

Granvik, Anton & Susanna Taimitarha. 2014. Topic-marking prepositions in Swedish: A corpus-based analysis of adpositional synonymy. *Nordic Journal of Linguistics* 37(2), 257–296.

Topic-marking prepositions in Swedish: A corpus-based analysis of adpositional synonymy

Anton Granvik & Susanna Taimitarha

This study analyses the relationship between four near-synonymous Swedish prepositions, namely *angående*, *beträffande*, *gällande* and *rörande*, which are used to establish what we call a topic-marking relation, as in *statens avtal angående finansieringen* ‘the agreement of the state regarding the financing’. By focusing on a single, loosely defined genre consisting of the written texts included in the Swedish PAROLE corpus, we address the question of what differences there are among these four prepositions, which intuitively seem highly similar and mutually interchangeable. In order to find out which contextual and grammatical factors might influence the choice of one preposition over the others, two complementary analyses were performed. First, a so-called collostructional analysis (see Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004) was performed on 791 cases of these prepositions found in the PAROLE corpus. Secondly, the corpus examples were annotated according to ten syntactic and four semantic criteria and a multinomial logistic regression analysis was performed on the annotated data set. The results show some tendencies pointing to differing usage patterns of the four prepositions. *Beträffande* stands out as the most frequent of them all and is also preferably used when no explicit head element is present, typically in sentence-initial position. *Angående* prefers words of communication while *rörande* is used when another topic-marking preposition is also present. On the other hand, neither of the two analyses leads to a clear distinction among the four prepositions, thus pointing to the fact that these topic-marking prepositions indeed constitute a fairly good case of adpositional synonymy, with few distinguishing factors separating one from the other.

Keywords collostruction analysis, corpus-based analysis, near-synonymy, quantitative analysis, Swedish prepositions

Anton Granvik, Centre for Languages and Business Communication, Hanken School of Economics, P.O. Box 479 (Arkadiagatan 22), FI-00101 Helsingfors, Finland. anton.granvik@hanken.fi
Susanna Taimitarha, Department of Finance and Statistics, Hanken School of Economics, P.O. Box 479 (Arkadiagatan 22), FI-00101 Helsingfors, Finland. susanna.taimitarha@hanken.fi

1. INTRODUCTION

Talking about something is probably one of the most basic linguistic activities. Hence, marking the topic of conversation is an essential part of the linguistic system in the

sense that most of what is said is usually said ‘in relation to, concerning or regarding’ a certain subject. This is what Halliday (1967:212) calls THEME, i.e. ‘theme is what is being talked about, the point of departure for the clause as a message’. In a similar sense, Langacker (2008:513) builds his characterization of TOPIC, defined as ‘what the sentence is about’, on what he calls the ‘notion of “aboutness”’.

The focus of this paper is precisely the notion of ‘aboutness’, but on a more specific, clause-internal level. What we call a (prepositional) topic-marking relation is the one established between, typically, certain verbs, nouns and adjectives denoting communication or mental processes and their ‘topic’, or ‘subject matter’ in the terminology of Quirk et al. (1985:709–710).¹ In a sentence such as *I enjoy talking/thinking about the future*, the *future* is the topic of discussion or thought, *about* is the topic-marking preposition (TMP), and *talk/think* are the head elements of the TMP, i.e. the words triggering the topic-marking relation.²

Languages typically have various means of expressing the topic-marking relation depending on context and communicative necessity. In English *about* is the default topic-marking preposition (Lindstromberg 2010:141), *om* being its Swedish equivalent, as in *tala om framtiden* ‘to talk about the future’. Beside unmarked *om*, Swedish has four originally participial expressions, *angående*, *beträffande*, *gällande* and *rörande* that can be considered marked topic-marking prepositions in much the same way as English *on*, *concerning* and *regarding* are marked in comparison with default *about*.³ Furthermore, there are several other expressions that can be used to establish a topic-marking relation in Swedish. Alongside *om* and the four TMPs *angående*, *beträffande*, *rörande* and *gällande*, the spatial preposition *kring* ‘around’ is frequently used, and there are also more complex expressions such as *vad X angår/beträffar/gäller* or, alternatively, *vad angår/beträffar/gäller X* (roughly equivalent to ‘when it comes to X’ or ‘with regard to X’), where X stands for the topic.

In this paper, we shall focus only on *angående*, *beträffande*, *rörande* and *gällande*, which share a number of characteristics making them a good case for investigating the issue of adpositional synonymy. Apart from belonging to the conceptual domain of topic-marking expressions, *angående*, *beträffande*, *gällande* and *rörande* are all originally present participle forms on the verge of becoming lexicalized as prepositions (*SAG:Particip* §42; *Prepositioner* §28, §38). Furthermore, they bear connotations of a formal, bureaucratic and legal register, and their use is widely discouraged as ‘rigid and formal’ among language planning professionals (*Ssb:s.v. beträffande*; *SLAF*:83, 127, 135). As example (1) shows, the interchangeability of these four expressions is uncontroversial in a context such as this one, where even *om* is possible without a significant change in meaning:

- (1) Till följd av statens avtal med Landstingsförbundet *beträffande* [*angående/gällande/rörande/om*] finansieringen av länsmusiken är det

rimligt att (med tanke på uppsägningstid m.m.) tillämpa det nuvarande systemet under 1997.

‘As a consequence of the agreement with Landstingsförbundet *regarding* the financing of regional music it is reasonable to follow the present legislation during 1997 (when it comes to let-off times, etc.).’

What is the difference between using one of the four participle prepositions in the topic relation as compared default *om*? And, related to this question, what are the differences between the four?

The first question can be answered, initially and in a simplified manner, by referring to stylistic reasons, since the four participle TMPs that concern us here belong to the formal register and are often thought of as an indication of complicated legal language, on the one hand, and as a symbol of education and high style, on the other. Another difference is that *angående*, *beträffande*, *gällande* and *rörande* are numerically quite rare – with a total of 929 prepositional cases in the PAROLE corpus, compared to 108,427 cases of prepositional *om*. Furthermore, whereas *om* can stand alone, i.e. as a so-called stranded preposition, *angående*, *beträffande*, *gällande* and *rörande* cannot (SAG:Prep. §35).

For the reasons presented above we will not address the first question in this study, but instead concentrate on the more delicate issue of what differences there are between the four participle-based prepositions. Specifically, are the expressions synonymous or not? If they are, to what degree? What differences in meaning and usage can be identified?

Underlying these questions is the assumption that there are differences in the usage of the four TMPs. This assumption rests on the idea that changing the expression also implies a change in meaning, i.e. the principle of ‘one form, one meaning’. As Gilquin (2010:97) states,

this ‘one form, one meaning’ principle has been adopted in several theoretical frameworks, most notably in functional grammar and cognitive linguistics [where] it has been expressed among others by Goldberg (1995:67) and her ‘Principle of No Synonymy’.

According to this principle ‘[i]f two constructions are syntactically distinct, they must be semantically or pragmatically distinct’ (Goldberg 1995:67).

Langacker (1987:110–113, 138–141; 2008:43) builds his notions of IMAGERY and CONSTRUAL on this principle of ‘one form, one meaning’, i.e. ‘[t]he term construal refers to our manifest ability to conceive and portray the same situation in alternate ways’. At the lexical level, the ‘one form, one meaning’ principle has led to the introduction of the term NEAR-SYNONYMY (see Taylor 2002; Vanhatalo 2002, 2003; Arppe & Järvikivi 2007:8; Desagulier 2014). From Langacker’s construal perspective,

synonymous expressions have identical conceptual content and impose the same construal upon that conceptual content, while near-synonyms share the same conceptual content but differ in terms of construal. (Desagulier 2014:145–146).

As Taylor (2002:265–266) states, in summarizing the opinions of earlier theoretical considerations regarding synonymy,

[i]f, depending on one's viewpoint, perfect synonymy is vanishingly rare, methodologically proscribed, or a logical impossibility, what we frequently do encounter are pairs of words that are 'near' synonyms. A characteristic of near synonyms, as opposed to words which are merely similar in meaning, is that near synonyms exhibit 'a low degree of implicit contrastiveness'. (Cruse 1986:266)

From the perspective of the TMPs under focus in this study, the question of synonymy is, in principle, a lexical one. Four near-synonymous prepositions are contrasted. However, due to the fact that *angående*, *beträffande*, *gällande* and *rörande* are used as prepositions, their syntactic behaviour is far more constrained than that of nouns, verbs or adjectives, which have been the subject of most previous studies on lexical synonymy (see Taylor 2002; Vanhatalo 2002, 2003; Divjak & Gries 2006; Arppe & Järviö 2007; Arppe 2008; Janda & Solovyev 2013; Glynn 2010, 2014; Gries & Otani 2010; Liu 2010, 2013; Desagulier 2014). Hence, it is not easy to see how the following consideration made by Divjak & Gries (2006:24) applies to the TMPs in question: 'no two words ever are exact synonyms, but instead always differ from a syntactic, semantic and/or pragmatic point of view'. Langacker's notion of CONSTRUAL, does not improve the situation significantly, because the inherently abstract notion of 'aboutness' is not evidently 'construed' differently in terms of *angående* as opposed to *beträffande*, and so on.⁴

As the title of this paper indicates, we have to deal with what we refer to as adpositional synonymy.⁵ This notion seems closely related to Klavan's (2012) work on GRAMMATICAL SYNONYMY. However, in comparing two Estonian locative constructions, the adessive case and adpositional *peal* 'on', Klavan's grammatical synonymy bears reminiscence of the extensive work on syntactic and grammatical alternations, as exemplified by Goldberg's (1995) work on the English so-called 'dative alternation' from a Construction Grammar perspective (see Gries & Stefanowitsch 2004). Our prepositional topic-markers, as exemplified in (1), however, do not instantiate distinct (grammatical) constructions, but can rather be seen as belonging to a single schematic topic-marking construction, in which both the topic-marking preposition and the topic-instantiating head elements can vary.

Under these conditions, then, we have adopted a broad working definition of the notion of synonymy, such as the one presented by Glynn (2014:10), who considers

that synonymy is, quite simply, the ‘formal variation of a concept-function’. This means that we view (near)synonymy not only as a ‘low degree of contrastiveness’ in terms of meaning, but also in terms of differences in ‘formal variation’, as might be evidenced in distinct collocation patterns for the four TMPs.

Having said this, we can introduce a couple more technical objectives for the study. First, what can a corpus-based analysis of mainly one linguistic register (namely the formal, written language; see Section 2) actually tell us about adpositional synonymy? Second, we will also discuss other information needed in order to reach a ‘full’ account of this question, envisaging directions for future research.

In the next section, we present the main points of interest for the realization of the study, introducing the corpus and the methods used for analysing the data. Section 3 focuses on the data, including an analysis of the four synonyms first from a collocation perspective (Section 3.1) and then from a statistical modelling perspective (Section 3.2). The main focus of Section 3 lies on the comparative analysis, i.e. we comment on the differences and similarities across the four TMPs as they are highlighted by the different analyses. Section 4 provides a discussion of the main findings of the study as well as a critical evaluation of the methodology used. In Section 5 we wrap it all up by offering some concluding remarks and discussing the issue of adpositional synonymy from the perspective of this study.

2. CORPUS AND METHODS

Given that the centre of attention of this paper are four prepositions which are known to belong in a formal, high-style register, focusing on a written corpus seems motivated. By excluding the potentially differentiating factor of genre and limiting the analysis to the use of *angående*, *beträffande*, *gällande* and *rörande* within a single genre, our main interest lies on specific lexicogrammatical differences between them.

With this in mind, we extracted all prepositional uses of *angående* (N = 187), *beträffande* (N = 466), *gällande* (N = 77) and *rörande* (N = 199) in the Swedish PAROLE corpus. This is a corpus of written contemporary Swedish, consisting of 24,331,936 tokens constituting 1,646,688 sentences. The material was collected within the PAROLE project during the 1990s, and it includes novels, newspapers, magazines and web texts. The PAROLE corpus is freely available on the internet at the Språkbanken web page, where it is included in the Korp corpus search interface (<http://spraakbanken.gu.se/korp/>). The corpus is tagged for syntactic features and the word class information was used in calculating overall frequencies for the words with which *angående*, *beträffande*, *gällande* and *rörande* co-occur.

The searches were made using the full lemma form of each TMP. A total of 1,719 cases were found, 929 of which are topic-marking prepositions (TMP) as

defined above. This means that almost half the retrieved cases had to be discarded. In the case of *angående* and *beträffande*, which function exclusively as prepositions, only a dozen or so examples that are repeated in the corpus had to be removed. For *angående* the original search provides 196 cases, but the actual total is 187 distinct examples, and for *beträffande* the corresponding numbers are 472 and 466. The search word *gällande* yields a total of 692 hits, a number that is reduced to 686 by eliminating the doublets. However, only 77 of these cases are prepositional uses. The most frequent use of *gällande* is as an adjective, as in *gällande regler* ‘current rules’ (N = 335), but the form also occurs frequently as a verb particle (N = 267), mainly in the expression *göra (sig) gällande* ‘to show/enforce (one/itself)’. Finally, seven cases of *gällande* correspond to verbal uses of the present participle. *Rörande* also presents other than prepositional uses. Out of a total of 365 hits, 199 are prepositional uses, while the rest are adjectival, as in *en rörande vers* ‘a touching (i.e. “emotional”) verse’.

Since we were interested in finding out to what extent context has an impact on the choice of one synonym over another, the data set was annotated using a number of factors. Prepositions typically occur as a nexus between two lexemes, so one of the fundamental differentiations that have to be made is to determine the word class of the head (prepositional trajector) and the governed (prepositional landmark) element. After making this initial distinction, the examples were annotated according to ten mainly syntactic characteristics of the usage context. When the head and governed elements were nouns, they were annotated for number, definiteness (two classes), modification (two classes) and complexity. When the syntactic head of the TMP was a verb, the formal attributes transitivity, tense, voice, and subject complexity were used.

The head element was also annotated using four semantic criteria, including semantic word class, agent animacy, abstract/figurative use, and uniqueness (see Thuilier & Danlos 2012).⁶ The semantic word class annotation was performed by both authors in order to minimize the influence of subjective judgements on the results. A total of five classes were included, representing COMMUNICATION, COGNITIVE PROCESSES, ACTIONS, INFORMATION SOURCE and GENERAL WORDS (see examples (11) to (15) in Section 3.1 below). Beside the residual class of general words, the information source class is worth a comment, since it comprises words, typically nouns, which are conceptually quite close to the communication words. However, nouns such as *uppgifter* ‘data, information’, *avtal* ‘agreement’, and *rapport* ‘report’, do not really communicate, but rather contain some sort of information, hence the label. The agent animacy classes, on the other hand, are more straightforward, including HUMAN, COLLECTIVE, OBJECTS (i.e. inanimate) and UNSPECIFIED subjects/agents. The uniqueness factor makes reference to the presence or absence of other topic-marking prepositions in the clause.

Nr	Ag	anim	uniq	fig	wc	H_lex	HEAD	H_num	H_def	H_det	H-modif	H_attr	H_compl	Marker	GOV	G_num	G_def	G_det	G_modif	G_attr	G_compl	
28	1	0	0	4		uppgifter	1	2	0	0	1	0	0	1	angående	1	1	1	0	0	1	1
242	1	1	0	1		synpunkter	1	2	0	0	0	1	0	0	beträffande	1	1	1	0	0	0	0
692	3	1	0	0		förslag	1	2	0	0	0	0	0	0	rörande	1	1	1	0	0	0	0
899	1	1	0	4		avtal	1	1	0	1	0	0	0	1	gällande	1	1	0	0	1	1	2

Figure 1. Excerpt from the annotated corpus data.

Figure 1 includes an excerpt of the annotated material including one line per TMP, the specifics of which are exemplified and explained in Appendix A. The data lines in Figure 1 correspond to the corpus sentences (2)–(5):

- (2) Det står nu utom allt tvivel att Autonova systematiskt har lämnat falska uppgifter inför koncessionsnämnden och regeringen *angående* innehållet av lösningsmedel i de vattenbaserade lacker man ska använda, skriver Björn Gillberg till miljödepartementet.

‘It is beyond doubt that Autonova has systematically been giving false information to the concession board and the government *regarding* the solvents included in the water based lacquers that are to be used, writes Björn Gillberg to the Ministry of Environment.’

- (3) Det hedrar konstitutionsutskottets ledamöter att de inte gick på finansminister Göran Perssons linje att köra över lagrådets synpunkter *beträffande* förslaget till nytt inkomst- och kostnadsutjämningsystem för kommunerna.

‘It does the deputies of the constitutional committee credit that they did not follow Finance Minister Göran Persson in overriding the considerations of the Council of Legislation *regarding* the proposed new municipal income and cost levelling system.’

- (4) Senare modifierade Winberg sina uttalanden till att det inte spelade någon roll om förslag *rörande* sysselsättningen kom under våren eller först till hösten.

‘Later on, Winberg modified his/her comments saying that it did not matter whether suggestions *regarding* employment were presented during the spring or only by autumn.’

- (5) Hur långt skall vi acceptera fackföreningarnas stridsåtgärder för att tvinga en oorganiserad arbetsgivare att skriva under ett avtal med facket *gällande* icke fackansluten personal mot deras vilja?

‘To what extent shall we accept the combat measures of the labour unions forcing an unorganized employer to sign an agreement with the union *regarding* non-unionist personnel against their will?’

The rationale behind the detailed annotation of our data is the Firthian idea that a word gets its meaning by the company it keeps (Firth 1968:179). That is, by assuming that the immediate syntactic context can predict the choice of synonym, we are interested in finding out which of a range of possible contextual factors are important for determining which TMP is used. An effective method that can be used

to test the relative contribution of a range of factors is logistic regression analysis (see Glynn 2010, Klavan 2012). As the dependent variable in this case is categorical and has four members (i.e. *angående*, *beträffande*, *gällande* and *rörande*), a multinomial version of the logistic regression analysis was used. What the multinomial logistic regression analysis (MLRA) does is to create a model that describes which of the included factors can be used to distinguish between the four TMPs. The advantage of the MLRA is that it simultaneously takes into account the impact of all the factors included in the data, determining their relative contribution to the model as a whole. It also reveals which factors are significant for a particular TMP and builds up a general classification model based on the data. The prediction capacity of the model provides an insight into how well the annotated factors predict the usage of the different TMPs.

In addition to the considerable detail the MLRA goes into by taking into account all the annotated factors, we also decided to look at context on a purely superficial level, i.e. the lexical co-occurrence of the TMPs with different head elements. The main objective was to find out whether there are frequent collocates or collexemes for any of the four TMPs, following the idea that word combinations actually constitute meaningful patterns of their own (Croft 2001:14–29; Croft & Cruse 2008:323). With this in mind, we turned to the collostructional and distinctive collexeme analyses (see Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004, Schmid 2010, Schmid & Küchenhoff 2013). This means considering each combination of a head element + TMP as a (sub-)construction of its own. The four sub-constructions thus analysed can then be compared to reveal possible differences and similarities.⁷

Following Schmid & Küchenhoff (2013:555), we decided to rank the collexemes of *angående*, *beträffande*, *gällande* and *rörande* according to their relative Odds Ratio (OR) ratings (see Table 4 in Section 3.1 below). In order to compute the Odds Ratio value for a given collocation, four values are needed, which can be expressed in a conventional contingency table, as shown in Table 1 for *angående* combined with its most frequent collexeme, the noun *fråga* ‘question’. As the table shows, the values of interest are the frequencies of the *angående* sub-construction, and the frequencies of one of its collexemes, *fråga*. The figures marked by boldface stem directly from the PAROLE corpus, while the others are calculated.

The OR value is achieved by making three calculations, using the values in the shaded area in Table 1. First, the relative frequency of the collexeme in question, *fråga*, is divided by the frequency of other words in the same construction, i.e. combined with *angående*, giving a score of $8/179 = 4,5\%$ (0.0447).⁸ The second relative frequency is obtained by dividing the remaining frequency of *fråga*, i.e. its use outside the *angående* construction (16,592 cases), with the frequency of all other prepositional constructions with any head other than *fråga* (2,540,574 cases). Finally, the first frequency is divided by the second frequency, giving the OR value,

	<i>angående</i>	\neg <i>angående</i>	Sum rows
<i>Fråga</i>	8	16,592	16,600
\neg <i>fråga</i>	179	2,540,574	2,540,753
Sum columns	187	2,557,166	2,557,353

Note: The negation sign in this table indicates, very simply, other elements than *fråga* (\neg *fråga*) and other prepositions than *angående* (\neg *angående*). The sum total of this table stands for all prepositions in the corpus, whereas 16,600 is the token frequency of the noun *fråga* and 187 the token frequency of *angående*. Since the relationship between the head element and the TMP is semantically and syntactically constrained, the collocation of the two elements is of course not limited to their linear combination.

Table 1. Contingency table for the collocation *fråga angående* ‘question regarding’ in the PAROLE corpus.

which for *fråga* and *angående* is 6.8.⁹ A positive side of the OR score is that it is directly interpretable. Thus, a score of 6.8 indicates that finding the word *fråga* in a Swedish text implies that it is almost seven times more likely for it to appear in combination with *angående* compared to any other Swedish word (Schmid & Küchenhoff 2013:554).

3. ANALYSIS

In this section the main results of the two analyses are presented. In Section 3.1 we go through the outcome of the collocation analysis, focusing on the different prepositions in order to set the basis for their comparison. In Section 3.2 we look into the results that emerge from the multinomial logistic regression analysis.

Before proceeding with the analyses it is worth taking a look at some general characteristics of the data. Tables 2 and 3 present the main distributional data regarding the syntactic word class of the head and governed elements combined with the four TMPs. As Table 2 shows, the governed elements (GOV) are very similar for all the four TMPs; *angående*, *beträffande*, *gällande* and *rörande* all combine preferably with nouns (N) (over 90% of the governed elements), and to a much lesser extent with proper nouns and pronouns.

GOV	<i>angående</i>	<i>beträffande</i>	<i>gällande</i>	<i>rörande</i>	mean
Noun	93%	90%	91%	93%	92%
Proper noun	5%	7%	6%	4%	6%
Pronoun	2%	2%	3%	3%	3%
Sum	100%	100%	100%	100%	100%
N	187	466	77	199	929

Table 2. Characteristics of governed element (GOV) ($p = .77$, $\chi^2 = 3.299$, $df = 6$).

HEAD	<i>angående</i>	<i>beträffande</i>	<i>gällande</i>	<i>rörande</i>	Mean
Adjective	3%	8%	1%	0%	3%
Noun	77%	48%	91%	96%	78%
Verb	20%	15%	1%	3%	6%
No head	0%	29%	6%	2%	9%
Sum	100%	100%	100%	100%	100%
N	187	466	77	199	929

Table 3. Characteristics of head element (HEAD) ($p < .001$, $\chi^2 = 228.95$, $df = 9$).

The numbers in Table 3, on the other hand, show significant differences among the four TMPs with regard to the syntactic category of the head element (HEAD). *Angående*, for example, is never used without a head, while up to 29% of the uses of *beträffande* lack a clear syntactic head. *Rörande* is never combined with adjectives, and together with *gällande* is almost exclusively used with nominal heads (96% and 91%, respectively). And, although *beträffande* exhibits the greatest variety of usage contexts and is the most frequent of all the TMPs, *angående* shows a higher relative frequency of combination with verbal heads.

These initial data seem to indicate a greater contribution to the overall differences and similarities in usage in the characteristics of the head elements (the prepositional trajector) than in the governed elements (landmark), which are highly uniform. This makes perfect sense, as it is natural to assume that what influences the choice of topic marker is not the topic itself, i.e. the governed element that indicates what is talked about, but rather the specific word or expression triggering the topic-marking relation, i.e. the head element. As we shall see, this initial hypothesis is largely confirmed in the logistic regression analysis, where no factor concerning the governed element was found to be significant in the regression model.

Before continuing a further observation needs to be made. Based on the data presented in Tables 2 and 3, the rest of our analyses will focus mainly on the different head elements of the topic-marking construction. This has an important consequence which might not be immediately apparent, and which affects the number of cases analysed. Since a total of 140 of the identified topic-marking prepositions in our data are analysed as lacking an explicit head element (133 cases of *beträffande*, five of *gällande* and two of *rörande*), these cannot be included in the analyses because they carry no information. Thus, eliminating the no-head cases from the following analyses means that our focus lies on ever more specific contextual factors. This is especially relevant to *beträffande*, whose use without a head (with 95% of the 140 cases) already clearly distinguishes it from the other TMPs.

3.1 Collostructional analysis

The main idea in carrying out a collostructional analysis is to find out which words are most salient for a given construction. For reasons presented above, the following collostructional analysis concentrates on the combination of different head elements with the four TMPs. In order to keep the presentation at a manageable size, Table 4 only includes the heads occurring at least twice with *angående*, *beträffande*, *gällande* and *rörande*. The collexemes are ranked according to their OR-score, which correlates very strongly with the Reliance (Rel) score.¹⁰ The Attraction score, on the other hand, correlates with absolute frequency, but neither has been taken into account in Table 4.¹¹

As the data show, most of the top ranked lexemes (those with an OR score of over 100, for example) are unique for each TMP. This might lead one to consider that they form some sort of formulaic sequence together with the TMP in questions. However, this is hardly the case since the total number of co-occurrences is very small, as the very low Attraction and Reliance scores indicate. Also, judging by examples (6)–(10) the contexts in which these high scoring collocations occur are not so specific as to lead one to think that there is a particular attraction between any two of them.

- (6) Redan förra året fick det svenska justitiedepartementet en förfrågan från EU-kommissionen, *angående* vilka regler som gällde för flaggföring och fartygsregistrering. (*förfrågan*, N = 154, two occurrences with *angående*)
‘Already last year the Swedish Department of Justice received an enquiry from the EU Commission *regarding* the rules that apply to flagging and ship registration.’
- (7) Vart en sådan attityd till sist leder har vi dock under den senaste veckan kunnat bevittna: först helomvändningen *beträffande* den höjda fastighetsskatten, sedan anpassningen till folkpartiets riktlinjer för den kommunala skatteutjämningen. (*helomvändning*, N = 126, two occurrences with *beträffande*)
‘Where such an attitude leads is something we have been able to witness in the past week: first the full-turn *regarding* the raise in property tax, then the adaptation to the guidelines of the People’s Party for municipal tax equilibration.’
- (8) Vidare slöts en andra överenskommelse, *gällande* EU-importen av ris och vete. (*överenskommelse*, N = 851, three occurrences with *gällande*)
‘There was a further agreement *regarding* the EU import of rice and wheat.’
- (9) Per-Erik Nilsson säger att det fanns flera tveksamheter *gällande* ansvarsfördelningen efter mordet. (*tveksamhet*, N = 179, two occurrences with *gällande*)
‘Per-Erik Nilsson says that there were many doubts *regarding* the division of responsibilities after the murder.’

<i>angående</i>	Freq	Attr	Rel	OR	<i>beträffande</i>	Freq	Attr	Rel	OR	<i>gällande</i>	Freq	Attr	Rel	OR	<i>rörande</i>	Freq	Attr	Rel	OR
förfrågan	2	1.07%	1.30%	181.9	gå från ord till handling	2	0.60%	7.14%	590.7	tvksamhet	2	2.78%	1.14%	419.9	dokument	7	3.55%	1.02%	139.3
ringa upp	2	1.07%	1.27%	178.3	helomvändning	2	0.60%	5.00%	404.1	överenskommelse	3	4.17%	0.35%	131.1	anmälan	2	1.02%	0.82%	108.8
klagomål	3	1.60%	1.17%	164.8	oklarhet	2	0.60%	1.59%	123.8	avtal	3	4.17%	0.09%	34.2	påverkan	2	1.02%	0.79%	104.1
data	3	1.60%	0.79%	111.2	tänka om	2	0.60%	1.55%	120.9	beslut	4	5.56%	0.05%	20.0	betänkande	2	1.02%	0.72%	95.4
bestämmelse	3	1.60%	0.66%	91.8	pessimistisk	2	0.60%	1.30%	101.0	plan	2	2.78%	0.05%	19.0	upptäckt	2	1.02%	0.55%	72.8
förhör	3	1.60%	0.62%	86.5	frågeställning	2	0.60%	1.10%	85.3	verksamhet	2	2.78%	0.04%	13.6	tvist	2	1.02%	0.53%	70.1
optimistisk	3	1.60%	0.60%	83.9	oenighet	2	0.60%	0.99%	76.4	förslag	2	2.78%	0.02%	8.8	granskning	2	1.02%	0.36%	46.8
yttrande	2	1.07%	0.50%	68.9	optimistisk	4	1.19%	0.80%	62.3	mål	2	2.78%	0.02%	8.0	ärende	4	2.03%	0.26%	34.7
detalj	2	1.07%	0.43%	59.2	förbättring	2	0.60%	0.78%	60.0	fråga	3	4.17%	0.02%	6.7	handling	4	2.03%	0.25%	33.1
tilfråga	2	1.07%	0.39%	54.3	förundersökning	2	0.60%	0.68%	52.4						regel	2	1.02%	0.24%	31.2
motion	2	1.07%	0.34%	47.2	upplysning	2	0.60%	0.56%	43.1						<i>lagstiftning</i>	2	1.02%	0.22%	29.4
tvivel	2	1.07%	0.32%	44.8	enighet	2	0.60%	0.53%	41.3						analys	3	1.52%	0.21%	28.1
regler	2	1.07%	0.24%	32.9	kriterium	2	0.60%	0.51%	39.6						<i>forskning</i>	5	2.54%	0.21%	28.0
presskonferens	2	1.07%	0.23%	31.7	synpunkt	5	1.49%	0.48%	37.0						prognos	2	1.02%	0.20%	26.2
lagstiftning	2	1.07%	0.22%	31.0	detalj	2	0.60%	0.43%	32.9						<i>utredning</i>	6	3.05%	0.18%	24.6
samarbeta	2	1.07%	0.20%	27.6	regel	3	0.90%	0.36%	27.6						<i>förhandling</i>	3	1.52%	0.14%	18.0
besked	5	2.67%	0.19%	26.7	tvivel	2	0.60%	0.32%	24.9						avtal	4	2.03%	0.12%	16.3
synpunkt	2	1.07%	0.19%	26.3	prognos	3	0.90%	0.30%	23.1						förslag	8	4.06%	0.10%	13.0
Slutsats	2	1.07%	0.16%	21.9	ändring	2	0.60%	0.30%	22.8						fråga	15	7.61%	0.09%	12.6
agera	2	1.07%	0.16%	21.7	insikt	2	0.60%	0.26%	19.6						information	2	1.02%	0.09%	11.7
uttalande	2	1.07%	0.15%	20.5	argument	2	0.60%	0.25%	19.2						föreställning	2	1.02%	0.09%	11.5
rapport	2	1.07%	0.14%	19.8	osakerhet	2	0.60%	0.25%	19.1						artikel	2	1.02%	0.08%	11.0
information	3	1.60%	0.13%	18.6	initiativ	2	0.60%	0.23%	18.0						studier	2	1.02%	0.08%	11.0
oro	2	1.07%	0.11%	14.7	tvksam	2	0.60%	0.22%	16.7						beslut	6	3.05%	0.08%	10.7

Table 4. OR-ranking of head collexemes for *angående*, *beträffande*, *gällande* and *rörande*.

<i>angående</i>	Freq	Attr	Rel	OR	<i>beträffande</i>	Freq	Attr	Rel	OR	<i>gällande</i>	Freq	Attr	Rel	OR	<i>rörande</i>	Freq	Attr	Rel	OR
<i>utredning</i>	3	1.60%	0.09%	12.7	<i>slutsats</i>	2	0.60%	0.16%	12.2						diskussion	2	1.02%	0.08%	10.0
<i>riåd</i>	3	1.60%	0.09%	12.6	<i>inflytande</i>	2	0.60%	0.14%	11.0						besked	2	1.02%	0.08%	10.0
<i>förhandling</i>	2	1.07%	0.09%	12.6	information	2	0.60%	0.09%	6.8						<i>undersökning</i>	2	1.02%	0.07%	9.6
<i>brev</i>	3	1.60%	0.09%	12.1	<i>krav</i>	5	1.49%	0.09%	6.8						<i>åtgärd</i>	2	1.02%	0.07%	9.3
<i>forskning</i>	2	1.07%	0.08%	11.6	studie	2	0.60%	0.08%	6.4						<i>program</i>	3	1.52%	0.07%	8.9
studie	2	1.07%	0.08%	11.6	<i>påpecka</i>	2	0.60%	0.08%	6.2						<i>brev</i>	2	1.02%	0.06%	7.6
<i>uppdrag</i>	2	1.07%	0.08%	10.7	diskussion	2	0.60%	0.08%	5.9						<i>debatt</i>	2	1.02%	0.05%	7.0
diskussion	2	1.07%	0.08%	10.5	besked	2	0.60%	0.08%	5.8						<i>samarbete</i>	2	1.02%	0.05%	6.5
<i>svar</i>	3	1.60%	0.07%	10.0	beslut	4	1.19%	0.05%	4.1						<i>lag</i>	2	1.02%	0.05%	6.1
beslut	5	2.67%	0.07%	9.3	<i>kritik</i>	2	0.60%	0.05%	3.9						<i>arbete</i>	3	1.52%	0.03%	4.1
fråga	8	4.28%	0.05%	6.8	förslag	4	1.19%	0.05%	3.7						<i>uppgift</i>	2	1.02%	0.03%	3.9
<i>tanke</i>	2	1.07%	0.03%	4.3	<i>gälla</i>	5	1.49%	0.03%	2.5						<i>problem</i>	2	1.02%	0.02%	2.6
förslag	2	1.07%	0.02%	3.3	fråga	4	1.19%	0.02%	1.8										
<i>skriva</i>	2	1.07%	0.01%	1.9	<i>göra</i>	3	0.90%	0.00%	0.4										
<i>säga</i>	2	1.07%	0.00%	0.3															

Note: The values used for calculating the different values are *angående* 187, *beträffande* 335, *gällande* 72, *rörande* 197. The numerical differences compared to Tables 2 and 3 are due to the lack of syntactic heads for some of the TMPs. The total number of constructions used is 2,557,353, which corresponds to all prepositions in the PAROLE-corpus (see Table 1). Nouns co-occurring with three or four of the TMPs are marked with boldface, italics indicate nouns co-occurring with two of the TMPs. The abbreviations used are the following: Freq = frequency, Attr = Attraction value, Rel = Reliance value, OR = Odds Ratio value.

Table 4. Continued.

- (10) Hittills har Skolverkets tillsynsinsatser huvudsakligen gällt anmälningar rörande enskilda elevers situation. (*anmälning*, N = 243, two occurrences with *rörande*)

‘Up until now the supervision efforts of the Swedish National Agency for Education have primarily concerned reports regarding the situation of particular students.’

In addition to the top-scoring collexemes in Table 4 the collostructional analysis brings up a couple of further issues. First, it needs to be pointed out that the head elements in Table 4 that are marked with boldface co-occur with at least three of the four TMPs. There are eight of them in total, all nouns, and only three of them co-occur with all four TMPs: *beslut* ‘decision’, *fråga* ‘question’ and *förslag* ‘proposal’. The remaining five, which are not found with *gällande*, are *besked* ‘notification’, *diskussion* ‘discussion’, *information* ‘information’, *regel* ‘rule’, and *studie* ‘study’. The words marked with italics co-occur with two of the TMPs: *avtal* ‘agreement’, *brev* ‘letter’, *detalj* ‘detail’, *forskning* ‘research’, *förhandling* ‘negotiation’, *lagstiftning* ‘law making’, *optimistisk* ‘optimistic’, *prognos* ‘prognosis’, *slutsats* ‘conclusion’, *synpunkt* ‘opinion’, *tvivel* ‘doubt’, and *utredning* ‘investigation’. This total of 19 nouns and one adjective co-occurring with two or more of the four TMPs make up 51 collexemes out of a total 791 cases, which amounts to only six percent of all collexemes. This can be interpreted as an indication that there seem to be no specific collocations for any of the TMPs, or even for the more schematic topic-marking construction.¹²

Second, most of the shared head elements are situated in the lower half of Table 4, meaning that they are ranked low in terms of occurring in the topic relation with any one of the TMPs in question. Only the adjective *optimistisk* ‘optimistic’ scores relatively high on the OR scale for *angående* and *beträffande*, and, interestingly, for *beträffande* its antonym, *pessimistisk* ‘pessimistic’ also scores quite high. *Optimistisk* seems to be of some importance also from an Attraction perspective, since for both TMPs it ranks among the top five in terms of Attr score (see Appendix B), but its overall frequency of seven occurrences in the four topic-marking sub-constructions is hardly enough for it to constitute a formulaic expression.

Third, the shared collexemes are fairly evenly distributed over the TMPs. For all four, they account for roughly 40% of the total amount of collexemes: out of a total of 39 collexemes for *angående*, 17 co-occur with other TMPs as well, amounting to 44%. For *gällande* the relationship is 4 to 9 (i.e. 44%), for *rörande* 15 to 36 (i.e. 42%); only *beträffande* shows a slightly lower dependence on shared collexemes, with 13 out of 38 (34%) being shared. This might be interpreted as indicating that *beträffande*, as the most frequent of the four TMPs, is the default alternative of them all, since it does not pattern up with any particular head.

In addition to the above, we wish to highlight one further aspect of the annotation of the corpus that is closely related to the head elements, namely their semantic class. As Table 4 shows, many of the collexemes represent words of communication or mental processes, which is of course to be expected. However, upon a more careful analysis, five different semantic classes can be distinguished, namely words that refer to Communication, Cognitive processes, Actions, Information sources, and General words, as exemplified in (11) to (15):

(11) *Communication*

Det gäller att välja sina slagfält, skriver Dala-Demokraten (s) *angående* den kommande partikongressen.

‘It is a question of choosing one’s battlefields, writes Dala-Demokraten (social democrat) *regarding* the upcoming party congress.’

(12) *Cognitive processes*

Inte ens Banque de France, annars hyperkänslig *beträffande* inflationssiffrorna, väntas knorra över det resultatet.

‘Not even Banque de France, otherwise overly sensitive *regarding* inflation figures, is expected to frown at this result.’

(13) *Actions*

Det rättsliga efterspelet *gällande* två andra ex-ministrar, Mona Sahlin (s) och Reidunn Laurén (opol) har kommit så långt att det i dag är dags för de första förhören.

‘The judicial aftermath *concerning* two other ex-ministers, Mona Sahlin (social democrat) och Reidunn Laurén (unpolitical) has reached the point that the first hearings will take place today.’

(14) *Information source*

Det konstaterar Finansinspektionen i en rapport till regeringen “*rörande* värdepappersfondernas avgifter och informationsgivning”.

‘This is what Finansinspektionen agency states in a report to the government “*regarding* the fees and information of investment funds”.’

(15) *General*

Datoranvändningen i svenska skolor varierar *beträffande* omfattning och innehåll såväl mellan som inom skolor.

‘The use of computers in Swedish schools varies with regard to the extent to and the way in which they are used across schools as well as within schools.’

Tables 5 and 6 show the distribution of these five semantic classes across the four TMPs. As the tables show, on the level of semantic classes there are some differences between the four TMPs. Following the figures in Table 5, *angående* most frequently co-occurs with heads belonging to the communication domain. *Beträffande* and *gällande* are quite evenly divided between Cognitive processes and Actions, while *rörande* is biased towards Actions and Information sources. From

			wc					
			Communication	Cognitive process	General	Action	Information	Total
Marker	<i>Angående</i>	N	76	32	4	38	37	187
		%	40.6%	17.1%	2.1%	20.3%	19.8%	100%
	<i>Beträffande</i>	N	57	93	70	76	37	333
		%	17.1%	27.9%	21.0%	22.8%	11.1%	100%
	<i>Gällande</i>	N	11	18	11	17	13	70
		%	15.7%	25.7%	15.7%	24.3%	18.6%	100%
	<i>Rörande</i>	N	33	40	20	49	55	197
		%	16.8%	20.3%	10.2%	24.9%	27.9%	100%
Total	N		177	183	105	180	142	787
	%		22.5%	23.3%	13.3%	22.9%	18.0%	100%

Table 5. Crosstabulation of topic-marking prepositions (Marker) and semantic classes (wc) with percentages calculated ‘within Marker’ ($\chi^2 = 95.3, p < .001$).

			wc					
			Communication	Cognitive process	General	Action	Information	Total
Marker	<i>Angående</i>	N	76	32	4	38	37	187
		%	42.9%	17.5%	3.8%	21.1%	26.1%	23.8%
	<i>Beträffande</i>	N	57	93	70	76	37	333
		%	32.2%	50.8%	66.7%	42.2%	26.1%	42.3%
	<i>Gällande</i>	N	11	18	11	17	13	70
		%	6.2%	9.8%	10.5%	9.4%	9.2%	8.9%
	<i>Rörande</i>	N	33	40	20	49	55	197
		%	18.6%	21.9%	19.0%	27.2%	38.7%	25.0%
Total	N		177	183	105	180	142	787
	%		100%	100%	100%	100%	100%	100%

Table 6. Crosstabulation of topic-marking prepositions (Marker) and semantic classes (wc) with percentaged calculated ‘within we’ ($\chi^2 = 95.3, p < .001$).

the TMP (Marker) perspective, the most important semantic classes seem to be Communication, Cognitive processes and Action.

The figures in Table 6 corroborate the preference of *angående* for words of communication, followed by *beträffande*. This TMP dominates in the domains of Cognitive processes, General and Action words, while the preference of *rörande* for Information heads is also confirmed. What is interesting to observe in Table 6 is that *beträffande* is the most frequent or second most frequent TMP for all five semantic classes, a fact due to its numerical dominance (42.2% of all TMPs). Conversely, the few cases of *gällande* are also enhanced in the figures in Table 6.

To sum up, the discussion of the lexical co-occurrence of the four TMPs with their head elements, most of which are nouns of communication and cognitive processes or verbs or nouns of action, has shown that there is great variation in collocation patterns. The most frequent of all the heads, *fråga* ‘question’ (with 8, 4, 3, 15 = 30 cases in total) ranks highest for *angående* and *rörande* in terms of Frequency and Attraction. *Fråga* actually seems to be a significant collexeme of *rörande*, differentiating it from the other TMPs. However, despite this relative importance of *fråga* with regard to *rörande*, with a Reliance score of only 0.09 percent (15 out of 16,600 uses of *fråga* combine with *rörande*), *rörande* can hardly be considered a highly salient sub-construction for *fråga*.

Except for *fråga rörande* ‘question regarding’, though, the collostructional analysis shows little evidence for any formulaic sequences including the TMPs. This conclusion is enhanced by the fact that only six percent of all analysed topic-marking constructions include collexemes that co-occur with two or more TMPs. Instead, since all the TMPs are used with a large number of different head elements, they clearly behave distinctly; on the other hand, the absence of clear collocation patterns also indicates that they are used very freely and in highly similar contexts.

Introducing the semantic classes of the head elements into the discussion reveals some more detailed preferences: *angående* lines up with Communication words and *rörande* prefers Action and Information words. On the other hand, the most frequent of the four TMPs, *beträffande*, is used with all five semantic classes, and the same goes for *gällande*, although with lower overall frequency.

3.2 Multinomial logistic regression analysis

Compared with the collostructional analysis presented in the previous subsection, the multinomial logistic regression analysis (MLRA) is a natural following step to take, since what it does is, in a sense, to take the idea of co-occurrence to another level and consider not only one element, e.g. the specific head element, but rather the whole set of co-occurring factors (syntactic and semantic) and see to what degree they map together with one or several of the focus expressions, i.e. the TMPs. Simply put, the MLRA performed on the annotated corpus data is actually a mathematically sound

Effect	Model Fitting Criteria	Likelihood Ratio Test		
	-2 Log Likelihood	Chi-square	df	p-value
Intercept Only	1024.54			
Final model ^a	757.496	267.052	27	.000
Reduced model ^b				
uniq	774.947	17.451	3	.001
fig	798.936	41.439	3	.000
H_num	766.516	9.020	3	.029
HEAD_noun	862.500	105.004	3	.000
wc_com	776.074	18.577	3	.000
wc_cog	769.417	11.921	3	.008
wc_gral	796.072	38.575	3	.000
H_compl	766.610	9.114	3	.028
Ag_anim_unspec	765.840	8.343	3	.039

^a The final model includes nine significant variables. The chi-square statistic is the difference in -2 log-likelihood between the final model and a 'blind' model (intercept only).

^b The reduced model is formed by omitting an effect from the final model. The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model.

Table 7. Model fitting information and Likelihood Ratio Test of the multinomial logistic regression model on Swedish topic expressions *angående*, *beträffande*, *gällande* and *rörande*.

Observed	Predicted				Correct %
	<i>Angående</i>	<i>Beträffande</i>	<i>Gällande</i>	<i>Rörande</i>	
<i>Angående</i>	82	71	0	34	43.9%
<i>Beträffande</i>	52	218	0	63	65.5%
<i>Gällande</i>	9	38	0	23	0.0%
<i>Rörande</i>	36	69	0	91	46.4%
Predicted %	22.8%	50.4%	0.0%	26.8%	49.7%

Table 8. Classification accuracy of the multinomial logistic regression model on Swedish topic expressions *angående*, *beträffande*, *gällande* and *rörande*.

comparison of 30 simultaneous collostructional analyses. The difference is not only that 11 factors with 30 different values are taken into account in parallel, but also that their importance is calculated relative to one another. The degree to which a particular factor is significant is thus calculated by taking all the other factors into account simultaneously (*ceteris paribus*).

Tables 7 and 8 show the main results of the MLRA that was carried out on the annotated data set. As can be seen in the Classification table in Table 8, the overall prediction capacity of the model is only 49.7%. That this value even approaches the 50% mark, however, depends almost entirely on *beträffande*, which attains a

reasonable accuracy (at 65.5%) while also being the most frequent TMP. *Angående* and *rörande* are also classified correctly to a certain extent on this scale with an accuracy of 44 and 46%, while *gällande* eludes the model completely and is mostly classified as either *beträffande* (38/70) or *rörande* (23/70). The figures in the Classification table also tell us that the model often (wrongly) classifies *angående* as either *beträffande* or *rörande*, while more than half of the instances of *rörande* are also classified as either *beträffande* or *angående*. This initial data thus corroborates the finding from the collostructional analysis that *beträffande* is the leading TMP; however, it also calls into question to what extent the other TMPs can be neatly distinguished with respect to each other and to *beträffande*.

The Likelihood Ratio Tests in Table 7 list the factors that contribute significantly to the model.¹³ Of these nine factors, which were arrived at by a step-wise approach, six are semantic (uniq, fig, wc_com, wc_cog, wc_gral and Ag_anim_unspec) and three formal (H_num, HEAD_noun, and H_compl).¹⁴ No factors concerning the governed element were found to be significant. The relative contribution of the significant factors to the different TMPs is presented in more detail in Table 9. As Table 9 indicates, the factors contribute differently to the different TMPs. One should also bear in mind that in a multinomial regression analysis, one of the four TMPs is taken as point of comparison. In this case, we decided to use *beträffande* as reference point, a choice motivated by it being the most frequent and versatile of the four. This means that the measures in the *p*-value column indicate significance of the factor in question for the relevant TMP in contrast to *beträffande*. A value above 1 in the Exp(B) column indicates that the odds for using the TMP in question with this factor are higher than for using *beträffande* (e.g. when the head element is a noun, as happens with *gällande* and *rörande*). Conversely, values below 1 indicate that the odds for using *beträffande* are higher.

As a first example, take the FIGURATIVENESS factor (fig), which makes reference to the figurativeness or abstractness of the head element. This factor is only significant for *angående* and the value of 0.098 for Exp(B) indicates that the odds for using *angående* is 1:10 compared to *beträffande* when the head element is used metaphorically. Conversely, the odds for using *beträffande* over *angående* are tenfold when the governed element is used in a figurative sense, as in *tryckte på* ‘pushed on’ in (16):

- (16) William Perry (förra försvarsministern) tryckte på *beträffande* Bosnienuppdraget, vilket han betraktade som en möjlighet att gjuta nytt liv i Nato och förändra USA:s relationer till Ryssland efter det att Moskva skickat en brigad som lyder under amerikanskt kommando.

‘William Perry (the former prime minister) put pressure on [regarding] the Mission in Bosnia, which he considered an opportunity to bring NATO to a new life and to change US relations to Russia after Moscow had sent a brigade operating under US command.’

Marker		B ^a	Std.		df	p-value	Exp(B) ^b
			Error	Wald			
Angående	Intercept	0.333	0.491	0.462	1	0.497	
	uniq	-0.197	0.373	0.279	1	0.597	0.821
	fig	-2.320	0.532	19.047	1	0.000	0.098
	H_num	-0.463	0.214	4.666	1	0.031	0.629
	HEAD_noun	0.331	0.273	1.465	1	0.226	1.392
	wc_com	0.670	0.238	7.917	1	0.005	1.954
	wc_cog	-0.615	0.264	5.408	1	0.020	0.541
	wc_gral	-2.448	0.541	20.458	1	0.000	0.086
	H_compl	0.053	0.128	0.171	1	0.679	1.054
Gällande	Ag_anim_unspec	-0.148	0.212	0.489	1	0.485	0.862
	Intercept	-2.509	0.923	7.384	1	0.007	
	uniq	-0.656	0.477	1.890	1	0.169	0.519
	fig	-0.066	0.373	0.031	1	0.860	0.936
	H_num	-0.390	0.286	1.858	1	0.173	0.677
	HEAD_noun	3.029	0.743	16.602	1	0.000	20.677
	wc_com	-0.305	0.392	0.605	1	0.437	0.737
	wc_cog	-0.368	0.352	1.095	1	0.295	0.692
	wc_gral	-0.345	0.396	0.758	1	0.384	0.708
Rörande	H_compl	-0.221	0.169	1.705	1	0.192	0.802
	Ag_anim_unspec	-0.275	0.283	0.939	1	0.332	0.760
	Intercept	-1.511	0.613	6.080	1	0.014	
	uniq	-1.297	0.331	15.371	1	0.000	0.273
	fig	0.238	0.256	0.860	1	0.354	1.269
	H_num	0.141	0.202	0.488	1	0.485	1.152
	HEAD_noun	3.197	0.492	42.203	1	0.000	24.462
	wc_com	-0.399	0.268	2.224	1	0.136	0.671
	wc_cog	-0.814	0.262	9.652	1	0.002	0.443
wc_gral	-1.011	0.315	10.287	1	0.001	0.364	
H_compl	-0.305	0.123	6.108	1	0.013	0.737	
Ag_anim_unspec	-0.591	0.209	7.987	1	0.005	0.554	

^a Coefficient estimates in the MLR model.

^b Odds ratio estimates.

Table 9. Parameter estimates for relative significance and contributions of the nine significant factors in the MLRA (multinomial logistic regression analysis). The significance for each topic-marking preposition is measured in relation to *beträffande*, and significant values ($p \leq .05$) are bolded. See Appendix A for an explanation of the abbreviations in the first column.

In the following paragraphs, we will comment on the remaining eight significant factors and their contribution to marking a difference between TMPs.

The UNIQUENESS factor (uniq) makes reference to the presence or absence of another TMP in the immediate textual context. In this case, a positive value implies

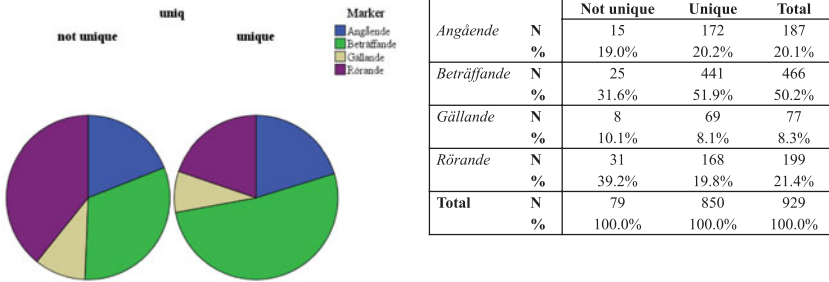


Figure 2. (Colour online) Two perspectives on the relative distribution of the uniqueness factor over the four topic-marking prepositions ($\chi^2 = 19.1, p < .001$).

the presence of another TMP, e.g. *om*, *kring*, *angående*, *beträffande*, *gällande* or *rörande*, in addition to the one explicitly analysed. As Figure 2 indicates, the co-occurrence of two or more TMPs in the same sentence is much less frequent than its counterpart, with less than 10% of the cases included in the count.

As shown in Table 9 above, the uniqueness factor is significant for *rörande*. The Exp(B) value is 0.27, which means that the odds for using *beträffande* are almost four times higher than for *rörande* when the TMP is UNIQUE. That is, *rörande* is almost four times more likely to occur when another TMP is present in the same sentence, as in (17):

- (17) I anslutning till valet 1994 uppstod fråga om lokal folkomröstning i Stockholm rörande Dennispaketet.

‘At the 1994 elections there was [a] question about a local referendum in Stockholm regarding the Dennis package.’

This can be interpreted as indicating that *rörande* functions as some sort of stylistic alternative rather than an apparent first-choice, where *beträffande* might be more suitable. This tendency is clearly visible in the pie chart and the crosstabulation table in Figure 2, where *rörande* accounts for almost 40% of the 79 not unique cases.

Another factor that is only significant with regard to one TMP is the class designating an UNSPECIFIED AGENT (Ag_anim_unspec) of the head element. Of the agent types, this was the only one that proved significant in the MLRA. As its Exp(B) value of 0.5 indicates, the odds for using *beträffande* are two to one compared to *rörande* when the agent is unspecified, as is the case in (18):

- (18) Men ett är säkert beträffande den kurdiska frågan: De omgivande länderna har alla sina intressen och investeringar i olika kurdiska organisationer, och använder dem gärna mot varandra.

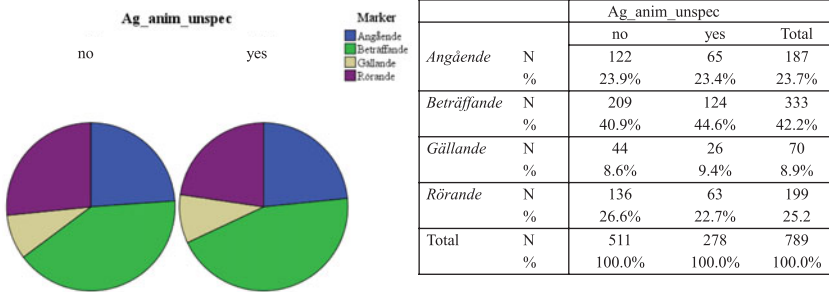


Figure 3. (Colour online) Two perspectives on the relative distribution of the unspecified agent factor over the four topic-marking prepositions ($\chi^2 = 1.83, p = .6$).

‘One thing is certain *regarding* the Kurds: the surrounding countries all have their interests and investments in different Kurdish organizations and they are not afraid to use them against one another.’

As Figure 3 reveals, the relative frequency of *rörande* with unspecified agents is lower (22.7%) than with other kinds of agents (26.6%), whereas the other TMPs are more evenly distributed over different kinds of agents. Note, though, that the data in the crosstabulation table in Figure 3 are not statistically significant. Despite this lack of statistical significance, unspecified agents is still the only one of the five different agent types included as significant in the MLR model.

The first of the three significant semantic classes of the head nouns, the words belonging to the domain of COMMUNICATION is a significant factor only for *angående*. With an Exp(B) value of 1.954, the communication domain positively distinguishes *angående* from the other TMPs, indicating that the odds for using *angående* over *beträffande* are almost two to one when the head element refers to some form of communication, such as, talking, writing, commenting, saying, etc. As the pie chart and the crosstabulation table in Figure 4 below show, with over 40% of the communication cases, *angående* clearly favors this domain, a detail that was already hinted upon in the collocation analysis (see Tables 5 and 6).

The semantic class of COGNITIVE PROCESSES is a significant factor for both *angående* and *rörande*. In both cases, however, the Exp(B) value is below 1.0, lying at 0.54 and 0.44, respectively. This means that the odds for using *beträffande* are the double as compared to *angående* and *rörande* when the head element is a word indicating a cognitive process, such as *tänka* ‘to think’, *insikt* ‘comprehension’, *enighet* ‘consensus’, *optimistisk* ‘optimistic’, and *befara* ‘to suspect’ as in (19):

- (19) Därefter ägnade han sig mest och med stor framgång åt vattenpolo, en utveckling som knappast kan befaras *beträffande* Anna.

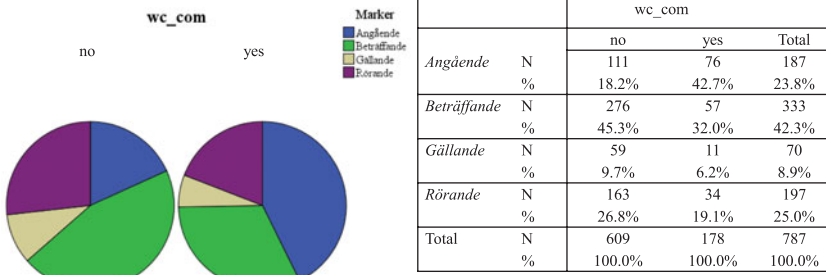


Figure 4. (Colour online) Two perspectives on the relative distribution of the communication words factor over the four topic-marking prepositions ($\chi^2 = 45.6, p < .001$).

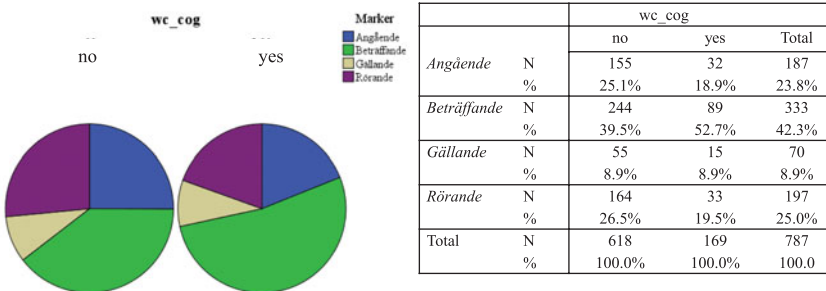


Figure 5. (Colour online) Two perspectives on the relative distribution of the words of cognitive process factor over the four topic-marking prepositions ($\chi^2 = 10.2, p = .017$).

‘After that he mostly dedicated himself to water polo with great success, a development that can hardly be suspected with regard to Anna.’

The higher percentage corresponding to *beträffande* with words of cognitive processes is evident in Figure 5, where it is seen to stand for as much as 52.7% of all the topic-marking relations belonging to this semantic domain.

The last semantic class that contributes significantly to the model is a residual one called GENERAL WORDS, which includes cases such as *alternativ* ‘alternative’, *aspekt* ‘aspect’, *använda* ‘to use’. This factor is also significant for *angående* and *rörande*, for which its Exp(B) value is lower than 1.0, lying at 0.086 and 0.36, respectively. This, of course, means that the odds for using *beträffande* greatly exceed the odds for using either *angående* or *rörande* when the head element is classified as ‘general’.

This seems to underline, once more, the relatively unmarked character of *beträffande* compared to the other TMPs, as *beträffande* stands for as many as two thirds of all cases. As the data in Figure 6 show, *angående* seems particularly repelled by heads classified as GENERAL WORDS (with an Exp(B) value of 0.086). Although statistically unimportant for the model as a whole, the relative frequency

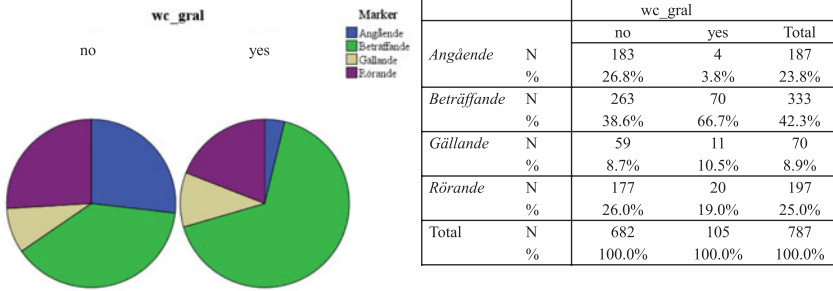


Figure 6. (Colour online) Two perspectives on the relative distribution of the general words factor over the four topic-marking prepositions ($\chi^2 = 39.4, p < .001$).

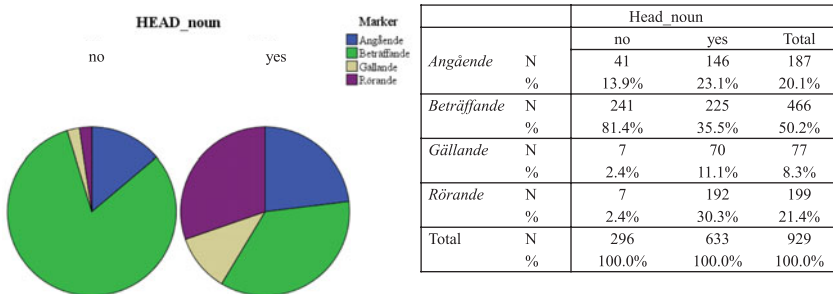


Figure 7. (Colour online) Two perspectives on the relative distribution of the nominal head factor over the four topic-marking prepositions ($\chi^2 = 185.2, p < .001$).

of *gällande*, on the other hand, actually goes up with general words as head elements compared to other heads.

Moving on to the syntactic factors, it seems natural to start off with the category of NOMINAL HEADS. This is the only factor that is significant for *gällande*, but it is also highly significant for *rörande* ($p < .001$ for both). The Exp(B) values are also very high (20 and 24, respectively), meaning that the odds for using *gällande* or *rörande* compared to *beträffande* are more than twentyfold when the head element is a noun. As the leftmost pie chart in Figure 7 shows, when the head element is not a noun *beträffande* is used in over 80% of the cases, while *gällande* and *rörande* are very seldom used with other than nominal heads.

However, as the figures in the crosstabulation table in Figure 7 indicate, due to its sheer numerical frequency *beträffande* is still the most frequent TMP even with nominal heads with 35.5% of the 633 cases. This table also clearly shows how much *gällande* and *rörande* prefer nominal heads: 70/77 and 192/199 cases. On a broader level, this result seems to indicate that *beträffande* is the most flexible of the TMPs and most readily used in varying contexts, i.e. with different head elements.

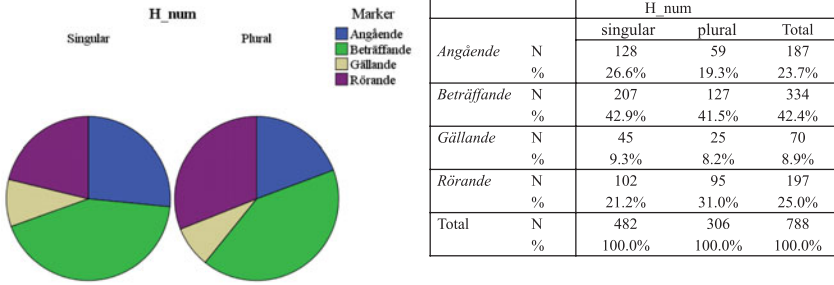


Figure 8. (Colour online) Two perspectives on the relative distribution of the head number factor over the four topic-marking prepositions ($\chi^2 = 11.9, p = .008$).

The following syntactic factor with a significant contribution to the model is the NUMBER of the head, i.e. whether it appears in singular or plural form. This factor is important only for *angående*. The Exp(B) value of 0.63 indicates that the odds for using *beträffande* are almost the double when the head is plural. As above, this result seems to underline the fact that *beträffande* is more flexible and readily used in different syntactic contexts. However, as can be seen in Figure 8, the relative frequency of *angående* is significantly higher with singular than with plural heads. Although this factor is not significant for *rörande*, the tendency for this TMP stands out as the opposite to *angående* in Figure 8, i.e. *rörande* is relatively more frequent with plural heads.

The final significant factor of the MLRA is HEAD COMPLEXITY, which refers to the formal complexity of the head element. A bare noun without syntactic arguments is classified as being low in complexity (gaining a score of 0), whereas nouns with one or several modifiers or verbs with adverbs and complex subjects are annotated as more complex with scores ranging from 1 to 3. Head complexity is only significant for *rörande* compared to *beträffande*. The *p*-value for *rörande* is .013, and its corresponding Exp(B) value is 0.74 (see Table 9), indicating that the odds for using *beträffande* are somewhat higher when the head element is a bare noun, as in (20), while *rörande* is more frequently used with more complex heads, as in (21):

(20) Regeringen avser därför att senare återkomma med förslag *beträffande* återstående regler.

‘The government intends to return later with proposals *regarding* remaining rules.’

(21) Det var inte förrän han hade fått det tredje brevet från advokaten *rörande* hustruns storstättliga legat som hon lyckades övertala honom att göra resan till Newcastle . . .

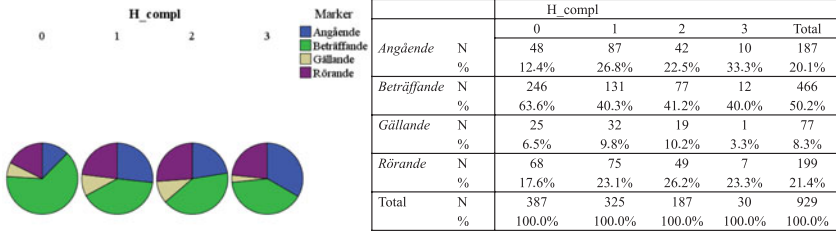


Figure 9. (Colour online) Two perspectives on the relative distribution of the head complexity factor over the four topic-marking prepositions ($\chi^2 = 54.8, p < .001$).

‘It was not until he had received the third letter from his lawyer *concerning* his wife’s grand legacy that she was able to convince him into making the trip to Newcastle.’

This is also clear in the data included in Figure 9, where *rörande* shows its lowest relative frequency with bare nouns (17.6%), whereas *beträffande* is most dominant with bare nouns (63.6%).¹⁵ If one considers that the bare noun pattern is typical of legal language, where cases such as example (20) are relatively common, *beträffande* stands out as the most salient formal TMP.

4. DISCUSSION

Looking back at the data analysis, what can be said about the four TMPs studied? The first and foremost finding regarding the specifics of the TMPs is that one of them rises above the rest, namely *beträffande*. It is the most frequent of the four, it is used with all four kinds of head elements, and with all five semantic classes. Furthermore, one must keep in mind that 29% of the cases of *beträffande* occur with no explicit head element, a fact that by itself distinguishes *beträffande* from its three companions, and, probably, also from default *om*.

In the MLRA, *beträffande* stands out as the first-choice alternative with a classification accuracy of over 65%. One of the factors in the MLRA model pointing to the special status of *beträffande* is the uniqueness factor, which reveals that *beträffande* is used more often than *angående* and *rörande* when no other TMP is present. The odds for using *beträffande* are very high compared to the other TMPs when the governed element is a bare noun (i.e. no modifier). Finally, one must not forget the significant impact of overall frequency on all corpus-driven analyses, which is also a salient feature of *beträffande*.

Gällande is used very similarly to *beträffande*, with the important difference of being very unsalient in most aspects. It can be used with heads belonging to any of the five semantic classes, but its overall frequency is a fifth of *beträffande*. Thus it is not surprising that *gällande* eludes a statistical modelling, yielding a prediction

accuracy of zero percent in the MLRA. The only distinguishing factor for *gällande* is its strong preference for nominal heads in contrast to *beträffande*.¹⁶ One reason for this lack of salience might be that *gällande* is the TMP that presents the most other uses, appearing as an adjective, a verb particle and verbally as a present participle. The prepositional use actually stands for only 11% of all cases of *gällande*, signaling its marginal status as a preposition.

What about *angående* and *rörande*? On the basis of the above analyses these two TMPs fall somewhere in between *beträffande* and *gällande*. In terms of frequency they combine for 40% of all the analysed expressions, and they have a classification accuracy of 44 and 46%. The main distinguishing factor for *angående* is that it strongly favours the communication domain and disfavours figurative heads and heads belonging to the domain of general words. *Rörande* strongly favours nominal heads, and it is repelled by many other factors, most importantly uniqueness. *Rörande* thus stands out as the TMP that is most frequently used when the context includes another TMP.

It is another matter, though, to determine to what degree *angående* and *rörande* differ from each other. A clear difference is that *angående* is used solely as a TMP, whereas *rörande* also functions as an adjective to a considerable extent (44%). In the analysis *angående* shows a bias towards Communication words, while *rörande* prefers words of Action and Information source.¹⁷

Having recalled these specifics, which help distinguish the TMPs from one another, there are two important things that must be kept in mind. Firstly, the overall predictive accuracy of the MLRA model is quite poor, not reaching even the 50% line. Only the most frequent TMP, *beträffande*, which stands for almost half the TMPs included in the data set, is described with reasonable success. The other TMPs are classified by the model as another TMP, mostly as *beträffande*: 38% of *angående*, 54% of *gällande* and 35% of *rörande* (see Table 8 above).

Secondly, the results of the collocation analysis are also unconvincing in terms of underlining differences between the TMPs, since no lexically specific collocations stand out. Although the OR, Attraction and Reliance values permit the ranking of head elements so that the order reflects a significant relationship between the TMP and the head element, the collocation frequencies are so low that they seem to impede any significant interpretation of the rankings. Instead, we found a considerable number of nouns, verbs and adjectives that occur repeatedly with many of the TMPs. Among these, the most salient are the nouns *beslut* ‘decision’, *fråga* ‘question’ and *förslag* ‘suggestion’, which co-occur with all four TMPs. This finding, though, is not significant from the perspective of distinguishing between the different topic-marking prepositions, but rather points at the importance of the topic-marking relation, in general, to these particular nouns.

Having constructed a sophisticated mathematical model (MLRA) and a detailed measure for lexicogrammatical attraction that are unable to bring forth clear

differences between the four prepositions studied, it seems reasonable to address where these apparent shortcomings might arise from.

The most obvious explanation to inconclusive results in a quantitative study would be that there was not enough data or that the data was too heterogeneous in some ways. This might well be true especially with regard to the different frequencies of the TMPs. For *gällande* only 77 cases were identified, compared to 466 cases of *beträffande*. This surely has an impact on the statistics, especially for the MLRA that uses group sizes as prior probabilities and thus favours the largest group. Another plausible explanation is that the annotated factors were insufficient or did not include the correct ones in order to single out the differences. What strikes us, though, when looking at the results of the two analyses is not, really, the lack of data to interpret, but rather its richness and the lack of salient patterns.

Another factor that is often thought to bring forth usage differences between elements is the presence of different genres. This factor is of course absent in the PAROLE corpus, which consists mainly of formal, written language. However, since it is well known that the TMPs are typically associated precisely with formal language, in this study we explicitly decided to overlook this factor and concentrate on possible differences within a loosely defined genre. This does not mean, of course, that taking genre into account is not of interest for future studies, quite the contrary.

In a sense, then, the fact that the two analyses we have performed have not produced the desired results does not seem to be due to inherent faults either in them or in the data sets. Instead, we find that the most serious impediment to finding factors differentiating *angående* and *beträffande* from one another lies in the language usage itself, where the data indicates not that they are clearly distinct in one way or another, but that they are very similar (see Glynn & Krawczak forthcoming). The differences that do show up tend to be very slight and highly detailed. Another factor influencing the interpretation of the results from the regression analysis is that in comparing four different expressions, the multinomial approach always uses one as reference point. This means that comparison with the reference point is straight-forward, but cross-wise comparison outside the reference-point expression is harder.

On the positive side, combining two complementary analyses aiming at revealing underlying differences between four near-synonyms has provided some converging evidence that there may not be that many factors that distinguish *angående* from *beträffande*, *gällande* and *rörande*. Seeing two methods produce similarly inconclusive results is in a sense less disappointing than seeing it happen with only one. In fact, having observed two approaches that are expected to mark a difference between a set of synonyms fail to do so is probably the single most important factor allowing us to emphasize the (near-)synonymous status of the TMPs studied, as will be evident in the final section.

5. CONCLUDING REMARKS

It is not easy to determine what specific differences there are among *angående*, *beträffande*, *gällande* and *rörande*. Although we have repeatedly pointed out that *beträffande* seems to be the most salient of the four, dominating when no head element is present, and *gällande* is a marginal alternative, their level of mutual contrastiveness is certainly very low. Rather, the major finding of our analyses points to subtle preferences for certain features such as *angående* preferring words of communication and *rörande* being the best option when another TMP is already present.

Considering further developments of the question under study here, a natural step would be to take the influence of genre into consideration. Bringing in the genre, though, would also automatically imply addressing the difference between default *om* and the four participle base TMPs, as well as, perhaps, *kring* ‘around’, which in present-day Swedish has become quite a popular alternative to the unwanted participles. The genre dimension could also focus on more subtle genre distinctions, such as contrasting legal, government and institutional texts with journalistic and academic texts, for example. Another possible way of identifying differences would be to take regional varieties into account, for example, the difference between the Swedish spoken in Sweden and Finland.

Another way of advancing the study would be to turn to other forms of data besides written corpora, such as the questionnaire approach advocated by Vanhatalo (2002, 2003) and Arppe (2006, 2008) or the experiments ran by Klavan (2012).¹⁸ One must keep in mind, though, that adpositional semantics is a very different matter compared to lexical semantics, in the sense that speakers’ intuitions tend to be much less reliable with function words than content words (see Granvik 2012). It would be interesting, however, to specifically test the interchangeability of the four TMPs using the most highly preferred contexts for all TMPs as control.

Furthermore, it does not seem intuitively plausible that *angående* and *rörande*, for instance, due to their lexical-semantic specifics, impose different construals on the topic-marking relation, as Taylor (2002) finds for the adjectives *high* and *tall* and Klavan (2012) for the adessive case and *peal* ‘on’.¹⁹ Most importantly, even though Klavan’s (2012) study addresses the abstract issue of grammatical synonymy, a locative relation such as the support relation is fairly concrete in comparison with the inherently figurative topic-marking relation, where an experiment such as an image rating task is difficult to imagine. On the other hand, the existence of lexically specific syntactic alternatives to the TMPs *angående*, *beträffande* and *gällande* (but not with regard to *rörande*), i.e. *vad X angår/beträffar* ‘in what concerns/regard X’, and *vad gäller Y* ‘with regard to Y’, is certainly something that could be explored further in an experimental approach such as sentence similarity or acceptability judgment tasks or the like.

Having said this, what do we make of the relationship between *angående*, *beträffande*, *gällande* and *rörande*? Quite obviously, the corpus-based statistical models do not tell us the story we wanted to hear. In fact, this was one of the motives for running the fine-grained collocation analysis. And still the picture does not seem as neat and clear-cut as one would desire. Could there be something with the whole setup, as hinted at before? Perhaps the denial of perfect synonymy and the introduction of its psychologically more valid substitute, near-synonymy, based as it is on the ideal of one meaning, one form, is what leads one partly astray. This is indeed how we set out addressing the differences between *angående*, *beträffande*, *gällande* and *rörande*; we expected that there were differences to be found. And we have gone to some detail trying to point out what these differences are.

Indeed, the detailed analysis of the specific patterning of *angående*, *beträffande*, *gällande* and *rörande* in formal written texts reveals certain idiosyncrasies. However, the classification model that emerges from the statistically sound MLRA is far from satisfactory. The same goes for the collocation analysis where few differences stand out beyond the level of individual collexemes. All of this seems to indicate that the four topic-marking prepositions indeed constitute a fairly good case of near-synonymy. There simply seem to be more similarities than differences between them: there are few unambiguously distinguishing contextual factors, the four TMPs are not semantically contrastive, their distribution is both highly similar and highly variable, and they are almost always interchangeable in the registers studied. Thus, despite being different words, we feel that *angående*, *beträffande*, *gällande* and *rörande* bear essentially the same meaning and behave syntactically in very similar ways.

In a case such as this one, then, it might not be so fruitful to dig increasingly deeper into the intricacies of particular lexemes in order to find subtle differences that are supposed to be there but do not seem to emerge. From a cognitive linguistic perspective, where redundancy in linguistic expression is accepted and even defended as the rules of parsimony and economy are frequently broken in the richness of actual language use, perhaps true synonymy is not to be ruled out after all, especially on an abstract level such as the prepositional topic-marking relation. The existence of a number of words for expressing the same meaning is but an example of the wide repertoire of our means of communication, which allows us to play around with words and meanings, using a variety of forms to express the same meanings that our thoughts swirl around as they keep us mentally occupied.

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for the *Nordic Journal of Linguistics* for their valuable comments on the original draft of this paper. All remaining inconsistencies and possible shortcomings are, of course, solely our responsibility.

APPENDIX A***Annotated factors, abbreviations, variables and annotation coding***

Syntactic factors	Abbreviations used in annotation		Annotation value(s)
		Variables	
HEAD syntactic category	HEAD	N(oun), V(erb), A(djective), 0 (no head)	1, 2, 3, 4
HEAD number	H_num	Singular, plural	0, 1
HEAD complexity	H_compl	N(0) vs. NP(1–3)	0, 1, 2, 3
HEAD definiteness	H_def	Definite, indefinite	1, 0 (i.e. yes or no)
HEAD determinacy	H_det	Determined, undetermined	1, 0 (i.e. yes or no)
HEAD modification	H_modif	Modified, unmodified	1, 0 (i.e. yes or no)
HEAD attribute	H_attr	Attribute, no attribute	1, 0 (i.e. yes or no)
GOV syntactic category	GOV	N(oun), P(roper)N(oun), Pron(oun)	1, 2, 3
GOV number	G_num	Singular, plural	0, 1
GOV complexity	G_compl	N(0) vs. NP(1–3)	0, 1, 2, 3
GOV definiteness	G_def	Definite, indefinite	1, 0 (i.e. yes or no)
GOV determinacy	G_det	Determined, undetermined	1, 0 (i.e. yes or no)
GOV modification	G_modif	Modified, unmodified	1, 0 (i.e. yes or no)
GOV attribute	G_attr	Attribute, no attribute	1, 0 (i.e. yes or no)
Semantic factors			
Semantic word class	wc	Communication, cognition, general word, action, Information source	0, 1, 2, 3, 4
Agent animacy	Ag_anim	Object, human, collective, unspecified agent/subject	0, 1, 2, 3

Syntactic factors	Abbreviations used in annotation		Annotation value(s)
		Variables	
Uniqueness	uniq	Presence, absence of other TMP	0, 1 (i.e. not unique or unique)
Figurativeness	fig	Concrete, abstract/figurative reading	0, 1 (i.e. concrete or abstract)

Continued.

APPENDIX B

Ranking of head collexemes for *angående*, *beträffande*, *gällande* and *rörande* according to their Attraction score

The relevant figures are identical to those presented in [Table 4](#), i.e. *angående* 187, *beträffande* 335, *gällande* 72, *rörande* 197. The total number of constructions used is 2,557,353, which corresponds to all prepositions in the PAROLE corpus. As in [Table 4](#), nouns co-occurring with three or four of the TMPs are marked with boldface, while italics indicate nouns co-occurring with two of the TMPs. The abbreviations used are the following: Freq = frequency, Attr = Attraction value, Rel = Reliance value, OR = Odds Ratio value.

<i>angående</i>	Freq	Atr	Rel	OR	<i>beskrifvande</i>	Freq	Atr	Rel	OR	<i>gällande</i>	Freq	Atr	Rel	OR	<i>rörande</i>	Freq	Atr	Rel	OR
fråga	8	4.28%	0.05%	6.8	<i>sympande</i>	5	1.49%	0.48%	37.0	bestut	4	5.56%	0.05%	20.0	fråga	15	7.61%	0.09%	12.6
besked	5	2.67%	0.19%	26.7	krav	5	1.49%	0.09%	6.8	överskottsmåle	3	4.17%	0.35%	13.1	förslag	8	4.06%	0.10%	13.0
beslut	5	2.67%	0.07%	9.3	gälla	5	1.49%	0.03%	2.5	anval	3	4.17%	0.09%	34.2	dokument	7	3.54%	1.02%	139.3
klagomål	3	1.60%	1.17%	164.8	opinionistisk	4	1.19%	0.80%	62.3	fråga	3	4.17%	0.02%	6.7	arrangang	6	3.05%	0.18%	24.6
data	3	1.60%	0.79%	111.2	beslut	4	1.19%	0.05%	4.1	veksambet	2	2.78%	1.14%	419.9	beslut	6	3.05%	0.08%	10.7
bestämmelse	3	1.60%	0.62%	91.8	förslag	4	1.19%	0.05%	3.7	plan	2	2.78%	0.05%	19.0	förklaring	5	2.54%	0.21%	28.0
förhör	3	1.60%	0.62%	86.5	fråga	4	1.19%	0.02%	1.8	veksambet	2	2.78%	0.04%	13.6	ärende	4	2.03%	0.28%	34.7
opinionistisk	3	1.60%	0.60%	83.9	regel	3	0.90%	0.30%	27.6	förslag	2	2.78%	0.02%	8.8	handling	4	2.03%	0.25%	33.1
information	3	1.60%	0.13%	18.6	prognos	3	0.90%	0.30%	23.1	mil	2	2.78%	0.02%	8.0	anval	4	2.03%	0.12%	16.3
arrangang	3	1.60%	0.09%	12.7	göra	3	0.90%	0.00%	0.4		2	2.78%	0.02%		analys	3	1.52%	0.21%	28.1
rid	3	1.60%	0.09%	12.6	gå från ord till handling	2	0.60%	7.14%	590.7		2	2.78%	0.02%		förhandling	3	1.52%	0.14%	18.0
brev	3	1.60%	0.09%	12.1	bestämning	2	0.60%	5.06%	404.1		2	2.78%	0.02%		program	3	1.52%	0.07%	8.9
svar	3	1.60%	0.07%	10.0	oklarhet	2	0.60%	1.59%	123.8		2	2.78%	0.02%		arbete	3	1.52%	0.03%	4.1
förfrågan	2	1.07%	1.30%	181.9	tanks om	2	0.60%	1.55%	120.9		2	2.78%	0.02%		anmaling	2	1.02%	0.82%	108.8
fråga upp	2	1.07%	1.27%	178.3	pejorativisk	2	0.60%	1.30%	101.0		2	2.78%	0.02%		palvetkan	2	1.02%	0.79%	104.1
ytrande	2	1.07%	0.89%	68.9	frågeställning	2	0.60%	1.10%	88.3		2	2.78%	0.02%		bestämmande	2	1.02%	0.72%	98.4
deklar	2	1.07%	0.43%	59.2	önsket	2	0.60%	0.99%	76.4		2	2.78%	0.02%		uppsäckt	2	1.02%	0.55%	72.8
tillfråga	2	1.07%	0.39%	54.3	förbitring	2	0.60%	0.78%	60.0		2	2.78%	0.02%		tvist	2	1.02%	0.53%	70.1
motion	2	1.07%	0.34%	47.2	förtvättning	2	0.60%	0.68%	52.4		2	2.78%	0.02%		granskning	2	1.02%	0.36%	46.8
nivå	2	1.07%	0.32%	44.8	upplysning	2	0.60%	0.56%	43.1		2	2.78%	0.02%		regel	2	1.02%	0.24%	31.2
regler	2	1.07%	0.24%	32.9	enighet	2	0.60%	0.53%	41.3		2	2.78%	0.02%		lagstiftning	2	1.02%	0.22%	29.4
presskonferens	2	1.07%	0.23%	31.7	kriterium	2	0.60%	0.51%	39.6		2	2.78%	0.02%		prognos	2	1.02%	0.20%	26.2
lagstiftning	2	1.07%	0.22%	31.0	deklar	2	0.60%	0.43%	32.9		2	2.78%	0.02%		information	2	1.02%	0.09%	11.7
sambete	2	1.07%	0.20%	27.6	tvist	2	0.60%	0.32%	24.9		2	2.78%	0.02%		föreläsning	2	1.02%	0.09%	11.5
sympande	2	1.07%	0.19%	26.3	ändring	2	0.60%	0.30%	22.8		2	2.78%	0.02%		artikel	2	1.02%	0.08%	11.0

<i>angående</i>	Freq	Atr	Rel	OR	<i>beröfjande</i>	Freq	Atr	Rel	OR	<i>gällande</i>	Freq	Atr	Rel	OR	<i>rörande</i>	Freq	Atr	Rel	OR
<i>status</i>	2	1.07%	0.16%	21.9	inskt	2	0.60%	0.26%	19.6						studer	2	1.02%	0.08%	11.0
<i>ägen</i>	2	1.07%	0.16%	21.7	argument	2	0.60%	0.25%	19.2						diskussion	2	1.02%	0.08%	10.0
<i>utlåtande</i>	2	1.07%	0.15%	20.5	osakerhet	2	0.60%	0.25%	19.1						besked	2	1.02%	0.08%	10.0
<i>rapport</i>	2	1.07%	0.14%	19.8	initiativ	2	0.60%	0.22%	18.0						undersökning	2	1.02%	0.07%	9.6
<i>ero</i>	2	1.07%	0.11%	14.7	ivelsam	2	0.60%	0.22%	16.7						algard	2	1.02%	0.07%	9.3
<i>förhandling</i>	2	1.07%	0.09%	12.6	<i>status</i>	2	0.60%	0.16%	12.2						<i>breve</i>	2	1.02%	0.06%	7.6
<i>forskning</i>	2	1.07%	0.08%	11.6	inlytande	2	0.60%	0.14%	11.0						debatt	2	1.02%	0.05%	7.0
studie	2	1.07%	0.08%	11.6	information	2	0.60%	0.09%	6.8						samarbete	2	1.02%	0.05%	6.5
uppdrag	2	1.07%	0.08%	10.7	studie	2	0.60%	0.08%	6.4						lag	2	1.02%	0.05%	6.1
diskussion	2	1.07%	0.08%	10.5	palpeta	2	0.60%	0.08%	6.2						uppgift	2	1.02%	0.03%	3.9
tanke	2	1.07%	0.03%	4.3	diskussion	2	0.60%	0.08%	5.9						problem	2	1.02%	0.02%	2.6
förslag	2	1.07%	0.02%	3.3	besked	2	0.60%	0.08%	5.8										
skrivna	2	1.07%	0.01%	1.9	kritik	2	0.60%	0.05%	3.9										
slåga	2	1.07%	0.00%	0.3															

Continued.

NOTES

1. Dirven (1993:87–89) uses the label ‘Area’ to cover the notion of ‘aboutness’. This label is motivated as an extension from the originally spatial meanings of the prepositions included in Dirven’s study, but in our view it does not capture the topic meaning in any transparent way.
2. The head elements (*talk/think*) correspond to Langacker’s (1992:289–290) notion of TRAJECTOR of the prepositional relation, while the governed element, *future*, functions as the so-called prepositional LANDMARK.
3. It is interesting to see that Hudson & Wiktorsson (2009:94) find a negative connotation in the formulaic structures that pattern around the TMP *about* in English, especially in combination with adjectives and nouns belonging to the domain of communication and opinion. They go on citing Chalker (1990:4), who states that ‘[b]ooks, articles and discussions can be *about* or *on* something. But *on* suggests a more serious study of the topic’. See Lindstromberg (2010:141) for a similar observation on the relationship between *about* and *on*.
4. It is interesting to see that Glynn & Krawczak (forthcoming), who are able to theoretically build up the testable hypothesis that *of* and *about* impose different construals on the complements of verbs of thinking and speaking, find no clear evidence of differences in construal despite thorough analyses.
5. To our knowledge, focusing on prepositional synonymy implies a move away from traditional cognitive linguistics studies on prepositions, since most previous studies on prepositional semantics have focused on the polysemous meaning structure and metaphoric extensions of single prepositions (see Lakoff 1987, Zelinsky-Wibbelt 1993, Sandra & Rice 1995, Tyler & Evans 2003, Martola 2007, Granvik 2012).
6. Since the word class of the governed element (GOV) proved to be insignificant for distinguishing between the TMPs (see Table 2, below), the semantic annotation was not performed on the GOV element.
7. As we have noted above, we consider that the topic-marking relation as defined in this study can reasonably be understood as a schematic construction in Goldberg’s (1995) or Croft’s (2001) sense. We refer to the instantiations of each individual TMP in the topic-marking relation as a sub-construction of the more general topic-marking construction, which would also include, at least, the prepositions *om* and *kring*. Including the more complex alternatives *vad angår/beträffar/gäller* ‘when it comes to’ is theoretically more problematic, since their prepositional status is disputable.
8. This score is numerically quite close to what Schmid (2000, 2010) and Schmid & Küchenhoff (2013) call ATTRACTION, which is the relative frequency of the use of a word in given construction. For *fråga angående* this would be $8/187 = 4.3\%$. A complementary frequency value, the so-called RELIANCE score, is obtained by dividing the frequency of *fråga* together with *angående* by the total frequency of *fråga*, 16,600, i.e. $8/16,600$, giving a frequency of 0.05% (0.0005).
9. As Schmid & Küchenhoff (2013:547–558) indicate, the exact same contingency table values are used for performing the Fisher Exact test, which constitutes the basic of Stefanowitsch & Gries’ (2003) and Gries & Stefanowitsch’s (2004) Collostructional Analysis.
10. See note 9 above. The Reliance score measures the degree to which the element in question ‘relies’ on the particular construction and tends to be much higher for low frequency words. Conversely, high frequency words typically score higher for Attraction, which measures

the degree of ‘attraction’ the construction exercises on the word, i.e. in how many of all uses of the construction it is included (see Schmid 2010:107–111).

11. See Appendix B for a table ranked according to the Attraction score.
12. As one of the reviewers remarked, this is perhaps not so surprising, since the numeric dominance of default *om*, thousandfold compared to our four TMPs, might well be reflected in the dispersed lexical-contextual distribution of *angående*, *beträffande*, *gällande* and *rörande*.
13. For the purposes of this study, only a main effects analysis was carried out. An attempt at including two-way interactions proved insignificant.
14. See Appendix A for an explanation of these abbreviations.
15. From the Marker-internal perspective, the 246 cases of *beträffande* with bare nouns stands for over 52% of its uses, whereas for the other TMPs the relative frequencies with bare nouns range between 25.7 and 34.2%.
16. In fact, a pairwise binomial logistic regression analysis (BLRA) performed on *beträffande* and *gällande* revealed one distinguishing factor, nominal heads, but the model could not meaningfully distinguish the two.
17. A binomial logistic regression analysis on *angående* and *rörande* found five significant factors: *angående* favours COMMUNICATION words and being the sole TMP (*uniq*), while *rörande* is associated with FIGURATIVE head elements, PLURAL heads and GENERAL words. Of these five factors, UNIQUENESS and FIGURATIVENESS are the strongest.
18. After discussing different approaches for measuring collocation strengths, Schmid & Küchenhoff (2013:562–563) proceed to compare the corpus-based measures (Attraction/Reliance, Odds Ratio and Fisher Exact Test) with experimental data (Gries, Hampe & Schönefeld 2005, 2010), only to find that the correlation between the different corpus analyses is much higher than the correlation between the corpus analyses and the experimental data.
19. Despite convincing theoretical argumentation, Glynn & Krawczak (forthcoming) fail to find convincing evidence of differences in construal between *of* and *about* as topic-marking prepositions, a contrast which is intuitively much clearer than the one between *angående*, *beträffande*, *gällande* and *rörande*.

REFERENCES

- Arppe, Antti. 2006. Complex phenomena deserve complex explanations – choosing how to think in Finnish. Presented at the QITL2 Conference, Osnabrück, Germany, 2 June 2006.
- Arppe, Antti. 2008. *Univariate, Bivariate, and Multivariate Methods in Corpus-based Lexicography – a Study of Synonymy*. Helsinki: Publications of the Department of General Linguistics, University of Helsinki.
- Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2), 131–159.
- Chalker, Sylvia. 1990. *English Grammar: Word by Word*. London: Nelson ELT.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, William & D. Alan Cruse. 2008. *Lingüística cognitiva [Cognitive Linguistics, Spanish edition]*. Madrid: Akal.
- Cruse, D. Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

- Desagulier, Guillaume. 2014. Visualizing distances in a set of near-synonyms: *Rather, quite, fairly* and *pretty*. In Glynn & Robinson (eds.), 145–178.
- Dirven, René. 1993. Dividing up physical and mental space into conceptual categories by means of English prepositions. In Zelinsky-Wibbelt (ed.), 73–97.
- Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1), 23–60.
- Firth, John R. 1968. *Selected Papers of J. R. Firth 1952–59*. Edited by F. R. Palmer. London: Longmans.
- Gilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions*. Amsterdam & Philadelphia, PA: John Benjamins.
- Glynn, Dylan. 2010. Testing the hypotheses: Objectivity and verification in usage-based Cognitive Semantics. In Glynn & Fisher (eds.), 239–269.
- Glynn, Dylan. 2014. Polysemy and synonymy: Corpus method and cognitive theory. In Glynn & Robinson (eds.), 7–38.
- Glynn, Dylan & Kerstin Fisher (eds.). 2010. *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*. Berlin & New York: De Gruyter Mouton.
- Glynn, Dylan & Karolina Krawczak. Forthcoming. Operationalising construal. Of/about prepositional profiling for cognition and communication predicates. In Bretones Callejas & Chris Sinha (eds.), *Construals in Language and Thought: What Shapes What?* Amsterdam: John Benjamins.
- Glynn, Dylan & Justyna Robinson (eds.). 2014. *Corpus Methods for Semantics*. Amsterdam & Philadelphia, PA: John Benjamins.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.
- Granvik, Anton. 2012. *De de: Estudio histórico-comparativo de los usos y la semántica de la preposición de en español* [On *de*: A historical-comparative study of the uses and semantics of the preposition *de* in Spanish]. Helsinki: Société Néophilologique. [Ph.D. dissertation]
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4), 635–676.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2010. Converging evidence II: More on the association of verbs and constructions. In Sally Rice & John Newman (eds.), *Empirical and Experimental Methods in Cognitive/functional Research*, 59–72. Stanford, CA: CSLI.
- Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34, 121–150.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics* 9(1), 97–129.
- Halliday, M. A. K. 1967. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3, 199–244.
- Hudson, Jane & Maria Wiktorsson. 2009. Formulaic language and the relater category: The case of *about*. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen Wheatley (eds.), *Formulaic Language*, 77–96. Amsterdam & Philadelphia, PA: John Benjamins.
- Janda, Laura & Valery Solovyev. 2013. What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. In Laura A. Janda

- (ed.), *Cognitive Linguistics: The Quantitative Turn. The Essential Reader*, 295–324. Berlin & New York: De Gruyter Mouton.
- Klavan, Jane. 2012. *Evidence in Linguistics: Corpus-linguistic and Experimental Methods for Studying Grammatical Synonymy*. Tartu: University of Tartu Press. [Ph.D. dissertation]
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: The University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, vol. 1: *Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 1992. Prepositions as grammatical(izing) elements. *Leuvense Bijdragen* 81, 287–309.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Lindstromberg, Seth. 2010. *English Prepositions Explained*, revised edn. Amsterdam & Philadelphia, PA: John Benjamins.
- Liu, Dilin. 2010. Is it a *chief, main, major, primary, or principal* concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics* 15(1), 56–87.
- Liu, Dilin. 2013. Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics* 24(1), 67–113.
- Martola, Nina. 2007. *Konstruktioner och valens. Verbfraser med åt i ett jämförande perspektiv* [Construction and valency: The preposition *åt* in Swedish]. Helsingfors: Institution för nordiska språk och nordisk litteratur, Helsingfors universitet. [Ph.D. dissertation]
- PAROLE corpus: <http://spraakbanken.gu.se/korp/#>.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York & London: Longman.
- SAG = Ulf Teleman, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik* [Grammar of the Swedish Academy], vol. 2. Stockholm: Svenska Akademien.
- Sandra, Dominiek & Sally Rice. 1995. Network analysis of prepositional meaning: Mirroring whose mind – the linguist’s or the language user’s? *Cognitive Linguistics* 6, 89–130.
- Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin & New York: Mouton de Gruyter.
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Glynn & Fisher (eds.), 101–133.
- Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3), 531–577.
- SLAF = *Svenskt lagspråk i Finland*. Föreskrifter, anvisningar och rekommendationer från statsrådets kansli [Swedish legal language in Finland: Norms, instructions and recommendations from the Prime Minister’s Office] <http://vnk.fi/julkaisut/listaus/julkaisu/sv.jsp?oid=296071> (accessed 16 December 2013).
- Ssb = Svenska språknämnden. 2003. *Svenskt språkbruk. Ordbok över konstruktioner och fraser* [Swedish language usage: Dictionary of constructions and phrases]. Norstedts.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–243.
- Taylor, John. 2002. Near synonyms as co-extensive categories: ‘High’ and ‘tall’ revisited. *Language Sciences* 25(3), 263–284.

- Thuilier, Juliette & Laurence Danlos. 2012. Semantic annotation of French corpora: Animacy and verb semantic classes. http://www.lreconf.org/proceedings/lrec2012/pdf/552_Paper.pdf (accessed 21 December 2012).
- Tyler, Andrea & Vyvyan Evans. 2003. *Semantics of English Prepositions*. Oxford: Oxford University Press.
- Vanhatalo, Ulla. 2002. «Naiset motkottaa aiheesta ja nalkuttaa syystä» Kyselytestit verbien semanttisten sisältöjen arvioinnissa [Using questionnaires to assess semantic content of verbs]. *Virittäjä* 3/2002, 330–353.
- Vanhatalo, Ulla. 2003. Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – Mitä vielä *keskeisestä* ja *tärkeästä*? [Evaluating the semantic content of near-synonyms: Population tests versus corpus linguistics]. *Virittäjä* 107(3), 351–369.
- Zelinsky-Wibbelt, Cornelia (ed.). 1993. *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*. Berlin & New York: Mouton de Gruyter.