowels, then they are palatal to start with ...]
- (12a) De høye fremre vokalene er palatale i utgangspunktet.  
 [The high front vowels are palatal to start with.]

However, coreference with the matrix clause subject accounts for only 17% of coreferential topics in Norwegian and 10% in English. Most identified topics refer to a non-subject, thereby pulling an otherwise rhematic constituent to thematic position. Examples (13)–(15) are more typical: in (13) and (14) the topic identifier thematises a constituent from a dependent clause, with a change of word class in (14) (*weighted* – *weight*). The identified topic in (15) represents a (potential) modifier of the subject NP, added in square brackets.

- (13) Indeed *in the case of PS* there is evidence that *he* relies on perceptual strategies ... (engling07)
- (14) *In the case of weighted abduction*, it is thus unclear how one can assign *the weights* that guide inference in a principled way. (engling09)
- (15) Når det gjelder *lokativ*, er Jakobsons klassiske *eksempel [på lokativ]* illustrerende: ... (noling06)  
 [When it concerns *locative*, is Jakobson's classic *example [of locative]* illuminating]

#### 5.4 *Pragmatic Functions of the Topic Identifier Construction*

A common pragmatic function of topic identifiers, regardless of the degree of coreference, is to announce the (topical) theme explicitly (Halliday 1994: 39). Those identified topics that resemble left-dislocated elements are especially likely to share their functions (Section 2), namely to “reintroduce or introduce the referents of the LD topical Themes in discourse” (Gómez-González 2001: 293). Aijmer (1989: 141 f) identifies three discourse functions of left-dislocated elements: (i) introducing a new topic in the conversation; (ii) exemplifying what the participants have been talking about; and (iii) switching the topic to something previously mentioned, but no longer active. The last point is also made by Lambrecht, who defines left detachment pragmatically as “a grammatical device used to promote a referent ... from accessible to active status” (1994: 183).

The material shows that single examples of topic identification can have several functions simultaneously. Thus, rather than attempting to quantify each function, I will discuss the roles of topic identifiers in the structure of Norwegian and English academic papers based on representative examples. The introduction of brand new topics appears to be an uncommon function: only a few instances were found. In (16) the identified topic *audience* appears for the first time in the text so it is discourse-new (Prince 1998) even if readers may be aware that *audience* is a relevant variable in register studies. The identified topic in (17) is also discourse-new although it may be inferable from the preceding context.

- (16) Dong (1998) describes doctoral theses which are based on a compilation of publishable research articles. These are quite different from other sorts of theses. The research article chapters are more concise than typical thesis chapters with less of the ‘display of knowledge’ that is often found in a thesis or dissertation. Further, *in terms of audience*, they are written more as ‘experts writing for experts’, than novices ‘writing for admission to the academy’. (engling11)
- (17) Alle har vel sine erfaringer med at den man genetisk står nærmest, ikke nødvendigvis er den man føler seg mest i slekt med. *Når det gjelder slekt-skap mellom språk*, opererer man tradisjonelt med Stammbaum – eller stamtremodellen. (noling27)
- [Everybody presumably has the experience that those you are genetically closest to are not necessarily those you feel most related to. *When it concerns relationships between languages*, one traditionally operates with the Stammbaum or family tree model.]

- (18) Investigating the visual dimension of *conference papers* will lead us to consider the important cognitive and rhetorical functions fulfilled by non-verbal representation in science in general, functions which the *conference paper* genre shares with other forms of scientific research discourse. Several analyses of these functions in written genres have highlighted the various semiotic resources (linguistic, mathematical, visual) used in science, enabling taxonomies to be drawn up which classify non-verbal representation into different types: tables, graphs, diagrams etc. *In the case of conference papers*, this “combinatorial semiotic principle” (Lemke, 1998: 92) is realised in a way specific to the spoken genre. (engling12)

A more common function is to reactivate a concept from the preceding discourse and give it topic status for one or more predications (Aijmer 1989: 143, 146). Example (18) occurs in an article introduction, and the identified topic is part of the discourse move of ‘establishing a niche’ (Swales 1990: 141) in relation to previous research. ‘Conference papers’ is mentioned at the beginning of the extract (as highlighted) and also before this, including the paper title (“Visual discourse in scientific conference papers. A genre-based study”). However, following a sentence with another topic, *in the case of* revives the concept of conference papers and highlights the contrast between these and the ‘written genres’ of the preceding sentence.

In (19), the identified topic, *tiltakseffekter*, is a compound noun whose components occur in the first sentence of the extract (in bold). The intervening sentences, however, make it necessary to reactivate this topic. The coreference between the identified topic and the rheme in (19) is indirect: the reader must infer that an extra month on an AMO course is a type of measure.

- (19) De estimerte koeffisientene for **tiltak** kan tolkes som gjennomsnittlige **marginaleffekter**. Ettersom utvalget består av søkere til AMO-kurs, fokuserer vi på effekten av deltakelse på AMO-kurs. <Four intervening sentences> *Når det gjelder tiltakseffekter*, finner vi at **en ekstra måned på AMO-kurs i tiltaksperioden** gir en økning i forventet varighet av bruttoledighet ... (noecon17)

[The estimated coefficients for **measures** can be interpreted as average **marginal effects**. Since the selection consists of applicants for AMO-courses,<sup>12</sup> we focus on the effect of participation in AMO-courses. <Four intervening sentences> *In terms of measures effects*, we find that

12 AMO-courses were labour-market oriented courses for unemployed people.

**an extra month on the AMO-course during the measures period** gives an increase in the expected duration of unemployment.]

A related function is that of maintaining coherence by navigating between two or several relevant discourse topics. This is demonstrated in (20), where two subsequent sentences with topic identifiers pick up different numbered topics from the preceding context.

- (20) ... rather what is at stake is (1) whether unidirectionality is a (testable) hypothesis at all, and (2) how the recognized directional tendency is to be explained [...]. *With regard to (1)*, the unidirectionality claim in its strong form that exceptions do not exist is clearly wrong, in light of counterexamples such as those presented in this paper. This does not negate a tendency towards directionality in grammaticalization changes, but does falsify the strong claim that the directionality is absolute or inviolable. *With regard to (2)*, I argue below ... (engling31)

Identified themes with weak or no coreference can have the same pragmatic functions as those with a coreferential item in the rheme, i.e. they can resume a previously discussed item and make it the topic for the current T-unit, as in (21), or mark a textual shift by explicitly navigating between different topics, as in (22). However, such themes connect more with the left than with the right context, while those with a coreferential rheme face both ways.

- (21) Resultatet viste signifikante forskjeller mellom gruppene i 3 av 5 situasjoner (Kjøkken, Haik, Politi). *Når det gjelder Politi-situasjonen*, er ikke spranget mellom nordmenn og gruppe 3 så stor som kurvene viser. (noling35)  
[The result showed significant differences between the groups in 3 of 5 situations (Kitchen, Hike, Police). *When it concerns the Police-situation*, the gap between Norwegians and group 3 is not as great as the graphs show.]
- (22) Since this does not shift his election probability sufficiently (the second part of Assumption 2), then he will not respond to the small shock even if he can pick up voters. *In the case of a large shock*, he can be responsive and gain 2 voters. (engecon14)

When there is no coreference between the identified topic and the rheme, the marked theme assumes a more clearly circumstantial meaning, as in (23)–(24),

and it may resemble a corresponding adjunct in rhematic position; i.e. it restricts the validity of the proposition in the matrix clause (Hasselgård 2018: 109). However, such adjuncts also thematise elements that are useful for the appropriate interpretation of the sentence in its context.

- (23) *In the case of hostile take-overs* even larger employment falls are noted and some gain in output per worker seems apparent. With friendly acquisitions employment falls are both smaller and experienced in the first year post acquisition. (engecon12)
- (24) *Når det gjelder tysk*, er forklaringen på det synkrone problemet enkel: (noling08)  
[When it concerns German, is the explanation of the synchronic problem simple.]

The element of contrast is salient in (23), where *hostile take-overs* evokes a concept in the preceding context and forms the first part of a contrastive pair with *friendly acquisitions* in the next sentence. The article in which (24) occurs constantly compares Old Norse to German and hence uses identified topics contrastively to navigate between two current topics. While the thematic adjuncts in (23) and (24) indicate aboutness, those in (25) and (26) approach temporal and spatial meaning, respectively (Hasselgård 2010: 39), and it is less obvious that they (re-)introduce discourse topics.

- (25) *I forbindelse med beregning av kostnads- og allokerings effektivitet* har vi også behov for prisinformasjon. (noecon21)  
[In connection with the calculation of cost – and allocation-efficiency, have we also a need for price information.]
- (26) Engel (1999) finds that *in the case of the U.S.*, almost all of the variation in the real exchange rate is due to failures in the law of one price for tradable goods, with a very small role played by the relative price of nontraded to traded goods. (engecon21)

In sum, Norwegian and English topic-identifiers seem to have similar discourse-functions: they situate and contextualise the current T-unit by activating a topic that may be discourse-new, or – more typically – previously mentioned but in need of reactivation, and the construction may involve a topic shift, possibly including a contrast in relation to the preceding context. However, given the different inclinations of both individual topic identifiers and the two languages in general to have a co-referent element in the rheme,

it appears that the languages select from the potential discourse functions to varying extents.

## 6 Discussion

Most of the topic identifiers are complex prepositions composed of preposition + NP + preposition (e.g. *in the case of*, *in terms of*, *i forhold til*, *med hensyn til*). However, the Norwegian *når det gjelder* contains a finite structure, despite Teleman et al. (1999: 719) classifying its Swedish cognate *när det gäller* as a complex preposition. Formally speaking, adverbials with *når det gjelder*, or the related *when it comes to* (27), are thus adverbial clauses.

(27) First, it shows that *when it comes to nominalization*, the supine pairs with the infinitive and not with the past participle. (engling24)

Adverbials realised by PPs have been observed to be more common in academic prose than in conversation in English, while adverbial clauses are more common in conversation (Biber and Gray 2016: 93–94). According to the Corpus of Contemporary American English (COCA), *when it comes to* is frequent in speech, magazines and newspapers (46, 42 and 32 pmw, respectively), but rare in academic prose (13 pmw). *In the case of* occurs 48 times pmw in academic prose and not above 20 elsewhere. *Når det gjelder*, by contrast, is frequent in academic prose (Table 3) as well as other registers, as revealed by searches in the Norwegian corpus for bokmål lexicography (LBK).<sup>13</sup> Thus, there is less of a register difference in Norwegian than in English in this area of lexical choice.

The analysis has shown that English and Norwegian topic identifiers share the function of thematising a participant to make it the topical theme of the current T-unit (and possibly also for subsequent T-units). In both languages, the identified topics may or may not have a co-referent item in the rheme. It may be argued that topic identifiers introduce three distinguishable but related constructions: those with full coreference between theme and rheme correspond to 'true' dislocation, while those with partial or indirect coreference are closer to what Geluykens calls 'quasi-dislocation' (1992: 22). Without coreference, the thematised constituent has a more clearly circumstantial

13 Websites with further information on COCA and LBK are listed at the end of this paper.

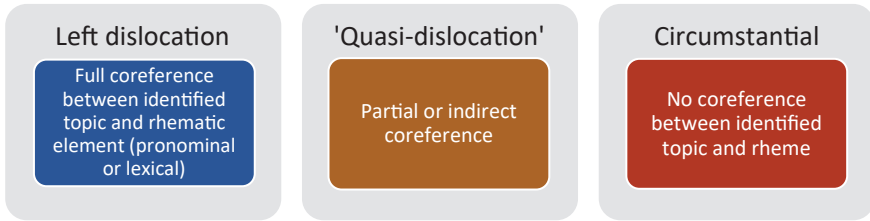


FIGURE 5 Coreference and construction type

meaning, as visualised in Figure 5. Norwegian has been shown to choose the first type more frequently.

The difference between the three types is reflected in the fact that it seems harder to move the dislocation type to clause-final position, as evidenced by (28)–(30), which are paraphrases on (8), (16) and (24) above. More precisely, the acceptability of (28) seems doubtful while (29) is slightly better and (30) is completely acceptable (albeit less suited to its context).

- (28) \* It is sometimes conceptualised as a living organism *in the case of the economy as a whole*, ... (full coreference with rhematic subject)
- (29) ? Further, they are written more as ‘experts writing for experts’, than novices ‘writing for admission to the academy’ *in terms of audience*. (partial coreference with non-subject part of rheme)
- (30) Forklaringen på det synkrone problemet er enkel *når det gjelder tysk*: “The explanation of the synchronic problem is simple when it concerns German.”

## 7 Summary of Findings and Concluding Remarks

The present study has illuminated pragmatic contrasts between English and Norwegian in three ways (cf. Verschueren 2016): by giving a discourse-functional analysis of a grammatical construction, comparing its topicalisation potential across languages, and by comparing academic disciplines in two languages. The first research question asked what topic identifiers occur in English and Norwegian academic texts, and how frequent they are. In a first step to answering this question (Section 4.2), the bidirectional translations in the ENPC were exploited to establish paradigms of corresponding expressions in English and Norwegian and thus obtain a set of semantically and functionally similar terms for further contrastive study on the basis of KIAP. A crucial point of this

exercise was to enhance the *tertium comparationis* for the comparable corpus investigation.

The resulting set of expressions formed the basis for examining topic identifiers in KIAP, where their overall frequencies were somewhat higher in Norwegian than in English (5.1). Most of the expressions were, however, far more common in rhematic than in thematic position in both languages, thus demonstrating that near-synonymous expressions can have very different colligational preferences. The most frequently thematic ones were *når det gjelder* and *in the case of*. The academic disciplines use topic-identifying expressions with different frequencies in both languages: in both English and Norwegian, medicine has very few thematic uses, while linguistics has more than economics. In rhematic position the picture is less consistent: while English has the same rank frequency as in thematic position, Norwegian economics is higher than linguistics and medicine, which have similar frequencies (Table 2).

As regards (thematic) topic identifiers, linguistics has higher frequencies as well as more individual variation (5.2). The disciplinary difference appears greater in English than in Norwegian: while English and Norwegian economics papers seem to use topic identifiers in similar fashion, English linguistics has a higher frequency and wider dispersion of topic identifiers than both economics and Norwegian linguistics. The favoured lexical realisations of topic identifiers, *når det gjelder* and *in the case of*, are both vastly more frequent than the second most frequent ones (*i forhold til* and *in terms of*) across the disciplines. In contrast to *når det gjelder*, *in the case of* is register-sensitive and characteristic of academic prose.

The most salient discourse function of both English and Norwegian topic identifiers is to focus on a (marked) theme by announcing it explicitly (Halliday and Matthiessen 2004: 67). Frequently, the identified theme represents a topic shift by reactivating a slumbering discourse participant or by navigating between two or more current topics. The same discourse functions of topic identification occur across languages and disciplines, and they are similar to those of left dislocation. However, the similarity depends on the presence of coreference between the identified topic and a rhematic element: when there is coreference, an element from the rhematic part of the T-unit is highlighted. Such constructions may be characterised as a variant of left dislocation which is acceptable in formal writing (Lambrecht 1994: 182). In cases of no coreference, the T-unit-internal highlighting does not occur, and the thematic adjunct functions in the same way as other thematic adjuncts: “[locating and orienting] the clause within its context” (Halliday and Matthiessen 2004: 64). The fact



that coreference occurs more often in Norwegian than in English suggests that the Norwegian construction is more akin to left dislocation than the English one, which is more commonly associated with either weak coreference or no coreference at all.

A further contrastive difference concerns the register-specificity and formality of the preferred topic marker. *In the case of* has a form generally favoured in academic writing (Biber and Gray 2016: 93) and shows great affinity with this register, while *når det gjelder* appears to be stylistically neutral. *When it comes to*, by contrast, belongs predominantly to spoken and journalistic registers and may be regarded as too informal for academic writing. Kranich (2016: 30) argues that academic writing is particularly interesting “for the contrastive study of communicative styles” due to the international character of this type of communication. Hence, it is noteworthy that English academic writing opts for a more formal style than Norwegian in its choice of topic identifiers. This observation ties in with previously observed differences, such as the more frequent use of first-person pronouns in Norwegian than in English academic writing (Fløttum et al. 2006: 157), including the use of *vi* (“we”) to include the reader as a collaborator in a joint activity (ibid.: 264), and the more extensive use in English than in Norwegian of lexical nominalisation, often as part of long and complex NPs (Nordrum 2007: 216). Norwegian thus appears to have a more colloquial communicative style than English in academic writing, although this is a hypothesis that requires further contrastive study.

In both languages, medicine was found to differ sharply from the other two disciplines in its sparse use of topic identifiers. This raises the question of whether medicine uses other focusing constructions, or simply abstains from them due to “a weaker need for medical writers than for linguists to exhibit text composition” (Fløttum et al. 2006: 262). Further studies of various thematisation and focusing devices, e.g. cleft constructions, fronting and focusing adverbials across languages and disciplines would contribute to a more exhaustive account of the ways in which academic writers construct their discourses. Preferably such studies should be based on larger corpora to get a fuller picture of low-frequency phenomena. Another avenue of research is the rhematic use of the expressions, as the frequencies reported in Figure 2 indicate that the cross-linguistic and cross-disciplinary differences there may be even greater. Finally, if it turns out that the register of academic writing is actually less distinct from other written registers in Norwegian than in English, as further contrastive discourse-pragmatic studies might ascertain, this will have implications for the teaching of academic writing in both English and Norwegian.

## References

- Aijmer, Karin. 1989. Themes and tails: The discourse functions of dislocated elements. *Nordic Journal of Linguistics* 12:2, 137–154.
- Biber, Douglas and Bethany Gray. 2016. *Grammatical Complexity in Academic English. Linguistic Change in Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: Cambridge University Press.
- Davies, Eirian C. 2017. The significance of theme. *English Text Construction* 10:2, 298–322.
- Dyvik, Helge. 2004. Translations as semantic mirrors. In: Karin Aijmer and Bengt Altenberg (eds.), *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*. Amsterdam/New York: Rodopi, 311–326.
- Faarlund, Jan Terje, Svein Lie, and Kjell Ivar Vannebo. 1997. *Norsk Referansegrammatikk*. Oslo: Universitetsforlaget.
- Fløttum, Kjersti, Trine Dahl, and Torodd Kinn. 2006. *Academic Voices*. Amsterdam: Benjamins.
- Fontaine, Lise. 2013. *Analysing English Grammar: A Systemic Functional Introduction*. Cambridge: Cambridge University Press.
- Geluykens, Ronald. 1992. *From Discourse Process to Grammatical Construction. On Left Dislocation in English*. Amsterdam/Philadelphia: Benjamins.
- Gómez-González, María. Ángeles. 2001. *The Theme–Topic Interface: Evidence from English*. Amsterdam/Philadelphia: Benjamins.
- Gosden, Hugh. 1992. Discourse functions of marked theme in scientific research articles. *English for Specific Purposes, Vol 11*, 207–224.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar. Second edition*. London: Arnold.
- Halliday, M.A.K. and Christian M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar. Third edition*. London: Routledge.
- Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Hasan, Ruqaiya. 1985. The texture of a text. In: M.A.K. Halliday and Ruqaiya Hasan, *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Geelong: Deakin University Press, 70–96.
- Hasselgård, Hilde. 2004. Thematic choice in English and Norwegian. *Functions of Language* 11:2, 187–212.
- Hasselgård, Hilde. 2005. Theme in Norwegian. In: Berge, Kjell Lars, and Eva Maagerø (eds.), *Semiotics from the North: Nordic Approaches to Systemic Functional Linguistics*. Oslo: Novus, 35–48.

- Hasselgård, Hilde. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press.
- Hasselgård, Hilde. 2018. Language contrasts, language learners, and metacognition. In: Åsta Haukås, Camilla Bjørke and Magne Dypedahl (eds.), *Metacognition in Language Learning and Teaching*. Routledge, 98–120.
- Hasselgård, Hilde. 2019. Phraseological teddy bears: frequent lexical bundles in academic writing by Norwegian learners and native speakers of English. In: Viola Wiegand and Michaela Mahlberg (eds.), *Corpus Linguistics, Context and Culture*. Berlin: De Gruyter, 339–362.
- Hasselgård, Hilde. Forthcoming. Corpus-based contrastive studies: Beginnings, developments and directions. To appear in *Languages in Contrast* 20:2.
- Holmes, Philip and Hans-Olav Enger. 2018. *Norwegian. A Comprehensive Grammar*. London and New York: Routledge.
- Hyland, Ken. 2012. *Disciplinary Identities. Individuality and Community in Academic Discourse*. Cambridge: Cambridge University Press.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora*. Amsterdam/Philadelphia: Benjamins.
- Johansson, Stig, Jarle Ebeling, and Signe Oksefjell. 1999/2002. English-Norwegian Parallel Corpus: Manual. University of Oslo. [www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/ENPCmanual.pdf](http://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/ENPCmanual.pdf).
- Kranich, Svenja. 2016. *Contrastive Pragmatics and Translation. Evaluation, Epistemic Modality and Communicative Styles in English and German*. Amsterdam/Philadelphia: Benjamins.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form. Topic, focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Nordrum, Lene. 2007. English Lexical Nominalizations in a Norwegian-Swedish Contrastive Perspective. PhD thesis, University of Gothenburg. <http://hdl.handle.net/2077/17181>.
- Prince, Ellen F. 1998. On the limits of syntax, with reference to left-dislocation and topicalization. In: Peter Culicover and Louise McNally (eds.), *The Limits of Syntax. Syntax and Semantics* 29. New York: Academic Press, 261–302.
- Swales, John M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska Akademiens Grammatik. 2, Ord*. Stockholm: Svenska Akademien.
- Thompson, Geoff. 2004. *Introducing Functional Grammar*. Second edition. London: Routledge.
- Thompson, Geoff and Susan Thompson. 2009. Theme, Subject and the unfolding of text. In: Gail Forey and Geoff Thompson (eds.), *Text Type and Texture*. London: Equinox, 45–69.

Verschueren, Jef. 2016. Contrastive pragmatics. In: Jan-Ola Östman and Jef Verschueren (eds.), *Handbook of Pragmatics Online*. Amsterdam/Philadelphia: Benjamins. <https://doi.org/10.1075/hop.20.com18>.

### Corpora Used

The Corpus for Bokmål Lexicography (“Leksikografisk bokmålskorpus”): [www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/](http://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/).

The Corpus of Contemporary American English (COCA): [www.english-corpora.org/coca/](http://www.english-corpora.org/coca/).

The English-Norwegian Parallel Corpus (ENPC): [www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/](http://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/)

The KIAP corpus (Cultural Identify in Academic Prose): [www.uib.no/fremmedsprak/23107/kiap-korpuset](http://www.uib.no/fremmedsprak/23107/kiap-korpuset).