Corpora et comparatio linguarum: **Textual and contextual perspectives**

Signe Oksefjell Ebeling, Hilde Hasselgård

University of Oslo (Norway)

1.   Introduction

This collection of papers arose out of the contrastive pre-conference workshop at the 38th ICAME[1] conference organized by Charles University in Prague in May 2017. The first part of the title of this issue of BeLLS, and indeed the title of the workshop, was inspired by the overall conference title: *Corpus et Orbis: Interpreting the World through Corpora*. The workshop theme – textual and contextual perspectives – was chosen in recognition of the fact that the conference took place in the home university of the Prague School and Functional Sentence Perspective. We were fortunate to be able to include a keynote presentation in the workshop by a prominent representative of this school, Prof. Libuše Dušková. Her own work reflects two of the main concerns of Functional Sentence Perspective: the close association between syntactic form and information structure and the text-based comparison of languages (Dušková 2015). Corpus-based contrastive linguistics is both related to and inspired by the text-based language comparison that existed before multilingual corpora, as acknowledged by Johansson (2009).

The call for papers invited scholars to make use of corpora to report on textual and contextual matters in a cross-linguistic perspective. As pointed out by Johansson (2011), one of the great advantages of multilingual corpora in contrastive linguistics is that they can make "possible a comparison of language use in context. We can compare not just structures, but their conditions of use" (2011: 125). The papers in this collection demonstrate that the contextual perspective may be taken at many levels of linguistic analysis, from the interpretation of single lexical items to the study of information structure. The contextual perspective is evident as well as inevitable in a cross-linguistic study of the FSP of English and Czech (Dušková), cohesive chains (Kunz & Lapshinova-Koltunski), and the use of marked theme (Rørvik & Monsen). Moreover, particular syntactic and phraseological constructions can only be defined, and thereby studied, by reference to their contexts (see studies by Ebeling, Hasselgård, and Ström Herold & Levin). The interpretations of individual lexical items, such as the Swedish verbs *skall* (Aijmer) and *orka* (Johansson & Nordrum), the English *fail to* (Egan), and the postfix *-pak* (Šebestová & Malá) are closely connected with the contexts in which they occur – and moreover, the translators are likely to draw on the

---

[1] International Computer Archive of Modern and Medieval English (`http://clu.uni.no/icame/`).

context in order to find appropriate counterparts in cases where the target language lacks a direct equivalent.

2.   Contents of the volume

All the papers in this collection compare English with at least one other language on the basis of parallel (translation) or comparable corpora. The languages studied, in addition to English, are Czech, German, Norwegian and Swedish.

Libuše Dušková's paper offers a critical account of the use of parallel corpora consisting of original and translated texts. The main pitfalls are associated with cross-linguistic differences related to information structure in English and Czech. The three aspects investigated − linear ordering of clause elements, FSP structure and distribution of communicative dynamism − are indeed shown to pose challenges in determining adequate translation counterparts. These challenges notwithstanding, Dušková maintains that the use of parallel texts "is irreplaceable insofar as it is the only methodology that provides expression of the same content worded in different languages" (p. 5).

Kerstin Kunz and **Ekaterina Lapshinova-Koltunski** look at lexical cohesion and chains of coreference in a study of four spoken and written registers of German and English: fiction, essays, interviews, and popular science. Chains of coreference are taken to reveal the development of discourse topics. A contribution of the study is that it considers the two types of cohesion together, in looking at the intersections of lexical cohesion and coreferential chains. The study uncovers cross-linguistic contrasts: In particular, there are more overlapping antecedents and more intersections in German than in English but the number of overlapping anaphors is higher in English than in German. In addition there are cross-register differences within each language. For example, fictional texts are distinct from other registers in both languages, though there are cross-linguistic differences. On the other hand, popular science texts show cross-linguistic similarities as regards chain intersection.

Sylvi Rørvik and **Marte Monsen** investigate the use of marked themes in English and Norwegian within the field of didactics. Using material from research articles written in L1 English and Norwegian and L2 English (by L1 Norwegian speakers), the authors aim to uncover potential contrastive differences (L1 vs. L1) in order to inform novice (L2) writers of good practices with regard to textual features. Rørvik and Monsen demonstrate that, although there are some significant contrastive differences both in terms of realizations and meanings of marked themes in didactics articles, the L2 writers are generally shown to adapt to English, discipline-specific discourse conventions in these respects.

Drawing on material from the English-Norwegian Parallel Corpus+, **Signe Oksefjell Ebeling** explores the cross-linguistic congruence of two stance frames: *it BE ADJ that* and *det VÆRE ADJ at*. The results indicate that, although there is a lot of similarity between the two languages in the use of these frames (55% intertranslatability, or congruence in translation), 45% non-congruence is noted, i.e. overt but formally different correspondences. The degree of congruence is found to depend on the type of attitude/evaluation expressed by the frames. Moreover, both languages are shown to have a number of other expressions of attitudinal stance at their disposal.

Hilde Hasselgård studies sentence-initial indefinite subjects in English and Norwegian. Since indefiniteness is associated with new information, such subjects appear to violate the information principle that co-exists with the SVO principle in both languages. The English-Norwegian Parallel Corpus is used for comparing original texts as well as originals and their translations. Both parts of the study indicate that English is more tolerant of indefinite subjects than Norwegians. However, certain contexts are favourable to indefinite

subjects in both languages, particularly so-called 'bare presentatives' which include a verb of existence/appearance and a place adverbial, generic sentences, and sentences with an indefinite NP in object position. The fact that more changes are made to indefinite subjects in translation from English into Norwegian than vice versa is related to the stronger influence of the light-subject constraint in Norwegian.

Jenny Ström Herold and **Magnus Levin** use a new resource – the Linnaeus University English-German-Swedish parallel corpus (LEGS) – to explore German and Swedish correspondences (translations and sources) of English supplementive *ing*-clauses, a clause type that lacks a productive equivalent in the target languages. It is shown that coordination is by far the most frequent correspondence in both languages, reflecting the compact and semantically indeterminate nature of supplementive *ing*-clauses. Other major correspondence types include subordination, main clause and prepositional phrase. Main clauses are found to be more frequent correspondences in German than in Swedish, which is attributed to the fact that there seems to be an increasing preference for parataxis (rather than hypotaxis) in German overall. Ström Herold and Levin also note some instances of explicitation in both the German and Swedish translations.

Karin Aijmer shows how translations can shed light on multifunctional expressions, and how their interpretation must be informed by the context. The Swedish modal auxiliary *ska/ll* is studied through its English translations found in the English-Swedish Parallel Corpus. Since *ska/ll* is known to differ markedly in meaning and use from its English cognate, it is no surprise that *shall* is an infrequent translation correspondence. However, the analysis reveals a wide array of other correspondences, thus displaying the multifunctionality of *ska/ll*. Apart from its most frequent use in future constructions (where it typically corresponds to *will*), *ska/ll* is often found in performative uses associated with authority and obligation. *Ska/ll* is also involved in other types of speech acts, such as offer, suggestion and advice, where the translations indicate that the imposition on the hearer is weakened.

Mats Johansson and **Lene Nordrum** investigate the Swedish auxiliary *orka* through the lens of its English correspondences in parallel corpora. Since English does not have a straightforward equivalent of *orka*, the various correspondences reveal its meaning components: most importantly, participant-internal ability and sufficient physical or mental strength/energy. The sufficiency component distinguishes it from other ability verbs such as *kunna* ('be able to'). The analysis also finds that *orka* tends to occur in non-assertive, often negative, contexts, which is related to the fact that it is more important to specify sufficiency in the case of unrealized events. Drawing on van der Auwera and Plungian's (1998) *Modality's semantic map*, the authors argue that sufficiency should be regarded as a layer of modality.

Thomas Egan investigates the English construction FAIL TO and its Norwegian correspondences in the English-Norwegian Parallel Corpus. Since Norwegian does not have a direct counterpart to FAIL TO, the correspondences are expected to reveal grammaticalized uses of the expression, particularly in contexts where FAIL TO does not imply effort, duty or expectations on the part of the Subject. FAIL TO is found to be much more frequent in English source texts than in translations from Norwegian, which is related to the absence of such a construction in Norwegian and the relative unlikelihood of translating a default negator such as *ikke* ('not') by a more wordy construction, such as FAIL TO. However, *ikke* is a relatively common translation of FAIL TO, both on its own and in combination with a verb meaning roughly 'manage'. Almost half of the instances of FAIL TO in translations have a negative element such as *ikke* or *ingenting* ('nothing') as their sources. The study thus shows that FAIL TO, at least in some of its uses, can be described as a periphrastic negative.

Using the Czech-English part of the InterCorp, **Denisa Šebestová** and **Markéta Malá** explore the discourse functions of the Czech postfix *-pak*. A number of both primary and

secondary functions are uncovered through an analysis of its English translation counterparts, lending support to the authors' assumption that such counterparts can indeed be used to shed light on this multifunctional postfix. Expressions ending in *-pak* are found to have the primary functions of marking epistemic modality, voicing an appeal, and marking a change in the speaker's previous assumption. In addition, *-pak* expressions are shown to have the ability to mark other pragmatic meanings, including politeness and textual cohesion.

3.   Acknowledgements

References

Dušková, L. 2015. *From Syntax to Text: the Janus Face of Functional Sentence Perspective*. Prague: Karolinum Press.

Johansson, S. 2009. Multilingual corpora: Possibilities and limitations. Unpublished manuscript for a lecture given at Charles University, Prague.

Johansson, S. 2011. A multilingual outlook of corpora studies. In *Perspectives on Corpus Linguistics*, V. Viana, S. Zyngier and G. Barnbrook (eds), 115–129. Amsterdam: John Benjamins.

van der Auwera, J. and Plungian, V. A. 1998. Modality's semantic map. *Linguistic Typology* 1998(2): 79–124.

Parallel corpora and contrastive linguistics: Where to look for pitfalls in the translation of information structure

Libuše Dušková

Charles University, Prague (Czech Republic)

Abstract: This paper deals with some of the problems encountered in English-Czech contrastive studies of information structure based on parallel texts. It largely focuses on those arising from the different hierarchy of the respective word order principles, the primary one being grammatical function in English, and information structure / functional sentence perspective (FSP) in Czech. Three aspects are considered: linear ordering of clause elements, FSP structure, and the basic distribution of communicative dynamism. Owing to the character and complexity of the FSP factors, word order, context, semantics and intonation, and the lack of distinctive realization forms of the carriers of the FSP functions, the study is based on manual excerpts from passages of digitalized running text of three English novels and their Czech translations drawn from the InterCorp. The results show that most problems arise in the case of different linear ordering as it may indicate either an identical or a different FSP structure.

Keywords: functional sentence perspective, linear ordering, basic distribution of communicative dynamism, translation counterpart, English/Czech

1.   Introduction

The present paper addresses some of the problems encountered in English-Czech contrastive studies of information structure based on parallel texts. Most of these problems result from the nature of this methodology itself: nevertheless, it is irreplaceable insofar as it is the only methodology that provides expression of the same content worded in different languages.

The main problems involved in this approach to language comparison have been outlined in a previous paper (Dušková, 2017), where this method was addressed from two aspects: a historical overview of English-Czech contrastive studies based on original texts and their translations since their beginnings in the fifties of the last century; and the variability of translation counterparts, discussed on the basis of two translations of the same novel.[1] The historical survey has shown three recurrent topics: the condensed structure of the English sentence vs. Czech subordinate finite clauses, English verbo-nominal predicates vs. Czech verbal ones, and arrangements of the information structure in English and Czech

---

[1] A passage from Kingsley Amis's *Lucky Jim*, translated by Jiří Mucha in 1954 and Kateřina Hilská in 2011.

within the theoretical framework of functional sentence perspective (FSP). The topic of the present article was suggested by the last point on the ground of its being the least elaborated and the most complex one.

As regards the pitfalls of this methodology, one of them, viz. the influence of the source language on the translation, was pointed out at the very beginning by Vachek (1955) and Hladký (1961), who studied the Czech counterparts of the English condensed sentence structure.

Another problem appeared in the choice of relevant translation counterparts in connection with their non-uniqueness, reflected in actual or potential variability. Contrastive studies of points from the level of clauses and sentences appear to yield relevant results where the comparison is based on such counterparts that reflect all semantic elements of the clause or sentence of the source language and whose adequacy is shown by the recurrence of the same translation patterns in all the samples making up the research material.[2] On the other hand, instances of free translation of clause and sentence structure do not as a rule contribute to displaying systemic relations. They may of course be of interest from other viewpoints, especially where their use indicates absence of parallel devices in the target language. The problem here is drawing a borderline between "close" and "free" translation. In this paper "free" translation counterparts have been excluded from the analysis. An example of free translation is given in (1).

(1) A strange asymmetry, stopped him being girlishly handsome. (Galbraith)
Tvář měl zvláštně asymetrickou, což ho zachraňovalo před vyloženě dívčím půvabem. (Šenkyřík)
[Face he-had strangely asymmetric, which him saved from patently girlish charm.][3]

A similar problem was encountered in the shifts observed in the rendition of the information structure. Translation counterparts deviating from the information structure of their source language originals, apart from possible mistakes or slips on the part of the translator, may be due to changes in the semantic structure. More frequently, however, it is the information structure alone that displays minor or greater shifts. It is this point that the present paper attempts to clarify in the following sections. As will be shown, a shift in the information structure need not reflect a translation failure, but may be due to a lack of means in the target language serving the same function, or it may be a case of indeterminacy / potentiality that offers more interpretations even in the source language.

2.   The theoretical framework of FSP

Information structure is here conceived within the theoretical framework of functional sentence perspective (FSP), developed by Firbas (1992), and further elaborated by his Brno co-workers.[4] Functional sentence perspective is defined as the distribution of degrees of communicative dynamism (CD) over the elements of the sentence. Degrees of CD are defined "as the relative extent to which a linguistic element contributes towards the further development of the communication" (Firbas, 1992: 8). If the elements of the sentence are arranged according to a gradual rise in their degrees of CD, the sentence displays the basic

---

[2] Cf. Johansson's concept of "translation paradigm" (2007: 23).
[3] Literal translation or exact formal rendition, where needed, is added in square brackets.
[4] The FSP theory goes back to Vilém Mathesius (1975), the founder of English-Czech contrastive studies. The original Czech text of *A Functional Analysis of Present Day English* dates from the 1920s and 1930s.

distribution of communicative dynamism (Firbas, 1992: 10). In terms of the principal FSP functions the sentence displays the ordering theme – transition – rheme. The transition is prototypically realized by the verb, the theme and the rheme are defined as the elements carrying, respectively, the lowest and highest degrees of CD (Firbas, 1992: 72-73). The definition of the rheme coincides with the concept of end focus (Quirk et al., 1985: 1356-57; Leech, 1983: 22, 64-65) but the conception of the theme differs. In the so-called British approach, the theme is defined by its initial position (Quirk et al., 1985: 1361-62; Halliday, 1994). On the other hand, the FSP structure is based on degrees of communicative dynamism irrespective of the position of the carriers of the FSP functions which, in the clausal FSP field, are represented by the respective clause elements. Their FSP functions are determined on the basis of four factors: context in/dependence, semantics, linear modification (word order) and intonation (prosody) in speech (Firbas, 1992: 10-11). An element can be disengaged from context dependence through the factors of selection, contrast, identification, purposeful repetition and the summarizing effect (Firbas, 1995: 22).

Correspondence in the FSP structure between the original and the translation counterpart is important as even an adequately rendered semantic structure of a clause, if perspectived differently, fails to convey the communicated meaning of the original.

So far, contrastive studies of the FSP structure have been mostly concerned with English and Czech with English as the source language. For treatment in the opposite direction see, e.g., Malá (2017). As regards other language pairs, e.g. Mojžíšová (2009) compared the cleft sentence in English and Norwegian, Dubec (2013) used Czech translation as a supporting device for determining the FSP structure of the Norwegian existential construction, and Aurová (2016) compared the FSP structure in Spanish and Czech. On the whole, however, there exist relatively few contrastive FSP studies of language pairs other than English-Czech few and far between.

In this paper the point under investigation is the FSP clausal field as such, i.e. the higher fields in complex sentences and the lower fields of phrases (except where acting as counterparts of clauses) are left aside. The main aim of the comparison is to find out whether or not the original clause and its counterpart express the same FSP structure, i.e. if the semantic elements constituting the theme and the rheme in the original correspond to the semantic elements that constitute, respectively, the theme and the rheme in the translation. The third main FSP function, transition, owing to its specific nature, has been left for further study.

Of the four factors determining the FSP functions, context dependence / independence (which largely coincides with given vs. new) and semantics are not language specific, and neither is the position of the intonation centre as it normally falls on the rheme. Unlike these factors, the fourth, word order (linear ordering, linear modification), is governed by different rules in English and in Czech. While in analytic English it primarily performs the grammatical function, in inflectional Czech it serves to indicate the information structure. In neutral, non-affective clauses the rheme as a rule stands at the end, irrespective of the syntactic function of its carrier. Accordingly, Czech linear ordering largely coincides with the basic distribution of communicative dynamism, or at least with the principle of end focus (taking into account the ordering in the transitional and thematic sections where it varies and often deviates from a gradual rise in CD). The principle of end focus is ascribed general validity and also operates in English, even though it may be, and often is, counteracted by the grammatical word order principle.[5] Consequently, all clauses and their translation counterparts are examined with respect to the following configurations: instances displaying

---

[5] For this point, see Chapter 10, Basic distribution of communicative dynamism vs. nonlinear indication of functional sentence perspective, in Dušková (2015: 82–92).

(a) the same linear ordering of clause elements and the same FSP structure, (b) different linear ordering and identical FSP structure, (c) different linear ordering and different FSP structure, and (d) the same linear ordering and different FSP structure.

3.　Material and method

The material was drawn from three English novels and their translations into Czech (see Sources: J. P. Barnes's *Nothing to be Frightened of,* Douglas Adams's *The Long Dark Tea-Time of the Soul*), Robert Galbraith's *The Silkworm*). Successive clauses of running text were excerpted from the opening pages of each source until their number reached 100, i.e. 300 clauses in total. The excerption was confined to the narrative parts, direct speech was excluded. Owing to the character and complexity of the FSP factors, word order, context in/dependence, semantics and intonation, and the lack of distinctive realization forms of the carriers of the FSP functions, the excerption had to be done manually, with the help of the digitalized versions of the texts available in InterCorp. Both finite and non-finite clauses were included on the ground that they differ only in the expression / non-expression of the subject and the operator, which are here inherently thematic, hence the information structure of the clause as a whole is not affected. Verbless clauses were taken into account only where they had finite or non-finite counterparts in the other language.

Each English clause and its Czech counterpart were considered with respect to agreement / disagreement from three aspects: linear ordering, FSP structure and basic distribution of communicative dynamism, the last being one of the linear arrangements of the FSP structure which coincides with the principle of end focus.

4.　Relations between linear ordering and FSP structure

4.1　Clauses with corresponding and non-corresponding linear ordering and FSP structure

The primary classification of the material into clauses whose linear ordering of semantic elements and the FSP structure correspond, and clauses in which the two variables disagree is presented in Table 1.

Table 1. Relations between the same linear ordering and (non-)corresponding FSP structure.

		Number of clauses	Corresponding linear ordering and FSP	Non-corresponding linear ordering and FSP
Adams	Finite	73	43	30
	Non-finite	26	17	9
	Verbless	1	1	–
Total		100	61	39
Barnes	Finite	69	50	19
	Non-finite	27	22	5
	Verbless	4	3	1
Total		100	75	25
Galbraith	Finite	80	43	37
	Non-finite	17	11	6
	Verbless	3	3	–
Total		100	57	43
Total		300	193 (64.3%)	107 (35.7%)

As shown by the figures, correspondence between English and Czech in linear ordering and FSP structure is found in almost two thirds of the examples. The ordering here represents the basic distribution of CD with the theme at the beginning, the rheme at the end, and the transition in between. Compare the examples listed under (2).

(2) a. Strike fished in his overcoat pocket (Galbraith)
Strike zalovil v kapse kabátu (Šenkyřík)

b. The taxi-driver had been bad-tempered (Adams)
Taxíkář byl mrzutý (Hollanová)

c. until she married my grandfather, Bert Scoltock (Barnes)
než se vdala za mého dědečka, Berta Scoltocka (Fantys)

Agreement between English and Czech in linear ordering and the FSP structure representing the basic distribution of CD is mostly found in clauses with one post-verbal clause element which is context-independent, as illustrated in (2): adverbial, subject complement, and object, respectively. The examples listed under (2) differ only in the realization forms of the clause elements. In (2) a. the form of the Czech adverbial differs in having the modifier construed as a genitive 'pocket of coat', (2) b. has different morphemic structures of the corresponding lexical items and different tenses (explicit reference to an anterior past action in English by the past perfect vs. the past tense in Czech), while (2) c. displays differences in the morhemic structure and government of the verb: the English verb takes direct object, whereas the Czech verb is reflexive (cf. the reflexive particle *se*) and takes a prepositional object governed by the preposition *za* (*marry* / *provdat se za*).

Similar instances with more than one post-verbal element are rarer since here not only does context independence play a role, but, in the case of adverbials, also their semantics. Temporal and locative adverbials with scene-setting semantics, whose basic position is at the end in English, are components of the thematic section. The examples given under (3) illustrate post-verbal adverbials that further specify the verbal action.

(3) a. that it arrived at your front door in a hot cardboard box (Adams)
že přichází k vašim dveřím v horké kartónové krabici (Hollanová)

b. as he walked down the slope towards Smithfield Market (Galbraith)
Když se pustil ze svahu ke Smithfieldské tržnici (Šenkyřík)

The Czech counterparts in (3) a. and b. differ in the expression of the subject, cf. the personal pronouns in English against personal endings of the verbs in Czech *it arrived* / *přicház-í*, *he walked* / *pustil-0 se*, further in the reflexive form of the verb in (3) b. *walked* / *pustil se* (which corresponds to *set out* rather than to *walk*), and absence of tense shift in Czech: past tense in English *it arrived* vs. present tense in Czech *přichází*.

The overall percentage 64.3% of correspondence in linear ordering and FSP structure is in good agreement with the results of previous studies (62.2%, cf. Note 5, Dušková, 2015: 184) and confirms the general validity of the end-focus principle, subject to the restrictions imposed by the grammatical system of English. In the three samples, however, the percentages of these clauses differ: while the Barnes sample considerably (almost by 10%) exceeds the average, in the Galbraith sample their representation is lower (57%), with Adams's percentage close to the average. These differences partly reflect the authorial styles and partly the translators' adherence to the original. While the prevailing simple clause

structure in Barnes's narrative offers straightforward counterparts, Galbraith's complex sentence structure makes greater demands on the translator's choices.

4.2   Clauses with identical FSP structure and different syntactic structure

The figures in the column Corresponding linear ordering and FSP in Table 1 include two groups of examples: the first comprises instances that display agreement between English and Czech not only in the linear ordering of semantic elements and the FSP structure, but also in the syntactic structure. This group was exemplified in Section 2.1 by examples (2) and (3). The second group displays agreement in the first two of the three variables, but the syntactic structure is different. The syntactic shifts between English and Czech can again be differentiated according to whether they concern the finite, non-finite or verbless form of the clause or the syntactic functions of the clause elements. Predictably, in the case of English finite clauses, no Czech non-finite counterparts were found. However, there were a few instances of correspondence between an English finite and a Czech verbless clause (a prepositional phrase), cf. (4).

> (4) He ate ... (Galbraith)
>   Při jídle [during meal] (Šenkyřík)

In contrast, also predictably, the correspondence between English non-finite and Czech finite clauses was frequent (Adams 10 out of 26, Barnes 13 instance out of 27, Galbraith 15 out of 17), cf. (5a), (5b), and (5c). Finite counterparts in Czech were also found in the marginal group of English verbless clauses, cf. (5d).

> (5) a. (Norway was not at all a good place) <u>for her to go</u>. (Adams)
>   (Norsko rozhodně není vhodným místem), kam by měla jet [where she should go]. (Hollanová)
>
> b. (He had decided) <u>to specialize</u> in the British Empire. (Barnes)
>   (Rozhodl se,) že se omezí [that he would specialize] na Britské impérium. (Fantys)
>
> c. beneath a stone griffin <u>standing</u> sentinel on the corner of the market building. (Galbraith)
>   Pod kamenným gryfem, který držel stráž [which stood sentinel] na rohu budovy tržnice, (Šenkyřík)
>
> d. (listening to the story) <u>of her infatuation</u>, (Galbraith)
>   (poslouchal příběh o tom), jak se bláznivě zamilovala [how herself foolishly she-fell-in-love], (Šenkyřík)

The shifts in the syntactic functions of clause elements largely display patterning that has been observed in previous studies in connection with English passive – Czech active and English verbo-nominal – Czech verbal predication (Dušková, 2015: 30-45, 57-83, 107-137), cf. (6) and (7), respectively.

> (6) a. most of her life had been spent at a constant distance from it. (Adams)
>   většinu života$_{accusative}$ strávila$_{active}$ ve stálém odloučení od něj.

      b. her Worker was now supplemented by China Reconstructs, (Barnes)
neboť Worker_{accusative.} nyní doplňoval_{active} časopis _{nominative} China Reconstructs (Fantys)

(7) a. they ... were of that generation (Barnes)
patřili [belonged to] ke generaci (Fantys)

      b. so Strike made a detour down a side alley (Galbraith)
a tak Strike odbočil [turned] do postranní uličky (Šenkyřík)

Another recurrent pattern was found in English initial subjects corresponding to a Czech initial adverbial or object (in an active clause), cf. (8):

(8) a. But my grandmother's life had contained another enormous change (Barnes)
V životě mé babičky_{In the life of my grandmother} se však udála_{happened} ještě jedna obrovská změna_{nominative.} (Fantys)

      b. All the way out of London to Heathrow she had suffered from doubt. (Adams)
Celou cestu z Londýna na Heathrow ji_{her accusative} sužovaly_{worried} pochybnosti_{doubt} _{nominative} (Hollanová)

Other syntactic shifts were documented by single instances, e.g. English copular qualifying predication – Czech existential sentence (which corresponds to the existential construction in English), cf. (9)

(9) and if it was remotely possible, (Adams)
a pokud je byť i jen vzdálená možnost [and if there is only a remote possibility] (Hollanová)

The recurrent patterns illustrated by (6), (7) and (8) have been noted and described in previous studies (Brůhová and Malá, 2017; Malá, 2014; Dušková, 2012; 2015: 30-45), the first and the third showing, in connection with the basic distribution of CD, systemic differences between English and Czech, due to the different function of word order: while in Czech the basic distribution of CD can be achieved by word order alone, final placement of the rheme in English may result from the subject construction of a thematic post-verbal element, which in the case of the object often involves the use of the passive.

      Different syntactic structure was also found in the configurations of non-correspondence between linear ordering and FSP structure, but here the two variables do not appear to be systemically related.

4.3   Clauses with non-corresponding English and Czech linear ordering and FSP structure

The different types of non-correspondence between English and Czech in the linear ordering of clause elements and FSP structure are summarized in Table 2.

Table 2. Relations between non-corresponding linear ordering and FSP structure.

		Different ordering - same FSP	Different ordering - different FSP	Same ordering - different FSP	Total
Adams	Finite clauses	25	4	1	30
	Non-finite clauses	6	3	–	9
	Verbless clauses	–	-	–	–
Total		31	7	1	39
Barnes	Finite clauses	15	3	1	19
	Non-finite clauses	4	1	–	5
	Verbless clauses	1	–	–	1
Total		20	4	1	25
Galbraith	Finite clauses	24	12	1	37
	Non-finite clauses	3	2	1	6
	Verbless clauses	–	–	–	–
Total		27	14	2	43
Total		78	25	4	107
%		72.9	23.4	3.7	100

4.3.1 *The largest group of non-correspondence*

The largest group of non-correspondence between English and Czech comprises clauses that display different ordering and the same FSP structure. This group accounts for almost three-quarters of all instances displaying non-correspondence between the linear ordering of clause elements and FSP structure. Although this might appear to be a major pitfall, most instances of this type show the non-correspondence to be a consequence of the grammatical function of English word order. The same FSP structure with a different linear ordering in English and Czech is mostly found in clauses displaying context-dependent post-verbal clause elements, adverbials and objects, realized by anaphoric proforms, which clearly indicate their appurtenance to the thematic section. In the Czech counterparts they appear in the preverbal position, cf. (10). The differences in the linear ordering between English and Czech are indicated by graphic marking: the corresponding clause elements are marked in the same way, by italics and underlining, respectively.

    (10) a.   the Italian owner *placed tea* <u>in front of him</u> in a tall white mug, (Galbraith)
                  italský majitel restaurace už <u>před něho</u> *postavil čaj* ve vysokém bílém hrnku, (Šenkyřík)

         b.   the pizza problem, which *drove* <u>her</u> crazy. (Adams)
                  až na známý problém s pizzou, který <u>ji</u> *doháněl* k šílenství. (Hollanová)

         c.   and that he *would call* <u>her</u> <u>from there</u>. (Adams)
                  a že <u>jí</u> [her] <u>odtamtud</u> [from there] *zavolá*. (Hollanová)

Example (11) shows a context-dependent object realized by a noun. The anaphoric character of the object is here indicated by the determiner (in underlined italics).

(11) My brother <u>did not compete</u> for *such offerings*, (Barnes)
Bratr se mnou o *tyto* lákavé nabídky <u>nesoupeřil</u>.

Another type of non-corresponding linear ordering and identical FSP structure is found in English presentation sentences with the rhematic subject in the pre-verbal or initial position (Firbas, 2010; Adam, 2013) as in (12a). Where the English sentence also contains a final scene-setting adverbial, the Czech counterpart displays a complete reversal of the positions of the theme and the rheme, which stand in their regular positions, the theme at the beginning and the rheme at the end, cf. (12b).

(12) a. In the late 1950s, <u>the Sino-Soviet Schism</u> *took place*, (Barnes)
Koncem padesátých let *došlo* <u>k čínsko-sovětské roztržce</u> (Fantys)

b. <u>a stern stone face, ancient and bearded</u>, stared back at him *from over the doorway*. (Galbraith)
z prostoru nad vstupem opětovala jeho pohled <u>strohá kamenná tvář, starověká a vousatá</u>. (Šenkyřík)

Less recurrent types of different ordering and similar FSP structure can be illustrated by adverbial modification of the verb, cf. (13).

(13) and <u>sank</u>, *with a grunt of satisfaction*, onto the hard wood and steel chair. (Galbraith)
a *se spokojeným zafuněním* <u>usedl</u> na dřevěnou židli s ocelovým kováním. (Šenkyřík)

However, since the adverbial may occur post-verbally and pre-verbally in both languages, the FSP aspect of these instances calls for more material-based treatment taking into account the relationship between the placement of the adverbial, its semantics where it is context-independent, and its realization form.

Non-corresponding linear ordering indicating the same FSP structure was also found in the Czech finite counterparts of English non-finite and verbless clauses, cf. (14a) and (14b), respectively.

(14) a. (they were ... of that generation) *advised* <u>by dentists</u> to ... (Barnes)
(patřili ke generaci,) jejímž příslušníkům <u>zubaři</u> *radili*, aby ... (Fantys)

b. <u>An embarrassed grunt</u> *later* (Barnes)
Následovalo <u>rozpačité zabručení</u> (Fantys)

4.3.2 *The second group of non-correspondence*
The second group of clauses with different ordering of clause elements and different FSP structure is represented by less than a quarter of all instances of non-correspondence between linear ordering and FSP structure. Yet it is this group that raises most questions. The most prominent shift is found where the translation counterpart differs from the original in the rheme. However, these instances are often liable to potentially dual interpretation in which

even the prosodic factor does not offer a conclusive answer. More or less clear-cut instances of different assignment of the rhematic function are illustrated in (15).

(15) a.   when <u>the top</u> *was down* (Barnes)
            když se *sundala* <u>střecha</u> (Fantys)

      b.   I *welcomed* this mysterious decision <u>with blunt self-interest</u>, (Barnes)
            <u>S neomaleným sobectvím</u> jsem toto záhadné rozhodnutí *uvítal*, (Fantys)

      c.   *not to think* about him <u>at all</u> (Adams)
            <u>vůbec</u> na něj *nemyslet* (Hollanová)

      d.   who ordered *tea* <u>with an air of defiance</u>, (Galbraith)
            který si <u>s mírným odporem</u> objednal *čaj* (Šenkyřík)

      e.   (In her savage desire for retribution against a man who ...) she would
            damage *herself and her prospects* <u>beyond repair</u> (Galbraith)
            (V zuřivé touze pomstít se muži, který ...) by dokázala <u>nenávratně</u> zničit
            sama *sebe i své další životní vyhlídky*. (Šenkyřík)

In (15a) *the top* / *střecha* is context-dependent as it occurs in a passage describing the car, where the novel element is the change of the position of the top, expressed by the predicative part of the clause. Moreover, in English its realization form clearly assigns the intonation centre to the last element, which corresponds to the Czech verb. The same FSP structure could be expressed by imitating the English verbo-nominal structure, viz. *když byla střecha dole*, but this is clearly a dispreferred rendition of the given content when compared with the verbal form. The verb here appears to be too 'weak' to carry the FSP function of rheme on its own. The deviation from the FSP structure of the original thus appears to be due to the character of the target language rather than to a slip of the translator.

In (15b) the reason for the shift in the rheme can hardly be sought in the character of the target language insofar as the final position of the Czech adverbial is equally possible. The novel elements in this clause are the verb and the manner adjunct, both nominal elements being context-dependent. While in the original the function of rheme is assigned to the adverbial, in Czech it is the verb that constitutes the rheme. Although the FSP structure of adverbial modification of the verb in Czech is a point for further study (cf. the comment on (13) in Section 4.3.1), in this case the realization forms of the two elements − univerbal verb *welcomed* / *uvítal*, expanded form of the adverbial *with blunt self-interest* / *s neomaleným sobectvím* dispose the adverbial to operate as the rheme. It is to be noted, however, that the initial placement of the adverbial in Czech may suggest a different semantic role, viz. that of the subject adjunct, qualification of the subject in the course of verbal action.

In (15c) two semantic features compete for the function of rheme: the negative polarity of the verb phrase and the maximum degree intensifier of the verbal meaning. The linear ordering and the position of the intonation centre in the English clause indicate the intensifier as the rheme, while in Czech it is the negated verb.

In (15d) the only new element is the manner adjunct, the act of ordering tea being fully derivable from the situational context: the action takes place in a café at the time of breakfast while the companion of the actor making the order is eating his breakfast and drinking tea.

Example (15e) illustrates a shift in the rhematic section, specifically the assignment of the function of the rheme proper. In the English clause the components of the rhematic section, the object and the adverbial, are ordered, without regard to the weightiness of their realization forms, according to their degrees of CD: the coordinated object is partly context-dependent, its first conjoin being a reflexive pronoun and the second conjoin through the

possessive determiner, while the final adverbial, even though less weighty in form, is an entirely new element. In the Czech counterpart, it is the weightiness of the realization forms of the two elements that determines the ordering within the rheme (Rh₁, R₂ ..., the last constituting the rheme proper).

Example (16) shows another factor that may contribute to a different FSP structure of a clause, viz. a change in its position within a higher textual unit, in this case a complex sentence. Moreover, the Czech complex sentence corresponds only to the first two clauses of an English multiple sentence that contains altogether four clauses.

> (16) *The slight unevenness in his gait* became more pronounced (as he walked down
> the slope towards Smithfield Market,) (Galbraith)
> (Když se pustil dolů ze svahu ke Smithfieldské tržnici), zvýraznila se *mírná*
> *nepravidelnost jeho chůze*. (Šenkyřík)

The subject of the English clause is presented as context-dependent, the actor's gait having been described as tramping in the preceding context. If the clause remained in its original position, the ordering of the elements might be preserved, but being placed after the subordinate clause, the final element appears to need a more weighty realization form than the verb. This case thus may be regarded as a result of the combined effect of a different placement in a higher unit and the realization forms of the elements whose FSP functions have been interchanged.

A different FSP structure due to a different linear ordering was also found among non-finite clauses, cf. (17). The factor of the change is again to be sought in the relatively light realization form of the rheme.

> (17) *from being rickety-gnashered to fully porcelained* in one leap (Barnes)
> – jediný skok *od vyviklaných zubů k čistě porcelánovému stavu*, (Fantys)

The examples adduced in this section have been classed as more or less clear-cut, while in the case of the less clear-cut there is some ground for potentiality, "which occurs when the interplay of FSP factors permit [sic] of more than one interpretation" (Firbas, 1992: 108). An instance of this kind is represented by (18).

> (18) *Two men in fleeces and waterproofs* had just vacated a table. (Galbraith)
> Od jednoho stolku právě vstali *dva muži ve fleecových vestách a nepromokavých*
> *bundách*. (Šenkyřík)

In (18) all nominal elements are situationally given: the scene is a café at the time of breakfast; what is situationally underivable is the kind of clothing and the quantifier. These elements are components of the realization form of the subject, which makes it weightier and disposes it to operate as the rheme. On the other hand, the final position and the immediately relevant situational context suggest the object for this FSP function: for the new arrival on the scene, the most important point is a vacant table, cf. the immediately preceding sentence: *Exhausted and hungry, he turned at last, with the pleasure that only a man who has pushed himself past his physical limits can ever experience, into the fat-laden atmosphere of frying eggs and bacon.* Seen in the light of both perspectives, the FSP structure appears indeterminate.

Libuše Dušková

4.3.3 *The last group of non-correspondence*
The last group of non-correspondence between the linear ordering and FSP structure, identical ordering of elements in the source and the target language expressing different FSP structures, appears to be marginal as it is represented only by four examples. The examples are adduced under (19).

> (19) a. (I have no idea) how strong her religious faith had been. (Barnes)
>   (Nemám vůbec představu o tom,) jak silná její víra bývala. (Fantys)
>
>   b. (He had just started on his sausages) when Dominic Culpepper arrived. (Galbraith)
>   Když Dominic Culpepper dorazil, (Strike se zrovna pouštěl do párků). (Šenkyřík)
>
>   c. (It was almost pathetically easy) to wind up the ex-public schoolboy (Galbraith)
>   Vytočit někdejšího žáka soukromé školy (bylo až dojemně snadné.) (Šenkyřík)
>
>   d. as eventually they had crawled past it. (Adams)
>   a když konečně projeli kolem místa neštěstí (Hollanová)

In (19a) the rheme in the English clause is the subject complement[6] (whose initial position is due to the obligatory fronting of the *wh*-element) on the ground of its context-independence; it is the only novel element of the clause as the subject is context-dependent not only through the possessive determiner but also owing to its actual occurrence in the immediately preceding context. The corresponding Czech counterpart would have the predicative adjective at the end: *jak byla její víra silná*. The actual Czech counterpart assigns this function to the final element of the English clause, the verb, as is usual in Czech, and since it carries the intonation centre, its neutral form *byla* is replaced by the longer iterative form *bývala*. The most likely source of the FSP shift is here the influence of the original whose linear ordering the Czech counterpart imitates.

In (19b) the interpretation of the FSP structure depends on whether or not the subject is context-dependent. If context-independent, the clause would be a presentation sentence with the Czech counterpart *když dorazil Dominic Culpepper*, i.e. the subject would be placed at the end. Culpepper is mentioned in the preceding context, but there are five intervening paragraphs between this mention and the occurrence in (19b). According to Firbas (1992: 23-31) and Svoboda (1981: 88-89), the retrievability span is generally limited to seven intervening clauses. Five paragraphs greatly exceed this limit, which supports the context-independent interpretation. Nevertheless, characters in novels are given elements throughout the whole texts, which often applies even to the opening passages, as may be the case here, with the FSP structure assigned to potentiality. Dual interpretation of instances of this kind was actually attested by an example in Dušková (2017: 210), which is here reproduced (for the two translations, see Note 1):

---

[6] Subject complements with rhematic function have been found in more than 90% of their occurrences (Uhlířová, 1974); their rare occurrence in the theme is due to their largely prevalent context-independence.

(19)   b.'   <u>Welch</u>, Dixon noticed, *had rejoined* the group (Amis)

Nyní, jak si Dixon všiml, se k nim opět *připojil* <u>Welch</u>, (Mucha)

Dixon si povšiml, že se k nim <u>Welch</u> *vrátil* (Hilská)

Here the actor of the action has not been mentioned in a stretch of text covering a page and a half. In these instances, however, a factor in the S-V order may also be the influence of the source language.

In (19c) the only novel element in the non-finite clause is the verb, the object referring to one of the interlocutors whose public-school education is a known fact. However, since the order of the clauses has been reversed, a weightier realization form of the object appeared to be a more suitable candidate for the rheme. A corresponding, fully acceptable Czech counterpart of the whole complex sentence constituting (19c) would be *Bylo až dojemně snadné někdejšího žáka soukromé školy vytočit*, in which the elements are ordered in the same way as in English, with the exception of the reversed position of the infinitive and its object.

The last example (19d) is a clear instance of misrepresentation of the information structure, resulting from neglect of the anaphoric character of the adverbial, univocally indicated by its realization form. A notable consequence is the use of a non-pronominal realization form *kolem místa neštěstí* 'past the place of the accident'.

Both in this Section and in Section 4.3.2 recurrent sources of the shifts in the FSP structure appear to be a potentially dual interpretation of the FSP structure in the original and preference of a differing structure because it offers a smoother rendition of the content than a possible corresponding structure, which may even be excluded altogether owing to systemic differences between the two languages.

5.   Conclusion

The preceding discussion attempted to show the intricacies of the semantic and information structure displayed by the translation counterparts at the level of finite, non-finite, and marginally also verbless clauses. Three aspects were considered, linear ordering of the clause elements, FSP structure, and within the latter the basic distribution of communicative dynamism, each contributing to the complexity involved in determining an adequate translation counterpart. Both the same and a different linear ordering in the source and the target language may indicate either a corresponding or a different FSP structure. Of the four configurations under discussion, same linear ordering – same FSP structure, different linear ordering – same FSP structure, different linear ordering – different FSP structure, and same linear ordering – different FSP structure, the least problematic seems to be the same ordering – the same FSP structure since identical ordering rarely constitutes a different FSP structure. According to the frequency of occurrence, it is the other two configurations that present more problems: different ordering in English clauses and their Czech translation counterparts largely indicates the same FSP structure, but also fairly often a different FSP structure. The complexity of the interpretation here results from the interplay of the FSP factors and the different function of word order in Czech and in English. Instances where the original clauses and their translation counterparts differ raise the question whether there are any factors contributing to the shift or whether the shift is to be ascribed to a slip of the translator. Recurrent shift-supporting factors have been found in the realization forms of the clause elements whose FSP function is changed and in the position of a clause within a higher textual unit. As regards instances of dual interpretation of the FSP structure in the original,

neither interpretation represents an FSP shift, the structure being indeterminate as such. In general, more extensive material and further research may reveal other factors. A question to be asked in particular is whether a different interpretation is not due to a lack of means in the target language for the expression of the FSP structure in the original. While the foregoing English-Czech comparison of the FSP structure does not provide much evidence in this respect, an approach from Czech to English, considering the FSP function of the free Czech word order, is likely to provide more ground for ascertaining systemic differences. In both directions, however, the extent of the research material needs to be enlarged if more insight into the questions raised here is to be gained. How to achieve this is a pitfall of the present state of research.

References

Adam, M. 2013. *Presentation Sentences* (*Syntax, Semantics and FSP*). Brno: Masaryk University, Faculty of Education Work 162.

Aurová, M. 2016. El orden de palabras y el dinamismo comunicativo: una contribución al estudio contrastivo del español y checo. *Hermenéus, Revista de Traducción e Interpretación* 18: 27–58.

Brůhová, G. and Malá, M. 2017. On English Locative Subjects. *Acta Universitatis Carolinae, Philologica* 2017(1) (*Prague Studies in English*): 19–38.

Daneš, F. (ed.). 1974. *Papers on Functional Sentence Perspective*. Prague: Academia.

Dubec, P. 2013. Syntactic and FSP Aspects of the Existential Construction in Norwegian. Doctor's Dissertation, Charles University.

Dušková, L. 2012. Vilém Mathesius and contrastive studies, and beyond. In M. Malá and D. Šaldová (eds), 21–48.

Dušková, L. 2015. *From Syntax to Text: the Janus Face of Functional Sentence Perspective*. Prague: Karolinum Press.

Dušková, L. 2017. Kontrastivní lingvistika a paralelní korpusy. *Časopis pro moderní filologii* 99(2): 193–213.

Egan, T. and Dirdal, H. (eds). 2017. *Cross-linguistic Correspondences. From Lexis to Genres*. Amsterdam: John Benjamins.

Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: University Press.

Firbas, J. 1995. Retrievability Span in Functional Sentence Perspective. *Brno Studies in English* 21: 17–45.

Firbas, J. 2010. Non-thematic Subjects in Contemporary English. In A. Svoboda, J. Chamonikolasová and L. Urbanová (eds), 281–298.

Hladký, J. 1961. Remarks on Complex Condensation Phenomena in Some English and Czech Contexts. *Brno Studies in English* 3: 105–118.

Halliday, M. A. K. 1994. *Introduction to Functional Grammar*. London: Arnold.

Johansson, S. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Leech, G. N. 1983. *The Principles of Pragmatics*. London and New York: John Benjamins.

Malá, M. 2014. *English Copular Verbs. A Contrastive Corpus-supported View*. Prague: Charles University, Faculty of Arts.

Malá, M. 2017. Non-prepositional English Correspondences of Czech Prepositional Phrases. In T. Egan and H. Dirdal (eds), 199–217.

Malá, M. and Šaldová, P. (eds). 2012. *A Centenary of English Studies at Charles University: from Mathesius to Present-day Linguistics*. Prague: Charles University, Faculty of Arts.

Mathesius, V. 1975. *A Functional Analysis of Present Day English on a General Linguistic Basis*. Prague: Academia.

Mojžíšová, K. 2009. The Cleft Sentence in English and Norwegian. Master's Dissertation, Charles University.

Quirk, R., Greenbaum, S, Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
Svoboda, A. 1981. *Diatheme*. Brno: Masaryk University.
Svoboda, A., Chamonikolasová, J. and Urbanová, L. (eds). 2010. *Collected Works of Jan Firbas*. Vol. 1. Brno: Masaryk University Press.
Uhlířová, L. 1974. On the Role of Statistics in the Investigation of FSP. In F. Daneš (ed), 208–216.
Vachek, J. 1955. Some Thoughts on the So-called Complex Condensation in Modern English. *Sborník prací Filosofické fakulty brněnské univerzity* (A 3): 63–77.

Primary sources

Adams, D. 1990 (1988). *The Long Dark Tea-Time of the Soul*. London: Macmillan Pocket.
Dlouhý, temný čas svačiny duše. 2016. Translation Jana Hollanová. Prague: Nakladatelství Argo.
Barnes, J. P. 2008. *Nothing to be Frightened of*. New York: Knopf Doubleday.
Žádný důvod k obavám. 2009. Translation Petr Fantys. Prague: Odeon.
Galbraith, R. 2014. *The Silkworm*. London: Sphere,
Hedvábník. Translation Ladislav Šenkyřík. Prague: Nakladatelství Plus, 2015.
InterCorp: Český národní korpus – InterCorp. Prague: Ústav Českého národního korpusu. http://www.korpus.cz [last accessed 12 May 2017]

Author's address

Libuše Dušková
Department of English Language and ELT Methodology, Faculty of Arts
Charles University
nám. Jana Palacha 2
116 38 Praha 1
Czech Republic

libuse.duskova@ff.cuni.cz

Libuše Dušková

English vs. German from a textual perspective:

Looking inside chain intersection

Kerstin Kunz, Ekaterina Lapshinova-Koltunski

University of Heidelberg, University of Saarland (Germany)

Abstract: This paper presents a cross-lingual corpus-based study on the intersection of chains of coreference and lexical cohesion. The two types of cohesion are often combined and thus play an important role for the development of discourse topics. We analyse chain intersection as cases where chain elements of lexical cohesion occur inside of coreference chains. We use a corpus of English and German original texts from four written and spoken registers which is annotated for both types of cohesion. Our analyses point to contrasts between the two languages and across the four registers under analysis in the types and the number of intersections in coreference chains. This variation has an effect on the way important topics develop in a text.

Keywords: coreference, lexical cohesion, discourse topics, register analysis, English, German

1. Motivation and state of the art

This paper presents a corpus-based analysis of particular types of interaction between chains of coreference and lexical cohesion that we call **chain intersection**. Our main focus is on the comparison of English and German and variation in written and spoken registers in these two languages. We argue that different types of chain intersection and the number of chain intersections reflect continuity and development of important discourse topics. They impact on how topics are perceived by text recipients as such.

There is general agreement in the literature that cohesion is an important linguistic device to explicitly establish coherence and continuity in texts. The interaction of coreference and lexical chains is regarded as essential to a text's cohesive harmony. The notion of chain interaction discussed in existing studies (Hasan, 1984; Hoey, 1991; Martin, 2015; Song *et al.*, 2015), however, differs considerably from our concept of chain intersection, as will be seen in Section 2 below.

For instance, chain interaction in Hasan's (1984, 1985) model of cohesive harmony applies when elements of different chains are realized as different constituents of the same clause – in the theme or the rheme – and as different participants involved in the same process (e.g. actor and beneficiary). Clauses or sentences in the same text containing such

reoccurring chain interactions of the same coreference and/or of lexical chains are considered to form key sentences. They are used in different approaches to measure local coherence, e.g. Grosz *et al.* (1995) in Centering Theory, or Strube and Hahn (1999) and Hoey (1991). Our concept does not consider intra-clausal interaction between different chain elements. It accounts for elements in coreference chains in which grammatical and lexical devices of cohesion are combined and where the lexical device is at the same time integrated in an element of a lexical chain. In most existing corpus-based studies, the two types of cohesion are either studied separately, are not distinguished or do not deal with discourse topics as an aspect of language contrast and register variation.

Most computational studies, such as Doddington *et al.* (2004) and more recently CoNNL 2011 (Pradhan *et al.*, 2011), are monolingual. They focus on automatic anaphora resolution and draw data from large corpus resources such as OntoNotes (Technologies, 2006). They contain information on coreference relations and bridging but do not analyse chain interaction. Computational models developed by Morris and Hirst (1991), Barzilay and Elhadad (1999) and other works building on them apply chain interaction to extract key sentences for automatic text summarization on monolingual corpora containing individual registers. Their notion of chain interaction is based on the study by Hoey (1991) and, again, differs from the concept proposed in the study presented here (see Section 2). Besides, these studies are restricted to lexical cohesion. There are also computational models which identify key sentences for automatic assessment of local coherence and are concerned with the interaction of coreference chains. They are largely based on Centering Theory (see Grosz et al. 1995). Several computational works analyse the interaction of coreference and lexical chains or bridging, such as Mesgar and Strube (2015). These, however, do not focus on the fine-grained analysis of language contrast and register variation as they develop algorithms for automatic text analysis.

There are few multilingual studies that base their analysis on corpora that are manually annotated with coreference and bridging (e.g. Zikánová *et al.*, 2015; Lapshinova-Koltunski and Kunz, 2014). They do not consider interaction between the two types of chains. The multilingual corpus-based studies by Kerremans (2014) involve fine-grained manual annotations and use lexical patterns in coreference chains to analyse terminological variation and equivalence in originals and translations to build terminological databases. This model differs from ours in that they do not analyse whether the lexical patterns in coreference chains are also part of lexical chains. So, a corpus-based account applying fine-grained annotations in order to see how languages and registers differ in terms of chain intersection in the sense explained below does not exist so far. Moreover, we argue that our approach offers new insights into the interplay of coreference chains and lexical cohesion and how this intersection contributes to the linguistic reflection of discourse topics.

The remainder of this paper is organised as follows: we start with a clarification of the main concepts used: Discourse topics, coreference and lexical cohesion as indicators of discourse topics, and chain intersection. We discuss the linguistic indicators of different aspects of chain intersection, allowing us to interpret the latter as an indicator of topic continuity and development. In Section 3, we describe our methods and resources. We use a corpus of English and German comparable texts from four registers (political essays, fictional texts, popular scientific texts and spoken interviews). The corpus is annotated for both lexical cohesion (Martínez Martínez *et al.*, 2016) and coreference chains (Lapshinova-Koltunski and Kunz, 2014) and allows filtering out the chains that represent cases of intersection. We report on the results of the above research questions in detail in Section 4 and interpret them in terms of contrasts in thematic progression between English and German, also encompassing register variation in Section 4.10.

2.   **Chain intersection**

As mentioned above, the focus of this paper is on the interplay between two types of cohesive chains: coreference and lexical chains. We therefore begin with a brief definition of the two types before we discuss our concept of chain intersection.

2.1   Discourse topics

It has previously been established that cohesive chains often interact in texts and that this interplay is an important factor influencing how textual coherence and the development of discourse topics are perceived by text recipients, see e.g. Tanskanen (2006) and Todd (2016). The term *dicsourse topic* still remains somewhat undefined and fuzzy as it has been used in the literature from a variety of different perspectives, sociological and pragmatic or cognitve and semantically oriented ones. It is understood here in the sense of Chafe (1976) and Brown and Yule (1983) from a textual rather than a grammatical perspective. Intra-clausal relations between 'theme' and 'rheme', 'topic' or 'comment' or 'topic' and 'background' are therefore not considered in this paper, although they interact with cohesive chains. Discourse topics are topics that unfold throughout the text.

For our purpose, Todd's approach seems most fitting, in which topics are defined as "clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create connectedness and relevance" (Todd, 2003: 2009). This ideational clustering of extralinguistic concepts may stay rather implicit and may depend heavily on the text recipient's inference of knowledge about the world and the context of situation. However, conceptual associations are indicated, at least to some extent, by linguistic patterns in the text. One essential mechanism to explicitly express connectedness in semantic space (see also Hoey, 1991) are cohesive chains.

A discourse topic may be global and concern the whole text or it may be rather local and be important to a smaller part of the text. But even on a more local level it often extends beyond clause boundaries. Coreference and lexical chains are employed as textual means to indicate both local and more global relations, depending on the number of elements and the distance between elements in a chain (see Kunz *et al.* 2016).

This paper is concerned with how these two types of cohesion interact. From a textual semantic point of view, we are interested in how and when coreference chains are integrated into chains of lexical cohesion. From a more conceptual point of view we investigate, how and when concepts about central individual referents (explicitly indicated by coreference) are integrated into clusterings of associated concepts, which are explicitly indicated by lexical cohesion. Let us therefore first take a look at both types of cohesive chains in turn and see how they contribute to the creation of discourse topics and then discuss how one type can intersect with the other.

2.2   Coreference and lexical cohesion

Most existing models refer to coreference and lexical cohesion as two separate relations. This distinction is grounded in formal criteria – grammatical vs. lexical devices of cohesion – and also in conceptual differences in the meaning relations established, e.g. identity vs. similarity (Hasan, 1985), or coreference vs. bridging (Clark, 1975; Poesio *et al.*, 1997). Our notions of coreference and lexical chains combine these two aspects, as illustrated in example (1).[1]

---

[1] Note that in all our examples, the extension of elements in coreference chains is marked by brackets; elements of lexical chains are underlined.

(1)   This past spring, the U.S. Department of Education issued < a <u>report</u>, The Condition of Education 2000>. Some of the trends < it> pinpointed offer evidence that .... < The <u>report</u>> found that the benefits of attending college are greater today than ever before. With significant increases in the number of students who may not speak English at home, < this <u>report</u>> suggests that ...

In **coreference chains**, **grammatical** devices signal a textual relation to other coreferring expressions in the same text. The conceptual association evoked in this way is **identity** between conceptual referents. The first element in a coreference chain, the **antecedent** introduces a new extralinguistic referent into the textual world. Linguistic forms of antecedents can be manifold: they quite often contain an indefinite article and, most essential to this paper, a lexical nominal head, as in example (1) above. The subsequent elements of the coreference chain, the **anaphors**, contain grammatical devices, signalling that the same extralinguistic referent is mentioned again (and again). These devices may either serve as a modifier of a coreferring noun phrase, like *the* and *this* in example (1), or they may function as a pronominal head, such as *it* in example (1). It is widely accepted in the literature that different anaphoric forms indicate different degrees of accessibility, or givenness (see e.g. Ariel 2001, Prince 1981, Gundel *et al.*, 2003) but variation in anaphoric forms may also reflect pragmatic choice subject to register (see e.g. Kunz 2010). These aspects are however not the focus of the present paper.

What is important here is, first, that coreference chains reflect linguistically that individual conceptual referents play a central role in the textual world. Most often they are not the only participants of the textual world but they contribute to the development of discourse topics. Second, these central referents are related to other concepts in the textual world. All elements in coreference chains that contain a lexical head, be it in the antecedent or in the anaphor, have a potential to intersect with lexical chains and thus to reflect a relation to other concepts in the discourse (see below).

In chains of **lexical cohesion**, the relation between lexical devices of the chain elements is relevant. Our study includes relations between nominal expressions, which may consist of multiple words. Adjacent elements in lexical chains are connected by repetition, as in example (1), or sense relations such as hyperonymy, synonymy and meronymy, antonymy, and relations between named entities (see Martínez Martínez *et al.* 2016 for more details about the sense relations analysed). While grammatical devices in coreference chains are employed to signal identity between individual instantiated referents, lexical devices signal conceptual **similarity** between types of referents. As will be explained in more detail below, the two types of chains may intersect under certain conditions. In any case, lexical chains are an explicit means to create semantic space (see also Hoey 1991) in a text. They indicate linguistically how concepts in the textual world are clustered, evoking associations between types of referents. They are an essential linguistic mechanism to reflect discourse topics. So for an operationalisation of our approach, we regard lexical chains as **explicit discourse topics**. Our aim is to see if, when and to which degree central individual referents evoked by coreference chains contribute to these explicit discourse topics.

A cohesive chain minimally consists of a tie between an antecedent (first element in a chain, see above) and an anaphor (subsequent element(s)). As can be seen in example (1), many chains consist of more than two elements and contain several anaphors. Typically texts contain both types of chains, although to varying degrees. Kunz *et al.* (2016) analysed the variation in cohesive chains with respect to three chain features, the number of elements in chains, the distance between members and the number of different chains, as well as the

interaction of these chain features. They further discussed how this variation impacts on the way discourse topics are structured linguistically in texts: whether the organization of cohesive chains reflect more topic continuity or more variation, whether there is an abrupt change, a continuous modification or a constant interaction of topics. However, chains of coreference and lexical cohesion were analysed separately.

2.3   Chain intersection

In this paper, we focus on the combination of coreference and lexical cohesion in cohesive elements which belong to both chain types. As already stated in Section 1, our concept of chain intersection differs from other approaches in that we do not investigate different chains linked by elements of chains that are realized as different syntactic constituents. Our concept does not consider intra-clausal interaction between different chain elements.

Generally speaking, chain intersection takes place whenever a lexical item that is part of an element in a lexical chain also occurs inside an element of a coreference chain. The two chains 'meet' at the point of the intersection. From the perspective of the coreference chain, our approach accounts for elements in coreference chains in which grammatical and lexical devices are combined and where the lexical device is at the same time integrated in an element of a lexical chain. This is illustrated in example (2).

(2)   Neurobiologists have long known that the euphoria induced by drugs of abuse arises because all these chemicals ultimately boost the activity of <the brain's <u>reward system</u>>: a complex circuit of nerve cells, or neurons, that evolved to make us feel flush after eating or sex... At least initially, goosing <this <u>system</u>> makes us feel good... But new research indicates that chronic drug use induces changes in the structure and function of <the <u>system</u>>'s neurons...

In example (2), we have a coreference chain (marked with brackets) and a lexical chain (marked with underlining) that intersect. Both chains consist of three chain elements. In this case, the intersection starts in the antecedent of the coreference chain, with the lexical head, a compound noun. The nominal expression forms the antecedent of a lexical chain. The conceptual relation of identity in the coreference chain is indicated in the two anaphoric chain elements by two grammatical devices (the demonstrative pronoun *this* and the definite article *the*) to the antecedent *reward system*. These devices do not serve as heads but function as modifiers of the whole chain element. So in this coreference chain not only the antecedent but also the anaphors contain a lexical nominal head.

These lexical heads establish a relation of lexical cohesion. In example (2), the noun *system* is a hyperonym of the antecedent, and the noun in the nominal phrase *the system* is a repetition of the preceding noun. Other possible sense relations are synonymy and hyponymy. So a lexical chain and a coreference chain meet or intersect because the lexical items in their chain members overlap. In the case of example (2), the intersection takes place already in the coreferential antecedent and the lexical antecedent and goes on in the coreferential and lexical anaphors. In this way strong bonds are established inside an explicit discourse topic. There is an explicit linguistic signal indicating not only that an individual referent plays an important role in the textual world but also that this referent is central to an explicit discourse topic. We will see below that there are different types of chain intersection depending on where in the coreference chain and where in the lexical chain the intersection takes place. Moreover, variation in chain intersection concerns the number of intersections. In this study

we are interested in how the two languages English and German as well as the four registers differ with respect to a number of variations discussed in more detail in the next section.

2.4 Features and types of chain intersection

In this section, we define the nine features of chain intersection we analyse in our study on a more refined level. The explanations provided here serve as background information for the overview of the operationalisations given in 3.1. Note that we cannot discuss all the features that may be important in the frame of this paper. We will address them shortly in our outlook and hope to deal with them in the future. We start here by a definition of 'shallow' features which serve as general indicators of the degree of explicit marking of discourse topics by chain intersection and explain how they impact on the continuity of explicit discourse topics in general. We then define features on the basis of which different types of chain intersection are distinguished.

2.4.1 *General features of chain intersection*
Generally, the higher number of intersections between coreference and lexical chains is measured per text, the more central referents contribute to the explicit discourse topics indicated by lexical cohesion in this text. So, the first feature we are interested in is whether there are differences between English and German and between the four registers in our corpus in terms of **the overall number of chain intersections** (feature 1 in 3.1). This feature is obtained by counting the total number of tokens (i.e. nouns and nominal phrases) that are included in intersections of coreference and lexical chains. It is important to note here that an element in a coreference or in a lexical chain may contain several lexical nominal items, e.g. in the case of compounding. They are counted as separate overlaps with this measure. We therefore add the two other features below, in order to obtain the number of intersections per chain elements.

Most of the coreferential antecedents in our corpus contain lexical nouns and therefore have a potential for overlapping chains. This does not always apply to the subsequent elements in coreference chains, the coreferential anaphors, which may consist of pronouns, as can be seen in example (3) below. **The number of coreference anaphors with a lexical head** informs about this general potential of intersection for coreferential anaphors (feature 2 in 3.1).

The next two features are an elaboration of the first one defined above, distinguishing the perspective taken: The length of a chain element may differ in both chain types. For instance, the coreferential antecedent *the brain's reward system* in example (2), which is one single element in the coreference chain, contains three different nouns. Each of these nouns may, however, overlap with one single element of three different lexical chains, or the nouns taken together may overlap with one element of a lexical chain. A coreferential antecedent may even be more complex. We therefore include two additional features, which account for the perspective taken, the extension of the coreferential element or of the lexical element (see below).

Feature 3, **the number of intersections per coreferring element**, counts how many coreferring elements contain intersections with elements in lexical chains (and not whether the same coreference chain is affected by chain intersections again and again).

Furthermore, feature 4, **the number of intersections per lexical chain element**, informs about how many lexical chain elements intersect with coreferring elements.

Feature 5 serves to see how often an intersection takes place in one coreference chain, counting **the number of coreference chains with only one intersection**. One intersection means that only one element (i.e. an antecedent or also one, two or more anaphors) in the

whole coreference chain is responsible for the intersection. The first element in a coreference chain is always involved whenever there is a chain intersection. This implies that the coreferential antecedent is solely responsible if there is one intersection only.

In order to account for differences in the number of intersections per chain and thus between types of intersection, an additional feature is relevant: **the position of the intersection in a coreference chain** (feature 6). We here explore whether the intersection takes place only in the antecedent (first) or additionally also in another element – an anaphor of the coreference chain (non-first).

2.4.2 *Features related to specific types of chain intersection*
Example (2) above and examples (3) and (4) below illustrate three different types of chain intersection, which can be distinguished on the basis of several chain features. Examples (2) and (3) share one characteristic feature distinguishing them from example (4): Chain intersection is observed for the antecedents – the first chain elements – in both chains. This implies that the newly introduced individual referent in the textual world is at the same time used to introduce an explicit discourse topic. Hence, an important feature is **the number of antecedents of coreference chains that intersect with antecedents of a lexical chain**, operationalised with feature 7.

(3)   Well, in Edinburgh most of the water comes from <u><reservoirs></u> which are more towards the a lot of <them> are more towards the Borders, and then <they>'re actually quite old, I think. They first tried to sort out water in Edinburgh, ... But now, a lot of the water comes from the hills on the outskirts of Edinburgh. And then it's brought into <u>holding reservoirs</u>, and then it's brought into the treatment works... And it's the company looks at different ways of helping water companies manage those assets better... if you've got a set of pipes and you've got a set of, say, <u>service reservoirs</u>, which is where you store the clean water, ...

Example (4) shows a type of chain intersection which exhibits more apparent differences from (2) and (3). It demonstrates the importance of one more feature: **the number of antecedent(s) in coreference chain(s) that are anaphors in a lexical chain** (feature 8 in Section 3.1). The example is taken from our English subcorpus of fictional texts.

(4)   This evening I find myself settled here in this comfortable <u>guest house</u> in a street not far from the centre of Salisbury.... <The <u>landlady</u>, a <u>woman</u>> of around forty or so, appeals to regard me as a rather grand visitor ... <She> informed me that <a <u>double room</u>> at the front was available, though I was welcome to <it> for the price of a single. I was then brought up to <this <u>room</u>>, in which ... On inquiring where the <u>bathroom</u> was, <the woman> told me ... I asked <her> to bring me up a pot of tea, and when <she> had gone, inspected <the <u>room</u>> further...

In example (4), we have one lexical chain starting with *guest house* and two coreference chains starting with *The landlady ...* and *a double room*. This type of intersection differs from the ones described above in that the antecedents of the two coreference chains are anaphors in the lexical chain. The two referents pointed at by the two coreference chains play a central role to the explicit discourse topic but they are not used to introduce it. It is introduced by the general concept *the guest house*, which is not itself mentioned again with coreferring expressions later in the text. Instead, other referents, which are mentioned for the first time in the text afterwards and which are related to *guest house* by meronymy are taken up again. In

this way the topic is broadly introduced and configurations of specific referents that are involved in it are specified afterwards.

There is a difference between examples (2) and (3) concerning **the number of anaphors in a coreference chain that are also anaphors in a lexical chain**, which is operationalised with feature 9. In example (2), an individual referent introduces an explicit discourse topic – the first intersection of the two chain types takes place in the antecedents. Moreover the anaphors of the coreference chain, apart from the grammatical devices signalling identity, contain a lexical head and therefore overlap with anaphors of the same lexical chain again. This type of intersection is comparable to Halliday and Hasan (1976, 277ff)'s notion of **reiteration**. It accounts for chains of lexical cohesion where the anaphors, not the nominal lexical head, are combined with (i.e. preceded by) a grammatical item – the definite article or a demonstrative determiner – that indicates coreference. Thus, the individual referent is conceptually enriched and contributes to a very great extent to the explicit discourse topic because there is an intersection between a coreference and a lexical chain not only with respect to the antecedents but also with respect to the anaphors.

In example (3) above, we have a coreference chain with the elements *reservoirs – them – they* and a lexical chain with the elements *reservoirs – holding reservoirs – service reservoirs*. As in example (2), the antecedents of both chains overlap: The referring expression *reservoirs* in example (3) serves as an antecedent for a coreference chain and a lexical chain. What is different is that the intersection holds for the antecedents only (i.e. the first mentions) whereas the rest of the coreference and the lexical chains do not overlap: The coreferential anaphors are made up of coreferential pronouns functioning as nominal heads. The lexical anaphors, which follow the coreferential anaphors are nominal expressions in the plural. So the coreferential anaphors do not contain a lexical element and the lexical anaphors do not contain a grammatical coreferential element. The semantic relation between *reservoirs – holding reservoirs – service reservoirs* is not that of identity, it is a relation of hyperonymy/ hyponymy between different conceptual referents. This type of chain intersection is a typical mechanism to establish a smooth transition from one central referent to other concepts within one explicit discourse topic.

To sum up, we aim to compare the two languages and the four registers with respect to the overall number of chain intersections as well as to the types of chain intersection which depend on where, when and how often a lexical chain 'meets' a coreference chain. As illustrated by the above examples, these variations have an effect on the development of explicit discourse topics. They reflect variation as to the general degree of importance central referents have for explicit discourse topics, at which point and how often they contribute to the discourse topic. With a final comparison of all features we want to explore whether contrasts are greater between languages or between registers, and also compare the registers language-internally to find out if the breadth of variation is greater in English or in German.

3.   Analysis design

In the following sections, we present the set of operationalisations formulated on the basis of the features presented in Sections 2.4.1 and 2.4.2 that we use in our analysis. Apart from that, we describe the corpus resources at hand.

3.1   Operationalisations of chain intersection

For the sake of convenience, we here provide a concise summary of the nine features, which were already introduced in 2.4.1 and 2.4.2, along with the operationalisations used for the

corpus linguistic analysis. This structure serves as a basis for the analyses presented in Section 4 below.

1. Overall number of chain intersections (`nr.inters`): obtained by computing the number of tokens involved in the intersection between coreference chain elements and lexical chain elements, i.e. the total number of overlapping tokens in coreference and lexical chains.

2. Number of coreference anaphors that have a lexical head calculated as the proportion of all anaphors of coreference chains whose head is a lexical item – full nominal phrases (`nr.corefana.lexhead`).

3. Number of intersections per coreferring element (`nr.intersec.percor`): obtained by computing the number of coreferring elements that also include elements of lexical cohesion chains.

4. Number of intersections per lexical chain element (`nr.inters.perlexcoh`): obtained by computing the number of lexical chain elements that intersect with coreferring elements.

5. Number of coreference chains with only one intersection (`nr.corefchain.one.inters.percor`).

6. Average position of intersection (`nr.intersec1st` and `nr.intersec.non1st`): In our study we distinguish between the number of intersections which take place in the first element/ position of a coreference chain (coreferential antecedents) and the number of intersections in a position different from the first position of a coreference chain (coreferential anaphors).

7. Overlapping antecedents (`nr.intersec.ante.ante`): measuring the number of antecedents in coreference chains that are also antecedents (first elements) of lexical chains.

8. Number of antecedents in coreference chains that are anaphors (not the first member) in lexical chains (`nr.intersec.ante.ana`).

9. Overlapping anaphors (`nr.intersec.ana.ana`): number of anaphors in coreference chain that are anaphors (not the first element) in a lexical chain.

In a final step, we analyse the overall variation in the two languages and the four registers comparing them with respect to all the nine features. This is done with the help of correspondence analysis, which was applied, for instance, by Kunz *et al.* (2017) in their study of cohesive features in English and German. The findings will be presented in Section 4.10.

3.2   Corpus design and annotation

The dataset we use for our analysis contains texts of both written and spoken discourse. The written part was extracted from the corpus described in Hansen-Schirra *et al.* (2012), whereas the spoken subcorpus was extracted from the corpus described in Lapshinova-Koltunski *et al.* (2012).

The registers included in our sub-subcorpus are political essays (ESSAY), popular-scientific articles (POPSCI), fictional excerpts (FICTION) and transcribed interviews (INTERVIEW). ESSAY and POPSCI represent written discourse, INTERVIEW represents spoken discourse, whereas FICTION is on the borderline, as it contains both written and spoken elements in the form of dialogues. INTERVIEW and FICTION additionally share narrative elements. The details on the analysed subset are provided in Table 1.

Table 1. Corpus description.

register	EO		GO	
	texts	tokens	texts	tokens
ESSAY	23	27171	20	31407
FICTION	10	36996	10	36778
INTERVIEW	9	30057	12	35036
POPSCI	8	27055	9	32639

The whole corpus is annotated on various levels of lexicogrammar, e.g. parts-of-speech (POS), chunks, clauses, sentences. As mentioned above, the corpus contains manual annotation of various cohesive devices, including coreference (Lapshinova-Koltunski and Kunz 2014) and lexical cohesion (Martínez Martínez *et al.* 2016).[2] The annotation of coreferential devices includes possessive determiners and pronouns, personal pronouns, demonstrative determiners and pronouns as well as coreferential adverbs such as *here* and *there, now* and *then, hereby* and *therewith* (pronominal adverbs). Moreover, annotation of situational coreference (or complex anaphors), where the antecedent consists of a longer textual chunk than just a noun phrase (e.g. a clause, sentence or text paragraph), is also included.

The subset of the corpus presented here additionally provides relational information about lexical chains. Adjacent elements in lexical chains were annotated manually for the type of semantic relation holding between them (e.g. synonymy, antonymy, hyponymy, etc.). All manual corrections and annotations were performed with the tool MMAX2 (Müller and Strube, 2006).

3.3 Analysis techniques

We apply a descriptive data analysis with bar plots as visualisation techniques in Sections 4.1 to 4.9 to observe frequencies of the selected features and to derive general tendencies in their distribution across English and German texts in our data. We use them to relate their frequencies to the total number of chains per language and register to obtain an insight into their distributions. The results are tested for significance using the Pearson's chi-squared test with Yate's continuity correction,[3] with the help of which we can prove if the observed differences between languages (English vs. German) and registers (ESSAY vs. FICTION, etc.) are significant. The Chi-square test measures how well the observed distribution of data fits with the distribution that is expected if the variables are independent.

In Section 4.10, we describe the results of correspondence analysis (CA, Nenadić and Greenacre, 2007) performed for all the features taken together. This technique is explorative and allows us to discover structures in the data in terms of groupings of observations Baayen (2008), for instance, groupings of subcorpora according to their similarities. Besides that, this technique helps to see possible correlation of dependent and independent variables. The

---

[2] More information about the corpus and how to gain access to it can be found at http://hdl.handle.net/11858/00-246C-0000-0023-8CF7-A

[3] A correction for the Chi-square test to use with small data sets.

correlation of our features with the corresponding subcorpora indicates the contribution of these features to the similarities between languages and registers. In CA, distances between dependent and independent variables are calculated and represented in a two-dimensional map, and the larger the differences between subcorpora or texts, the further apart they are on the map. Likewise, dissimilar categories of features are further apart. The correlations between dependent and independent variables are transformed into a set of uncorrelated variables, called principal axes or dimensions. The first two principal axes account for as much variation as possible in two dimensions. In the present paper, this technique will provide a better overview of the interdependence of the features as well as over the breadth of variation between registers and languages.

4.   Analyses, results and discussion

In the present Section, we describe the results of our analysis. As already mentioned in Section 3 above, the section is structured according to the features under analysis that we operationalised in Section 3.1. We will restrict ourselves to a mere description of the findings for each separate feature.

4.1   Overall number of intersections

In Figure 1, we provide a bar plot representing the proportion of the overlapping tokens against the total number of tokens that are elements in coreference and lexical chains.
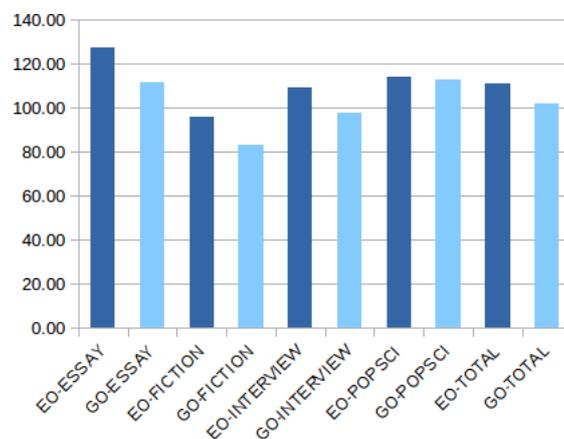

Figure 1. Intersecting tokens in coreference and lexical cohesion chains.

As seen from the plot, English texts reveal a higher number of total intersections when separate tokens (nouns) are counted. This applies to all registers except popular-scientific texts. The latter show similar tendencies in both English and German. The highest number is observed in English essays. Overall, the difference between English and German is significant, as we achieve a very low p-value ($p < 0.00001$, $\chi^2 = 57.369$, df = 3) with Pearson's Chi-squared test.

As noted in Section 2.4.1, an element in a chain may contain several lexical tokens that are nouns. These may intersect with elements of different chains. The higher number of intersections measured in English as compared to German may in part be explained by the fact that all nouns separated by a white space (which is more often the case in English) are

counted as separated tokens whereas compounds without a white space (more common in German) count as one word. More frequent compounding would explain why popular scientific texts and political essays contain a higher number of intersections than the fictional texts and the interviews.[4] The findings for political essays seem to point to the frequent repetition of compounds. This serves as a precision of information (as in scientific texts) and reflects the ideational function of persuasion.

4.2   Number of coreference anaphors with a lexical head

The proportion of all anaphors of coreference chains whose head is a lexical item measured against the total number of coreferring expressions is presented in Figure 2.
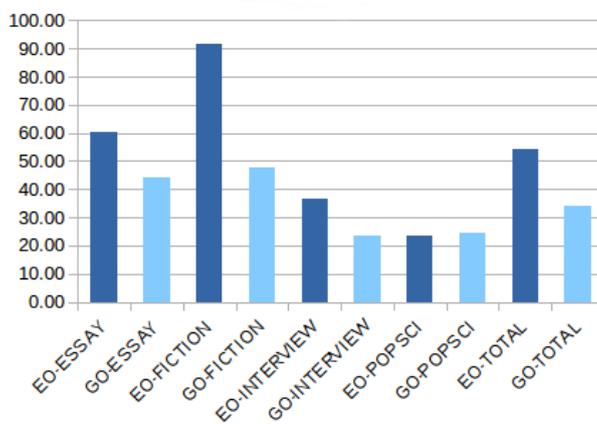

Figure 2. Coreference anaphors with lexical heads.

This number is much higher for all English texts if compared to the German ones, implying that the potential of a coreferential anaphor to take part in an intersection with a lexical chain is generally higher in English than German. In other studies, we could observe two language-specific factors, in addition to the ones mentioned above, which seem to be at play here: a higher number of coreferential anaphors are made up of pronouns and there is a higher amount of extended co-reference with non-nominal expressions (clauses, sentences or textual paragraphs) in German than in English. As for register variation, the popular-scientific articles show very similar, yet low proportions. These texts primarily have an informative communicative purpose, and high information density is expressed by high lexical density. The texts of this register contain many nouns (mostly terms) building chains of lexical cohesion. However, there are few coreference chains and many elements of the lexical chains do not intersect with coreference at all: They are often related by meronymy and repetition and indicate generic relations between types of referents (not instantiated ones) within a discourse topic. The English fictional texts show the highest number of coreferential anaphors with nouns. This is surprising, as fiction contains many coreferential pronouns (a feature of narrative style and spoken language). Looking into the texts reveals that the narrative parts of the English texts frequently contain descriptions of the settings in which the protagonists act, similar to example (4). The different components of the settings are mentioned again but alternate throughout stretches of text. They thus have to be resumed by a

---

[4] The proportion of compound nouns in our corpus comprises 25% in political essays, 19% in popular-science, 16% in fictional texts and 11% in the transcribed interviews.

fully lexical phrase. In German fictional texts, there is less alternation and the focus is more on the main protagonists.

The results of Pearson's Chi-squared test confirms that the differences across registers between the two languages are significant (p < 0.00001, χ^2 = 88.771, df = 3).

4.3 Number of intersections per coreferring element

The proportion of intersecting coreferring elements measured against the total number of (both lexical and coreference) chains is given in Figure 3.
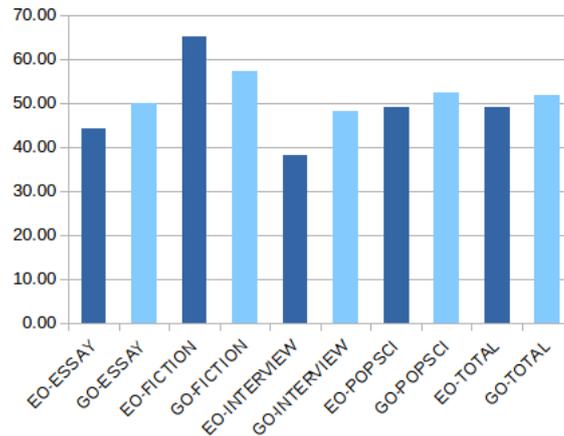

Figure 3. Intersections measured per coreferring element.

We observe an opposite tendency for this feature: German texts show a slightly higher number of intersections. However, an exception is provided by English fictional texts – here we have the highest number of intersections amongst all the texts analysed in both languages. The number in this register is higher than for all others across languages. This seemingly has to do with the exceptional length of the coreference chains as well as the high number of different coreference chains. One possible reason for the general differences to the above findings in terms of general language contrast and register variation has already been suggested above: Chain elements rather than tokens serve as a basis for the feature here. Hence, another reason for the high value for fiction seems to be the lower number of multiple nouns contained per coreferring element. So in German, more coreferring elements overlap with lexical chains elements than in English although the lexical potential in anaphors is lower. This may even strengthen the explicit effect, from a contrastive perspective. Individual referents thus seem to be connected more strongly to an explicit discourse topic in German than English, except for the fictional texts. The difference between the two languages is also significant in this case (p < 0.00001, χ^2 = 48.843, df = 3).

4.4 Number of intersections per lexical chain element

Figure 4 illustrates the proportion of intersecting lexical chain elements calculated against the total number of all chains in the corpus. Since this feature is measured on the basis of lexical cohesion elements rather than coreference chain elements, it conveys a different perspective on chain intersection than that in 4.3 above.

In this case, we observe a similar tendency as in Section 4.3, if all German and English texts are considered: The German texts use more intersections than the English ones, except

for the fictional texts. This difference is also significant (p < 0.00001, $\chi^2 = 81.064$, df = 3). Thus, more elements of lexical chains are connected to a central referent in German than in English. However, the register-specific figures show that political essays and popular-scientific texts contribute the most to the language contrast. In the latter register, language contrast is more pronounced for this feature than for all other features. This time, the numbers for the fictional texts lie below those of the other registers in both languages.
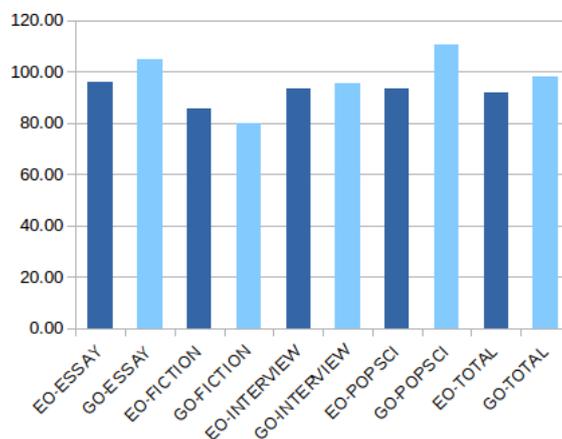

Figure 4. Intersections measured per lexical chain element.

4.5 Number of coreference chains with only one intersection

In Figure 5, we provide the proportion of coreference chains with one intersection measured only against the total number of intersections.
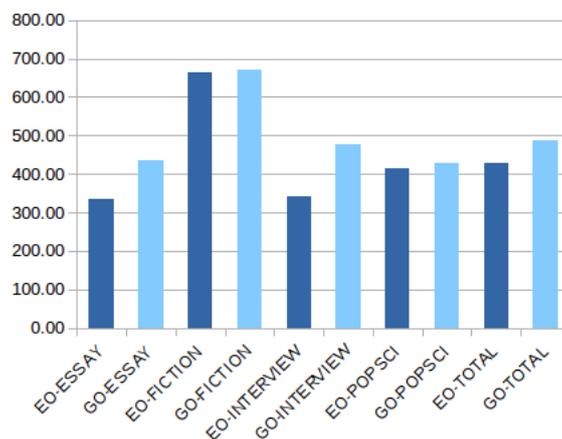
Figure 5. Chains with one intersection only

The findings show a higher number of chains with only one intersection in German compared to English. This means that there are more cases in German than English where only the coreferential antecedent but not the rest of the coreference chain intersects with a lexical chain, as shown in example (3). We observe similar tendencies for both languages in popular-scientific and fictional texts. The fictional texts outperform all other registers again. This may generally be connected to two facts: First, a low number of multiple nouns in noun phrases and second, a very high number of different chains of both chain types in both languages

(based on the findings by Kunz *et al.* 2016). In interviews and political essays, the number of coreference chains with only one intersection is higher in German than in English. Overall, languages turn out to differ significantly across registers (p < 0.00001, χ^2 = 46.478, df = 3).

4.6   Average position

The average position of the intersections in a coreference chain is defined as a binary category: first and non-first. In Figure 6, we present the proportions of these two types of intersections calculated against the total number of intersections.

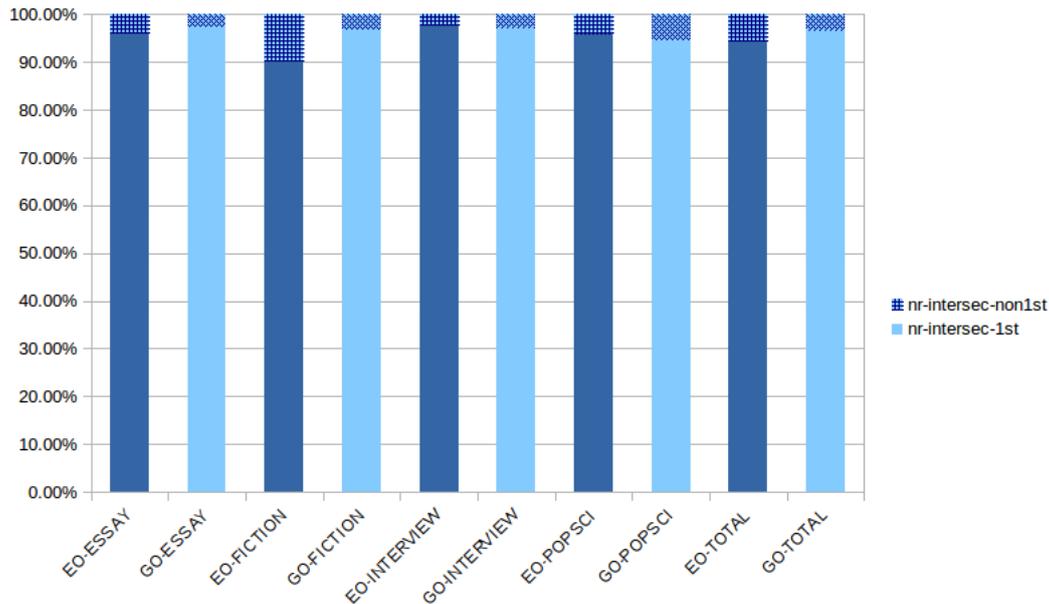**Figure 6. Average intersection position.**

We generally note a much higher number of intersections in the first position (the first chain elements) than in all other positions of coreference chains. This is of course due to the fact that the coreferential antecedent is always involved whenever chain intersection takes place, no matter which type of chain intersection. More interestingly, the proportion of non-first chain element intersections in relation to other chain elements is higher in English than in German, i.e. more coreferential anaphors are involved in English than German, leading to an intersection type such as in example (2). The English fictional texts contain the highest numbers of all texts, whereas the English interviews show the lowest frequencies. The significance test shows that the difference between English and German texts is significant (with a p-value of ca. 0.002, which is considerably higher than the results for the other features, but still below 0.05).

4.7   Overlapping antecedents

Figure 7 presents the proportion of overlapping antecedents calculated against the total number of intersections in English and German registers. In German fictional texts, interviews and political essays, antecedents tend to overlap more often than in the corresponding registers in English, which means that explicit discourse topics are introduced more often with central referents here. In popular-scientific articles, whose discourse structure is more standardised than in the other registers, we observe a similar number of

overlapping antecedents in both languages. However, the overall difference between different registers in both languages is significant (p < 0.0001). Within each language, fictional texts reveal most frequent cases of an overlap, with the highest number again shown for the English fictional texts and the lowest for English Interviews.
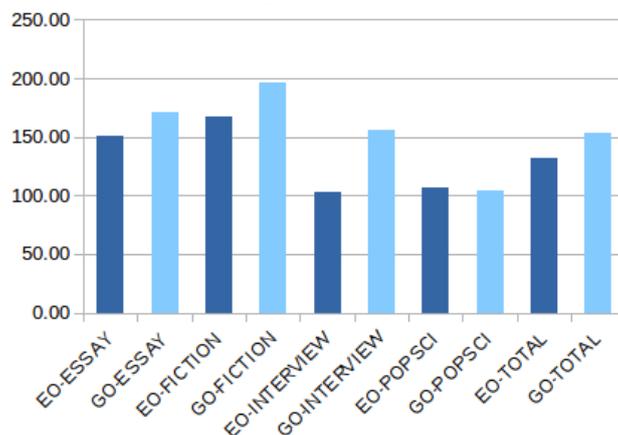
Figure 7. Overlapping antecedents in coreference and lexical chains.

4.8   Number of antecedents of coreference chains that are anaphors of lexical chains

Figure 8 displays this proportion which is also measured against the total number of intersecting elements. What has to be noted first is that the numbers for this feature are generally higher than those for overlapping antecedents for all registers in both languages. The degree of the difference is register-specific. For instance, it is less pronounced in the political essays.
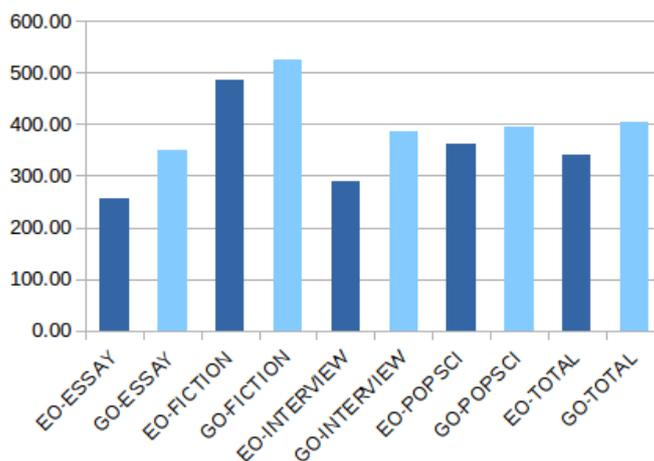
Figure 8. Intersections between coreference antecedents and lexical chain anaphors.

Second, in all German texts, there are more coreferential antecedents that intersect with anaphors (rather than with antecedents in lexical chains) than in English. Therefore, explicit discourse topics are introduced more often with a lexical antecedent in an earlier stretch of text, preceding the whole coreference chain, as exemplified in (4). This lexical chain element

is related to the coreferential antecedent by similarity of sense. It implies that central referents less often serve to introduce explicit discourse topics and more often play a role as topics unfold. Overall, we observe a significant difference between the languages (p < 0.00001, χ^2 = 31.74, df = 3).

4.9 Overlapping anaphors

The proportion of all overlapping anaphors in coreference and lexical chains presented in Figure 9 is measured against the total number of chain intersections.
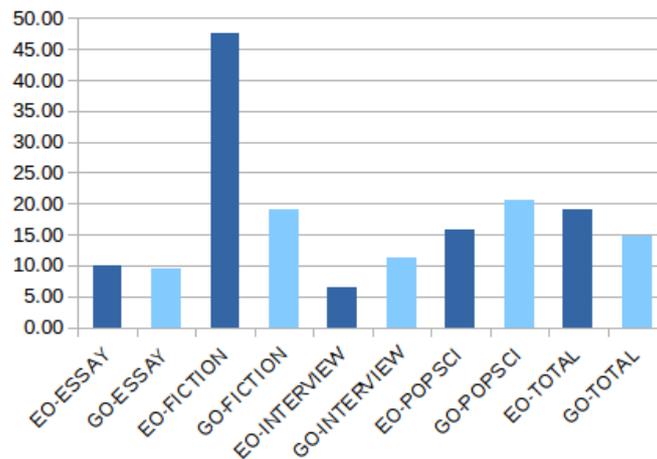

Figure 9. Overlapping anaphors in coreference and lexical chains.

Generally speaking, the numbers of overlapping anaphors is very low in both languages, when compared to the numbers for overlapping antecedents and coreferential antecedents that intersect with anaphors in lexical chains. Hence the intersection type as shown in example is not very frequent. The feature also indicates significant differences between the two languages (p < 0.001, χ^2 = 16.72, df = 3) if analysed across all registers. But the tendency of German numbers lying below those of English holds only for two registers, ESSAY and FICTION. In addition, we observe different rankings of registers within the languages: while in English, fictional texts show the highest amount of overlapping anaphors, popular-scientific articles occupy the first position in German. These texts have more central referents with an important role for the explicit discourse topic that are also conceptually enriched throughout the text.

4.10 Overall variation

Figure 10 illustrates the output of the correspondence analysis. As seen from the two-dimensional plot (which explains 89.1% of variation in our data), we observe heterogeneous tendencies across languages and registers. English and German popular-scientific texts seem to be very similar, as they are situated very close to each other on the x-axis and even overlap on the y-axis. This coincides with the tendencies we observed for these texts analysing individual features. Interestingly, the x-axis separates fictional texts in both languages from the other registers, which again concurs with the results observed for individual features – fictional texts in both languages often behave differently from the other texts. However, they do reveal some language-specific features found along the y-axis. Correspondence analysis does not show a clear distance between languages, i.e. a consistent language contrast: It is

rather observed for each register separately. On the y-axis we observe a very heterogeneous grouping of the registers: EO POPSCI and GO POPOSCI are close to GO FICTION, and GO INTERVIEW a bit further away; GO ESSAY and EO FICTION almost overlap, and EO INTERVIEW and EO ESSAY here show more resemblance to each other than to the other subcorpora. Whether the breadth of variation is bigger in one language than the other cannot be told on the basis of these data, as it heavily depends on one register.

Figure 10. Correspondence analysis.

5. Interpretative summary

This section is dedicated to the interpretation of the results described in Section 4 in terms of discourse topics, where the findings of the features will be related to each other.

5.1 General features

From a general perspective, the findings for the first two features seem to be in contrast to the other features. This can mainly be attributed to the fact that these features are obtained on a lower linguistic level than the others, on the basis of lexical tokens, and more specifically, nouns. First, we note a higher number of intersections per token in English than in German. However, this implies only that single nouns in coreference chains contribute more often to the development of explicit discourse topics in English than in German (with the exception of POPSCI). This tendency seems to be in line with the overall potential of anaphors to enter into such an intersection, which is again higher in English than in German (the exception again is POPSCI). As explained above, there are several influencing factors: More grammatical heads with a demonstrative pronoun and more extended anaphors are used in German than in English to establish coreference. In addition, the nouns in compounds are counted separately more often in English than in German.

For the other general features, we note an opposite tendency with respect to language contrast: First, German exhibits a higher number of intersections per coreferring element and per lexical chain element. A comparison of the findings in Sections 4.3 and 4.4 reveals that there is a higher number of intersecting lexical chain elements than of intersecting coreferring elements, both in English and in German. This mainly has to do with the fact that the

extension in coreference chains is longer than those of lexical chain elements. Both measures taken together show the same tendency, namely that central referents seem to be more relevant to explicit discourse topics in German than in English texts. The next two features reveal that the higher number of intersections in German mainly stems from intersections in which coreferential antecedents are involved: The higher number of coreference chains with only one intersection in German along with the higher number of intersections in the first position as compared to English means that more topics in German are introduced by a central referent directly. This observation finds further support by the three remaining features, which are more closely linked to specific types of intersection and thus to variation in the development of explicit topics by central referents.

5.2   Types of chain intersection

In both languages, the number of coreferential antecedents that intersect with anaphors in lexical chains is higher than the number of overlapping antecedents. In addition, overlapping antecedents play a more important role for intersection than overlapping anaphors. This entails a general ranking with respect to the types of intersection: More often a smooth transition is preferred, in which the explicit discourse topic is not introduced by a central referent directly. The antecedent of a lexical chain introduces a configuration of concepts. The central referent established by coreference, which constitutes one important concept within this configuration, is mentioned later. Explicit discourse topics that are introduced by a central referent directly are less frequent. In any case, further continuity in explicit discourse topics is not upheld to a great extent by central referents, as there are few anaphors of coreference chains overlapping with anaphors in lexical chains. Hence, most explicit discourse topics are reflected by lexical relations without conceptual identity being involved.

As for language contrast, we observe more overlapping antecedents in German than in English, hence more discourse topics are introduced by important individual referents directly. The same tendency however applies to the number of antecedents of coreference chains that are anaphors in lexical chains. Again we find more intersections in German than English, which possibly results from the fact that German has more intersections than English in general. In these constellations, the discourse topic is introduced by a nominal expression that is a holonym or a meronym of the following central referent in most cases. Contrary to the other two features, the number of overlapping anaphors is higher in English than in German.

Furthermore, our findings show that there is at least one register which is in contrast to the overall tendency observed in terms of language contrast, for most of the features. General features and particular types of intersection heavily depend on the register. They may thus relate to specific configurations of field, tenor and mode. Quite interestingly, the tendencies for registers in the two languages sometimes coincide but sometimes they do not. The fictional texts quite often stand out and seem to bear least resemblance to the other registers. Hence chain intersection and its impact on discourse topics seem to be a reflection of the distinction between fiction and non-fiction. Within the fictional texts, we note a dramatic difference with respect to the number of overlapping antecedents, the numbers for the English texts being much higher than those for German. This difference contributes the most to the general language contrast observed for the feature. A reason for this could be that settings and interaction between objects play a greater role in English, and main protagonists are favoured in German. The German fictional texts also contain more dialogic parts than the English ones. However EO FICTION is in even sharper contrast to EO INTERVIEW, a register within the same language, which contains a very low number of overlapping anaphors. EO INTERVIEW is also the register with the lowest number of overlapping antecedents and a

relatively low number of intersections between coreference antecedents and lexical chain anaphors, so there is not much chain intersection in general. This may be caused by the mode of spoken language in that there is a more frequent use of grammatical anaphors of coreference and a frequent occurrence of extended reference. It may also stem from colloquial style. The differences across languages with this spoken register are greater than for the written registers ESSAY and POPSCI, the latter being more standardised than all other registers.

6.   Outlook

In this study we could not integrate all features of coreference and lexical chains that are relevant to the development of discourse topics. These deserve further exploration in the future. For instance, we only differentiate intersections that are contained in coreferential antecedents vs. coreferential anaphors, but we do not specify further which position the anaphor has in the coreference chain and the lexical chain (the second, the third or another element). This would, however, be interesting for longer chains and inform about 'interrupted' intersection, which may have an interpersonal function in argumentative and persuasive texts (e.g. introduction and synopsis). Moreover, a more precise interpretation of the role of intersections to explicit discourse topics can be obtained if the features introduced in this study are related to the other chain features (as it was discussed in Kunz *et al.* 2016): chain length, distance in chains and number of different chains. The features of this study have to be brought together with the features of other models dealing with chain interaction and cohesive harmony, as mentioned in Section 1.

References

Ariel, M. 2001. Accessibility theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*, T. Sanders, J. Schliperoord and W. Spooren (eds), 29–87. Amsterdam/Philadelphia: John Benjamins.

Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Barzilay, R. and Elhadad, M. 1999. Text summarizations with lexical chains. In *Advances in Automatic Text Summarization*, I. Mani and M. Maybury (eds), 111–121. Cambridge, MA: MIT Press.

Brown, G. and Yule, G. 1983. *Discourse Analysis*. [Cambridge Textbooks in Linguistics.] Cambridge: Cambridge University Press.

Chafe, W.L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and Topic*, C. N. Li (ed.), 25–55. New York: Academic Press.

Clark, H. H. 1975. Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.) *Proceedings of the 1975 Theoretical Issues in Natural Language Processing*, 169–174. New York: Association for Computing Machinery.

Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S. and Weischedel, R. 2004. Automatic content extraction (ace) program – task definitions and performance measures. In *Proceedings of LREC-2004: 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Grosz, B., Weinstein, S. and Joshi, A. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2): 203–225.

Gundel, J., Hegarty, M. and Borthen, K. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information*, 12(3): 281–299.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London/New York: Longman.

Hansen-Schirra, S., Neumann, S. and Steiner, E. 2012. *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin/New York: de Gruyter.

Hasan, R. 1984. Coherence and cohesive harmony. In *Understanding Reading Comprehension*, J. Flood (ed.), 181–219. Delaware: International Reading Association.

Hasan, R. 1985. The texture of a text. In *Language, Context and Text: Aspects of Language in a Social-semiotic Perspective*, M.A.K. Halliday and R. Hasan (eds), 70–96. Oxford: Oxford University Press. Reprinted 1989.

Hoey, M. 1991. Another perspective on coherence and cohesive harmony. In *Functional and Systemic Linguistics: Approaches and Uses*, E. Ventola (ed.), 385–414. Berlin and New York: Mouton de Gruyter.

Kerremans, K. 2014. *Terminological Variation in Multilingual Europe. The Case of English Environmental Terminology Translated into Dutch and French*. PhD thesis, Vrije Universiteit Brussel.

Kunz, K. 2010. *Variation in English and German Nominal Coreference: A Study of Political Essays*. PhD thesis, Saarland University, Frankfurt am Main.

Kunz, K., Degaetano-Ortlieb, S., Lapshinova-Koltunski, E., Menzel, K. and Steiner, E. 2017. Gecco – an empirically-based comparison of English-German cohesion. In *Empirical Translation Studies: New Methodological and Theoretical Traditions*, volume 300 of *TILSM series*, G. De Sutter, M.-A. Lefer and I. Delaere (eds), 265–312. Berlin and New York: Mouton de Gruyter.

Kunz, K., Lapshinova-Koltunski, E. and Martínez Martínez, J. M. 2016. Beyond identity coreference: Contrasting indicators of textual coherence in English and German. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes co-located with NAACL 2016*, San Diego, California. ACL.

Lapshinova-Koltunski, E. and Kunz, K. 2014. Annotating cohesion for multilingual analysis. In *Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Reykjavik, Iceland. LREC.

Lapshinova-Koltunski, E., Kunz, K. and Amoia, M. 2012. Compiling a multilingual spoken corpus. In *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*, H. Mello, M. Pettorino and T. Raso. (eds), 79–84, Firenze: Firenze University Press.

Martin, J. 2015. Cohesion and texture. In Tannen, D., Hamilton, H. and Schiffrin, D. (eds), *The Handbook of Discourse Analysis*, Blackwell Handbooks in Linguistics, 61–81. Malden, Mass: Wiley.

Martínez Martínez, J. M., Lapshinova-Koltunski, E. and Kunz, K. A. 2016. Annotation of lexical cohesion in English and German: Automatic and manual procedures. In *Proceedings of the Conference on Natural Language Processing (Konferenz zur Verarbeitung natürlicher Sprache) – KONVENS-2016*, Bochum, Germany.

Mesgar, M. and Strube, M. 2015. Graph-based coherence modeling for assessing readability. In *Proceedings of STARSEM 2015: The Fourth Joint Conference on Lexical and Computational Semantics*, 309–318, Denver.

Morris, J. and Hirst, G. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1): 21–48.

Müller, C. and Strube, M. 2006. Multi-level annotation of linguistic data with MMAX2. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, S. Braun, K. Kohn and J. Mukherjee (eds), 197–214. Frankfurt a.M.: Peter Lang.

Nenadić, O. and Greenacre, M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: The *ca* package. *Journal of Statistical Software*, 20(3): 1–13.

Poesio, M., Vieira, R. and Teufel, S. 1997. Resolving bridging references in unrestricted text. In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, ANARESOLUTION '97, 1–6, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pradhan, S., Ramshaw, L., Mitchell, M., Palmer, M., Weischedel, R. and Xue, N. 2011. CoNLL-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*.

Prince, E. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, P. Cole (ed.), 223–255. New York: Academic Press.

Song, W., Fu, R., Liu, L. and Liu, T. 2015. Discourse element identification in student essays based on global and local cohesion. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2255–2261, Lisbon, Portugal. Association for Computational Linguistics.

Strube, M. and Hahn, U. 1999. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3): 309–344.

Tanskanen, S.-K. 2006. *Collaborating towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam/Philadelphia: John Benjamins.

Technologies, B. 2006. *Coreference Guidelines for English OntoNotes – Version 6.0*. Linguistic Data Consortium. BBN Pronoun Coreference and Entity Type Corpus.

Todd, R. 2003. *Topics in Classroom Discourse*. PhD thesis, University of Liverpool.

Todd, R. 2016. *Discourse Topics*. Pragmatics & Beyond new series. Amsterdam/Philadelphia: John Benjamins.

Zikánová, Š., Hajičová, E., Hladká, B., P., J., Mírovský, J., Nedoluzhko, A., Poláková, L., Rysová, K., Rysová, M. and Václ, J. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. ÚFAL, Prague, Czech Republic.

Authors' addresses

Kerstin Kunz
Institut für Übersetzen und Dolmetschen
Universität Heidelberg
Plöck 57a
69117 Heidelberg

kerstin.kunz@iued.uni-heidelberg.de

Ekaterina Lapshinova-Koltunski
Department of Language Science and Technology
Universität des Saarlandes
Building A2.2, Room 1.02
Campus, 66123 Saarbrücken

e.lapshinova@mx.uni-saarland.de

Marked themes in English and Norwegian academic texts in the field of didactics[1]

Sylvi Rørvik, Marte Monsen

Inland Norway University of Applied Sciences (Norway)

Abstract: This study investigates the use of marked themes in the field of didactics. The material comprises research articles written in L1 English and Norwegian, and L2 English. The aim is to shed light on contrastive differences between Norwegian and English that may be used to inform novice writers about central textual features of academic texts within the field of didactics, and about potential transfer-related features that might give their texts a "foreign accent". The results show that there are contrastive differences in the realization of and meanings expressed by marked themes, but that these differences cause few problems for L2 writers. The study further supports findings from previous research that show the importance of text type and academic discipline for thematic structure.

Keywords: thematic structure, marked themes, contrastive analysis, English/Norwegian

1.   Introduction

Text structure is a recurring topic in the literature, and studies have suggested that the structure of a text may be influenced by text type (Davies, 1997, Hasselgård, 2014), the discipline within which the text is written (Hyland, 2000, North, 2005b, Ebrahimi *et al.,* 2014, Ebrahimi, 2016, Babaaii *et al.*, 2016), the language in which it is written (Hasselgård, 2005) and the language background of the writer (Hasselgård, 1998, Rørvik, 2013). It has also been shown that different languages may have different means of realization for the same text-structuring function (Moyano, 2016). On the basis of contrastive studies, conflicting evidence has emerged regarding the existence of differences in this area in different text types: for instance, in a comparisons of fiction texts in English and Norwegian it has been shown that Norwegian permits a higher frequency of non-subjects in initial position than does English (Hasselgård, 1998, 2004, 2005). On the other hand, a study of argumentative newspaper texts found no contrastive difference as regards the frequency of marked themes (Rørvik, 2013: 51-52). Given this variability, contrastive corpus-based studies of academic writing in a range of disciplines and including all text types within the

---

disciplines are essential as a starting point for research-based teaching of academic writing (cf. Carter-Thomas & Rowley-Jolivet, 2008), not least because such studies may contribute to the definition of individual disciplines: if the writing within one putative field is found to differ from the writing in other fields with respect to the variable(s) under investigation, one may argue that the examined field has a set of writing conventions to which authors must adhere and of which novice writers should be made aware (Gosden, 1992: 207, 221). It is of course possible for a discipline to encompass many different text types, and it is therefore desirable to broaden the knowledge base regarding writing conventions within a field or discipline through a range of studies examining the ongoing development of text types published within the field (cf. e.g. Gray, 2015). For pedagogical purposes it is also important to investigate whether transfer from the writers' L1 can influence their production in the L2, and it is therefore necessary to carry out contrastive studies of material that includes native-speaker texts in the learners' L1 as well as in the L2. It is of course an added bonus if one can also include L2 texts, i.e. non-native target-language texts, as this enables the study of transfer-related difficulties in practice and not just as a potentiality based on differences between native-speaker L1 and native-speaker L2 (cf. Lado, 1957: 70). The aim of such a comparison would thus be to move from the "strong" hypothesis of contrastive analysis to a weaker version that can then be complemented by a contrastive interlanguage analysis at a later stage (cf. the discussion in Gilquin, 2000/2001, 2008 and Granger, 1996).

The present study is a first attempt at providing insights about contrastive differences between Norwegian and English that may be used to inform novice writers about central textual features of academic texts within the field of didactics, and about potential transfer-related features that might give their texts a "foreign accent" (Shaw, 2004: 81). To this end, a contrastive study has been carried out of published academic texts written in L1 Norwegian and L1 and L2 English. The design of the study is thus similar to that employed by Shaw (*ibid*.), but the discipline is different, as well as the languages involved and language backgrounds of the writers: Shaw investigated articles within the field of economics written by Danes in Danish and English and in English by native speakers, or at least articles written by authors with Anglo-Saxon names (*ibid*.: 77; see further Section 2 below). More specifically, the focus of the present study is on the frequency and realization of marked themes (roughly defined as non-subjects in initial position, see further Halliday, 2004: 64-105), as well as the meanings expressed by these. The following research questions are addressed:

1. What are the similarities and differences between L1 Norwegian and L1 English writers in the field of didactics as regards the following features:

- the frequency of marked themes

- the functions of marked themes

- the realizations of marked themes

- the meanings expressed by marked themes

2. To what extent have Norwegian writers of L2 English in the field of didactics managed to adapt to English conventions regarding the use of marked themes? Is there any evidence of transfer from Norwegian in their use of marked themes?

The remainder of this article is structured as follows: Section 2 presents an overview of relevant previous research, while Section 3 defines the central theoretical concepts and

analytical framework employed in the present study. The material and method are introduced in Section 4, and Section 5 contains the discussion of the results. Finally, Section 6 comprises a brief summary and conclusion.

2.   Previous research on thematic structure

Previous studies typically focus on text structure or thematic structure in general, and not marked themes in particular, so this is necessarily reflected in our presentation of previous research, which focuses on studies discussing factors influencing thematic structure, such as text type, field/academic discipline, language, and the language background of the writers. It should also be noted that not all studies define theme in the same way as the present study does, nor is the unit of analysis always the same (or even always clearly stated). In some cases we have therefore "translated" the terminology used in previous studies to make it fit the terminology applied in the present study. For instance, Shaw (2004) investigates sentence-initial elements, and uses the orthographic sentence as his unit of analysis, not the T-unit.

Davies (1997) suggests that theme analysis may help distinguish between different text-types and genres. She investigates 14 texts within genres like "textbooks", "academic papers" and "the modern novel", and illustrates that different text types make use of different themes, but without formulating any clear-cut distinctions between different text types. Further support for the claim that text type influences thematic structure can be found in a study of initial adjuncts in news and fiction (Hasselgård, 2014), where it is shown that these two genres differ both as regards the frequency of initial adjuncts, which are more common in fiction than in news, and in the meanings expressed by them, with the most prominent difference being that initial adjuncts expressing 'time' are more frequent in fiction than in news.

Several previous studies have also shown that the discipline within which the texts are written may influence text structure. North (2005a) actually questions the extent to which communication skills can be transferred across disciplines, and in a study of L1 English student texts she found that students from the fields of arts and sciences differed in their use of theme according to their subject background (North 2005b). Differences in the frequency/proportion of marked themes have also been found to exist between published research articles from various disciplines. For instance, Heng and Ebrahimi (2012) found that abstracts in research articles in applied linguistics had a lower proportion of marked themes than research articles in economics. Similarly, Babaii *et al.* (2016) investigated the frequency of marked themes in research articles in mechanical engineering, biomedical engineering, horticulture, and environmental sciences, concluding that marked themes are more frequent in mechanical engineering than in the three other disciplines. One final example comes from a study of research-article introductions (Valipour *et al.*, 2017), where the results indicate that chemistry exhibits a higher proportion of marked themes than does linguistics and software engineering.

There are also previous studies that investigate the influence of academic discipline on the meanings expressed by marked themes, for instance the above-mentioned Heng and Ebrahimi (2012), and Ebrahimi (2016). Heng and Ebrahimi (2012) show that linguistics and economics are fairly similar regarding the proportions of marked themes expressing 'time', 'cause', 'means', and 'condition', but that marked themes express 'contrast' more frequently in economics than in linguistics. Ebrahimi (2016) finds that chemistry has more themes expressing 'time' than does psychology and applied linguistics, and that 'condition' is more frequently expressed by marked themes in applied linguistics than in psychology and

chemistry. A further finding is that 'location' is a frequent meaning in both applied linguistics and psychology, but less frequent in chemistry.

A further set of studies investigate the frequency of marked themes and meanings expressed by marked themes either in one discipline or in several, but without distinguishing between them. We mention these briefly here because they involve disciplines different from the one investigated in this study, and the findings may therefore be compared to ours in an attempt to pinpoint whether and, if so, how, the writing conventions in didactics differ from those of other disciplines. First, Gosden (1992) found that the proportion of marked themes in a corpus consisting of research articles from physics, chemistry, and biological sciences was approximately 32%. It should be noted, however, that Gosden's definition of marked themes appears to include some structures that are are excluded from the framework applied in the present study (principally this concerns adverbial conjuncts). Furthermore, Gosden found that the most frequent meanings expressed by marked themes were 'place', 'contrastive'/'concessive' meanings, 'time', and 'cause'. A second study worth mentioning investigates theme in the method and discussion sections of biology journal articles (Martínez, 2003). In this material approximately 17% of the themes are marked, and nearly all of these are circumstance adverbials. The most frequent meanings expressed are 'purpose', 'time', and 'place'. The results presented by Gosden (1992) and Martínez (2003) thus support the idea that academic discipline influences the proportion of and meanings expressed by marked themes.

We now turn to relevant previous contrastive studies. Of particular relevance for the present study are studies of fiction material from the English-Norwegian Parallel Corpus that indicate a greater tolerance for non-subjects in initial position in Norwegian than in English (Hasselgård, 1998, 2004, 2005). On the basis of these studies, it would seem that there are clear cross-linguistic differences between Norwegian and English, in that Norwegian writers use more marked themes than do English writers. However, an investigation of texture in English and Norwegian argumentative texts found no such contrastive difference, suggesting that text-type differences play a part contrastively as well (Rørvik, 2013). Similarly, Shaw (2004) found more initial circumstance adverbials (i.e. one type of marked theme) in Danish economics articles than in English economics articles. Interestingly, Shaw also included articles written in English by Danish writers, and found that this L2 English material "is more like Danish than like English written by [native speakers]" (*ibid.*: 79) as regards features that correspond to the marked themes investigated in the present study. Danish and Norwegian are very closely related languages, so Shaw's results form a very important basis for comparison with the results of the present study. The final contrastive study to be included in this overview has already been briefly introduced above, where its results were mentioned as evidence for text-type/genre differences: Hasselgård (2014) not only examined differences between news and fiction, but also between English and Norwegian texts belonging to these two genres/text types. She found slightly higher proportions of initial adjuncts in Norwegian than in English, and also some differences in the types of meanings expressed by these adjuncts: in the fiction material, 'time' was more frequently expressed in Norwegian than in English. This difference also occurred in news, but was less pronounced there. In addition, the meaning of 'manner' was much more frequent in Norwegian news than in English news. Hasselgård also looked at the realization of adverbials, and found that there seem to be more initial adverbials realized by single adverbs in Norwegian (both genres) than in English. The same is true for prepositional phrases.

As we have seen, previous studies have indicated that thematic conventions may vary from text type to text type, and there may also be different expectations in different academic disciplines. This comes in addition to the potential cross-linguistic differences between Norwegian and English. Hence, in order for teachers to be able to teach academic writing

within the field of didactics, an investigation into the specific conventions of this field is necessary. Additionally, by adding L2 English material, we can examine whether Norwegian academics adapt their writing style when writing in a foreign language.

3.   Terms and definitions

This section presents the analytical framework and definitions employed in the present study.

3.1   Theme and the unit of analysis

The present study employs a definition of theme that is largely in line with that of Halliday (2004: 79), according to which theme comes first in the clause and "ends with the first constituent that is either participant, circumstance or process". A strict application of this definition would entail the clause as the unit of analysis, but, in line with previous studies in the field (cf. e.g. Hasselgård, 1998, 2004, 2005 and Rørvik, 2013), we have chosen to use T-units instead of clauses as the basis for our analysis. T-units are defined as a main clause together with any associated dependent clauses (Hunt, 1965). This has two implications for the thematic analysis, both of which are illustrated in example (1):

(1)   a. *She* knew that videos were being taken of Lila at the setting

b. *but unless a teacher took the time to show them to her* she did not see them. (L1Eng)[2]

First, the choice of T-units as the unit of analysis means that any sentence with more than one main clause will be split into several T-units, each with its own theme. Thus, the sentence in example (1) consists of two T-units (a and b), since there are two main clauses, and each of these has a theme (in italics). Secondly, a T-unit can have a dependent clause as theme, which is illustrated in the second T-unit in example (1). A strict Hallidayan approach here would mean that we analyzed the themes in the dependent clause and the main clause separately, but such a micro-level approach would mean that we would lose track of the development of the text.

Next we will introduce the distinction between unmarked and marked themes, using declarative sentences as examples (for details regarding thematic structure in other sentence types, see Halliday, 2004: 75-79). In declarative sentences, the unmarked choice of theme is to conflate it with the grammatical subject, as illustrated in example (2):

(2)   *Samisk* er et offisielt språk i Norge, […] (L1Nor)
        "*Sami* is an official language in Norway, […]"

In example (2) the first constituent is the grammatical subject *Samisk*, which constitutes an unmarked theme. By contrast, example (3) has the circumstance/adverbial *In the United States (U.S.)* in initial position:

---

[2] Each example has been provided with a tag identifying the corpus from which it has been extracted.

(3)   *In the United States (U.S.),* assessment has received a great deal of attention from both the popular press (Ravitch, 2011) and education scholars (Darling-Hammond, 2014). (L1Eng)

Since the first element in example (3) is not the subject, the theme in this T-unit is considered to be marked.

3.2   Functions of marked themes

There is only a limited set of options as regards the functions of marked themes. In cases where the marked theme is a circumstance/adverbial, the function will be 'Adjunct', as exemplified in (4):

(4)   *Dersom jeg kun hadde sittet og observert i klasserommet*, hadde det ikke vært mulig å få så detaljerte analyser. (L1Nor)
        "If I had only conducted classroom observation, it would not have been possible to get such detailed analyses."

Whenever the marked theme is realized by a non-adverbial (i.e. an object or predicative of some kind), the function is 'Complement'. In example (5), for instance, the author has fronted the subject predicative:

(5)   *Perhaps more important to the participants in this study than rank or location,* was that engagement in CoP resulted in the creation of trusting relationships which served to break their feeling of academic isolation. (L1Eng)

The third potential function of marked themes is 'Predicator', used for any fronted process (i.e. verbal). However, this did not occur in our material, so it will not be discussed further here.

3.3   Realizations of marked themes

Numerous structures functioning as marked themes were found in the material, but most with very low frequencies. We have therefore chosen to limit the present overview to the three most frequently occurring constructions: dependent clauses, prepositional phrases, and adverb phrases (includes single adverbs). These categories do not require a lot of explanation, but for the sake of completeness we illustrate each of them here. First, example (6) illustrates a theme realized by a dependent clause:

(6)   *Når det gjelder metoder* finner en så godt som alle tilnærminger innen samfunnsforskning. (L1Nor)
        "When it comes to methods you can find practically every approach within social sciences."

In example (7) the theme is realized by a prepositional phrase:

(7)   *At the pedagogical end of the continuum*, we hold that teachers conceive of assessment as serving the purpose of informing instruction and improving student learning. (L1Eng)

Finally, the theme in example (8) is realized by an adverb phrase:

(8)   *More recently,* Brown, Chaudhry, and Dhamija (2015) investigated 1,645 Northern Indian secondary (primarily private) school teachers' conceptions of assessment. (L1Eng)

Having defined the concepts related to the form, function, and realizations, we now turn to the meanings expressed by the marked themes.

3.4   Meanings of marked themes

As was the case with the realizations of marked themes, a range of different meanings were expressed in the corpora. In this paper we will include those that were among the top three in each subcorpus, resulting in the following list of meanings (all adverbial in nature): 'time', 'place', 'condition', 'instrument', 'concession', 'reason', and 'purpose'.
    The meaning of 'time' is illustrated in example (9), where the prepositional phrase in italics functions as theme:

(9)   I løpet av disse samtalesekvensene som totalt dekker 5 minutter, stiller læreren 29 spørsmål. (L1Nor)
       "During these speech sequences which cover in total 5 minutes, the teacher asks 29 questions."

Marked themes expressing 'place' may have a non-literal meaning, or be geographical, as is illustrated by example (10):

(10)   *In New Zealand* narrative formative assessments are used widely to support children's learning in ECE settings (Carr, 2009) […] (L1Eng)

'Condition' is frequently expressed by means of dependent clauses starting with *if*, as illustrated by example (11):

(11)   *If one believes that assessments are used to hold teachers and schools accountable*, then teachers can use this perspective to frame or limit their focus when they need to make decisions about assessments in their classrooms. (L1Eng)

The 'instrument' meaning entails a theme that describes the procedures used to achieve the objectives of the investigation. Example (12) is a typical example, where the marked theme describes the method employed in the study:

(12)   *Using exploratory principal axis factoring and cluster analysis*, we examined teachers' conceptions of assessment based on their responses to the COA instrument. (L1Eng)

Marked themes expressing 'concession' are exemplified by the dependent clause in (13):

(13) *Selv om de svarer anonymt*, kan svarene deres reflektere hva de tror er forventete eller ønskelige holdninger i en gitt situasjon, mer enn det de virkelig føler og mener om sakene som skjemaet dreier seg om. (L1Nor)
"Even though they respond anonymously, their answers may reflect what they think of as expected or desired attitudes in a given situation, more than what they really feel and think about the issues that the form asks about."

The two final meanings included here are 'reason' and 'purpose'. These may be somewhat similar in meaning, in the sense that 'purpose' meanings can be paraphrased as, for instance, "because X is so, we need to do Y to avoid it", but in practice the coding proved relatively unproblematic, as the occurrences were nearly all prototypical. The following two examples are representative of the vast majority of cases:

(14) *Because our factors were moderately correlated*, using an oblique rotation should "theoretically render a more accurate, and perhaps a more reproducible solution" (Costello & Osborne, 2005, p. 3). (L1Eng)

Example (14) is a clear case of a marked theme expressing 'reason', as signaled by *because*.

(15) *For å oppnå tilfredsstillende reliabilitet for denne* faktoren måtte én av påstandene i settet tas ut. (L1Nor)
"To achieve satisfactory reliability for this factor, one of the statements in the set had to be removed."

Equally, example (15) is an uncontroversial instance of 'purpose', where the dependent clause would require substantial changes to give it a 'reason' meaning instead.

3.5   The cross-linguistic applicability of terms and definitions

English and Norwegian are very similar languages, and the functions, realizations and meanings expressed by marked themes discussed in Sections 3.2 through 3.4 are among the inventory of possibilities in both languages, with the exception of non-finite *–ing* clauses, which do not occur in Norwegian. The same thing technically applies to the definition of marked themes as non-subjects in initial position, but there is a potential complication that might increase the proportion of marked themes in Norwegian: Norwegian is a V2-language (Faarlund *et al*., 1997: 859), so when a sentence starts with a non-subject element that does not exhaust the thematic potential (e.g. an adverbial conjunct), the finite verb is placed before the subject. This could potentially result in a higher frequency of marked themes, since initial processes would be counted as such. However, as mentioned in Section 3.2, this did not occur in our material, so it will not be discussed as a potential factor influencing the quantitative results presented in Section 5.1.

4.   Material and method

The study was designed to enable a contrastive analysis of L1 material in Norwegian and English, with an extra dimension being provided by L2 material in English written by native speakers of Norwegian, which afforded the opportunity to investigate whether L1 Norwegian writers transfer Norwegian patterns of marked-theme usage into their construction of English texts. Thus, the material comprises three subcorpora: 'L1Eng' and 'L1Nor', which contain published academic articles written by native speakers of English and Norwegian, respectively, and 'L2Eng', which contains published academic articles written in English by native speakers of Norwegian.[3] There are 11 complete texts in each subcorpus, collected from the following journals: L1Eng: *Teaching and Teacher Education*; L1Nor: *NOA Norsk som andrespråk*, *Studies in Education*, *Norsk Pedagogisk Tidsskrift*, *Nordisk Pedagogik*, *Heimen*, and *Nordic Studies in Education*; L2Eng: *Acta Didactica Norge* and *Nordic Journal of Modern Language Methodology*.

Table 1 provides an overview of the material as regards the number of words and number of T-units in each subcorpus.

Table 1. Overview of the material.

Subcorpora	Number of words	Number of T-units
L1Eng	90,052	3,414
L1Nor	56,161	2,732
L2Eng	75,529	3,130
Total	221,742	9,276

In total, the material comprises some 221,000 words and nearly 9,300 T-units, but as is obvious from Table 1 the English corpora contain longer texts than the Norwegian corpus, both in terms of number of words and number of T-units.

The texts were segmented and coded manually, since there is no automatic procedure for the identification of themes. The categories employed were introduced in Section 3 above.

After the manual segmentation and coding, statistical calculations were carried out. The material is relatively small in size, but the findings from the present investigation could nonetheless be used as the basis for hypothesis-forming with future studies in mind. It is therefore important to ensure that any differences identified are not due to chance. The statistical calculations have been carried out using RStudio (RStudio Team, 2015), in accordance with the following general procedure:

1.   The frequency for the occurrence of each variable per T-unit or marked theme was calculated for each text (e.g. the number of marked themes divided by the number of T-units).[4]

---

[3] It should be noted that the authors have been assigned native-speaker status primarily on the basis of their names, following the practice outlined by Shaw (2004: 72). However, with regard to the writers with a Norwegian background it is possible to say for certain that they are native speakers of Norwegian, since the relatively-speaking smaller context makes it possible to know more about the people behind the names.
[4] In section 5, frequencies per 100 T-units or 100 marked themes serve as the starting point for the discussion. This represents an upscaled version of the numbers employed in the statistical calculations, but has the obvious advantage of reducing the number of decimal points and therefore represents an improvement from the point of view of readability.

2. Tests were carried out to check whether the frequencies for each variable had normal distribution in each of the subcorpora. These tests included both the Shapiro-Wilks test and the function *gvlma* from the R package 'gvlma' (Peña and Slate, 2014), since the Shapiro-Wilks test is less than ideal for small samples (cf. Jensen, 2017: 84). In addition to these tests the density plots for each variable were inspected visually.

3a. In cases where all indicators listed under step 2 suggested that the distribution was normal and thus satisfied the requirements for a parametric test, step 3 was to carry out a one-way ANOVA with a Tukey post-hoc test.

3b. In cases where the distribution was not normal, a logit transformation was carried out following the procedure and R code outlined by Jensen (*ibid.*). After the transformation the data were once again tested using the *gvlma* function in R, and, upon confirmation that the distribution was normal following the logit transformation, a one-way ANOVA with a Tukey post-hoc test was carried out.

4. In those cases where the ANOVA and Tukey post-hoc test indicated a significant difference ($p<0.05$) between two subcorpora, the effect size and 95% confidence interval were calculated using the function *cohen.d* from the R package 'effsize' (Torchiano, 2017). Following Sawilowsky (2009: 599), these rules of thumb for interpreting effect sizes (the d-value) were applied: 'very small' for d=0.01, 'small' for d=0.2, 'medium' for d=0.5, 'large' for d=0.8, 'very large' for d=1.2, and 'huge' for d=2.0.

5. Results

The presentation of the results will be structured as follows: first, the proportions of marked themes in the three corpora will be discussed. Next come the functions and realizations of marked themes, followed by the meanings expressed by the marked themes. We will present the results in table form, with a visual representation of the corpus-internal distribution added for those variables where one or more significant differences were identified.

5.1 The proportion of marked themes

Table 2 shows the frequencies of marked themes per 100 T-units in each text of the three corpora. The final row shows the average frequency for each subcorpus.

Table 2. The frequency of marked themes per 100 T-units in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	20.8	19.6	19.3
Text 2	25.4	30.6	25.7
Text 3	17.6	19.4	18.3
Text 4	19.8	31.4	9.8
Text 5	28.5	18.1	15.8
Text 6	34.8	27.3	16.6
Text 7	28.0	17.3	15.4
Text 8	35.0	28.8	21.9
Text 9	31.0	26.9	17.1
Text 10	25.8	17.3	19.9
Text 11	29.8	22.3	21.9
Average frequency	27.0	23.6	18.3

The contents of Table 1 seem to suggest that there is a slight cross-linguistic difference, with marked themes being more frequent in L1 English than in L1 Norwegian. The frequency of marked themes is even lower in L2 English than in Norwegian, suggesting that if transfer plays a role here, it cannot be the only factor affecting the L2 writers. However, the statistical calculations show that the only significant difference is between the L1 English texts and the L2 English texts (p=0.0015121), with a very large effect size (Cohen's d ≈ 1.7, CI(95%) ≈ [0.6, 2.8]. The difference between L1Eng and L2Eng is easier to grasp when presented visually, as in Figure 1.
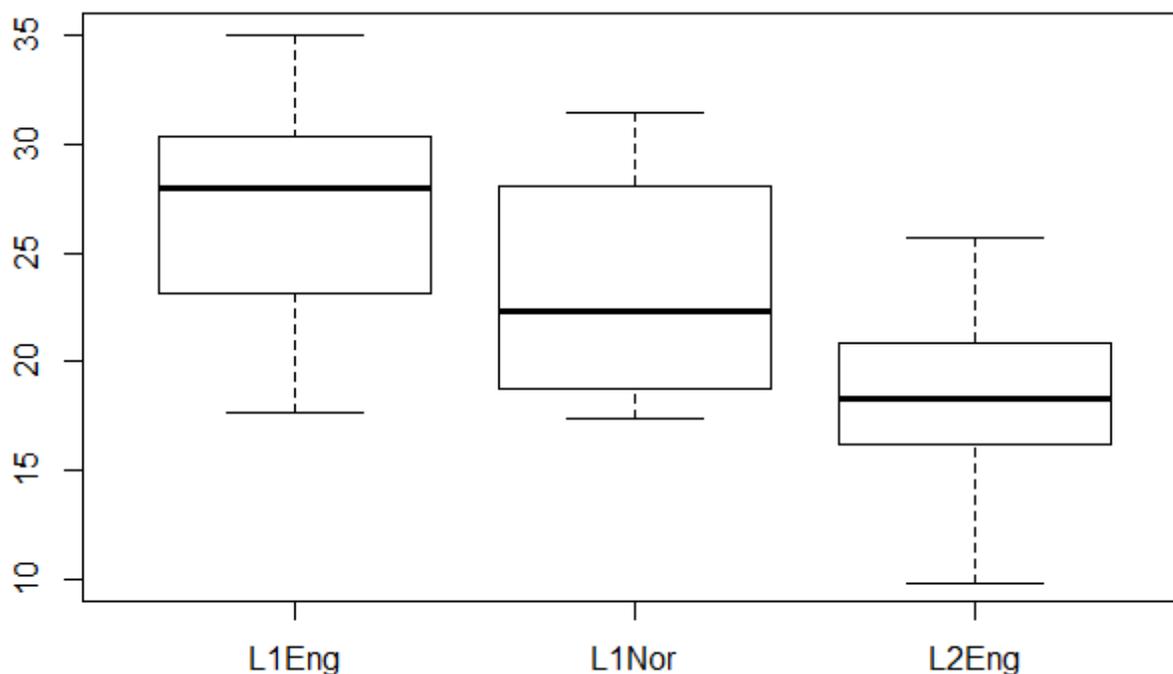

Figure 1. Distribution of marked themes per 100 T-units in each subcorpus.

The boxplots in Figure 1 clearly show that the writers in the L1Eng corpus use more marked themes overall than the L2 writers of English do. There is a certain amount of overlap between the plots, due to the corpus-internal variation in each group of writers, but it is nonetheless clear that the majority of L1 writers use more marked themes than the majority of L2 writers. Given the lack of a significant contrastive difference between the L1 texts in English and Norwegian as regards the frequency of marked themes, we must conclude that this is one area where the L2 writers of English have not successfully adapted to English discourse conventions, but probably not due to (direct) transfer from Norwegian.

Many of the previous studies discussed in Section 3 report their findings in percentages, e.g. the percentage of marked themes out of all the themes in a corpus. It is therefore difficult to compare the results from previous studies with the normalized frequencies presented in the present investigation. We can relate the findings here to those from previous contrastive studies, however. On the one hand, we might have expected a cross-linguistic difference between the two sets of L1 writers as regards the frequency of marked themes, as Hasselgård (1998, 2004, 2005) found that Norwegian fiction accommodates more non-subjects in initial position than does English. As we have seen, this does not hold true for the present material, which appears more similar to argumentative texts in that there is no such contrastive difference (Rørvik, 2013). Moreover, Shaw (2004) found a difference between English and

Danish economics articles, in that the Danish texts had more initial circumstance adverbials than the English texts, and he found that Danish L2 writers of English behaved similarly to the L1 writers of Danish in this regard. The writers in the present L2Eng corpus underuse marked themes where Shaw's L2 writers overused them, though it should be remembered that marked themes may encompass more than just initial adverbials, so the results are not 100% comparable.

5.2 The functions of marked themes

We now turn to the functions of marked themes. We have calculated the frequencies of themes functioning as Adjuncts per 100 marked themes, and these are presented in Table 2.

Table 3. The frequency of marked themes functioning as Adjunct per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	98.3	95.4	98.4
Text 2	95.0	98.5	97.7
Text 3	100.0	93.4	98.2
Text 4	100.0	95.9	100.0
Text 5	96.3	86.6	100.0
Text 6	98.0	96.4	94.1
Text 7	99.0	93.3	93.9
Text 8	98.9	96.0	100.0
Text 9	96.7	94.8	97.7
Text 10	95.7	97.4	97.1
Text 11	100.0	94.2	98.1
Average frequency	98.0	94.7	97.7

It is suggested by the average frequencies in Table 3 that there is a cross-linguistic difference in the use of marked themes that function as Adjuncts. The L1 Norwegian texts have a lower frequency than the two English subcorpora, although Adjuncts are slightly less frequent in the L2 English corpus than in the L1 English corpus. The statistical calculations confirm the impression created by the average frequencies, in that there is a significant difference between L1Eng and L1Nor (p=0.0058306) and between L2Eng and L1Nor (p=0.0265318). In both cases, the effect size is very large (d ≈ 1.44, CI(95%) ≈ [0.38, 2.5] and d ≈ -1.44, CI(95%) ≈ [-2.5, -0.38], respectively). Given these significant differences, we will include Figure 2 showing the distribution of marked themes functioning as Adjuncts.

Figure 2 below clearly shows that the majority of Norwegian texts have a frequency of marked themes functioning as Adjuncts that is below the frequencies in the majority of both the L1 and the L2 English texts. The shape of the plots themselves indicate that there is less variation among the L2 writers than among the two sets of L1 writers (but note the two outliers in the L2Eng corpus). With regard to the use of marked themes functioning as Adjuncts, then, we may conclude that the L2 writers have been able to conform to the pattern of usage employed by L1 writers of English, despite the potential for a transfer-related effect.
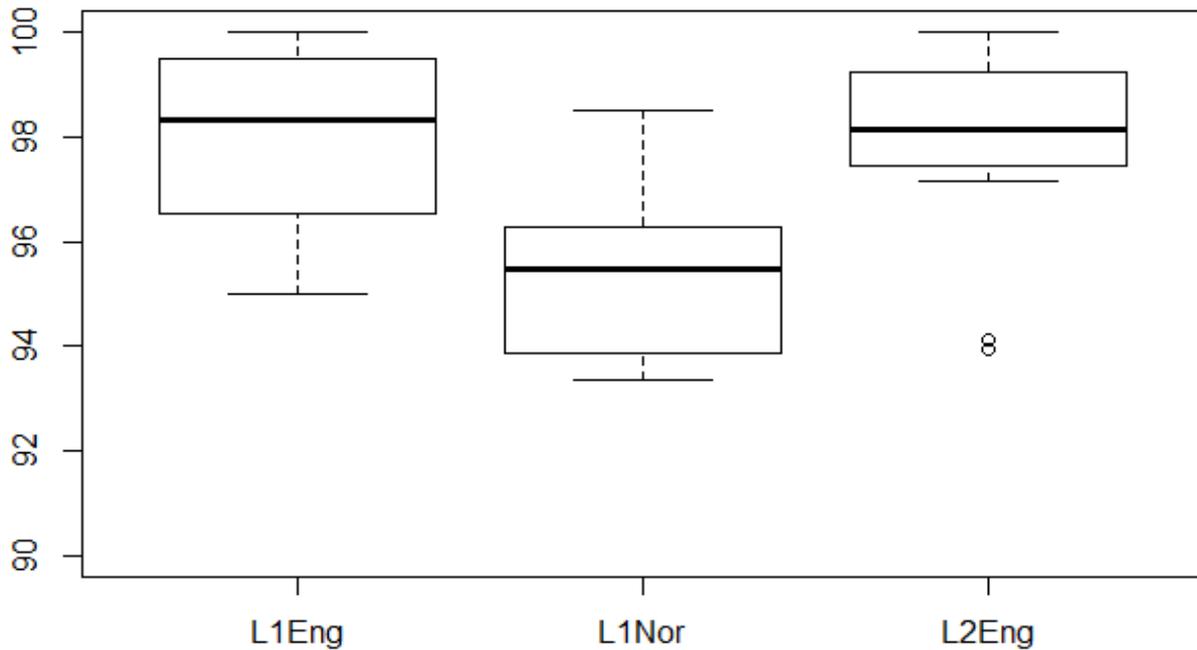
Figure 2. Distribution of marked themes functioning as Adjunct per 100 marked themes in each subcorpus.

The results for marked themes functioning as Adjuncts are more comparable to those presented in Shaw (2004), since initial circumstance adverbials will always be classified as marked themes using the present analytical framework. But here it is clear that Norwegian L2 writers of English have adapted better to English discourse conventions than the Danish L2 writers investigated by Shaw, and indeed the present L2 writers have managed to comply with English discourse conventions despite the cross-linguistic difference resulting from the lower frequency of marked themes functioning as Adjuncts in L1Nor. Hasselgård (2014) found more initial adjuncts in Norwegian than in English, in both news and fiction, which indicates that the text-type or genre also affects the use of marked themes functioning as Adjunct.

Since there were no marked themes realized by processes in the current material, the only other possible function of marked themes is that of Complement, for which the frequencies are presented in Table 4.

Table 4: The frequency of marked themes functioning as Complement per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	1.6	4.5	1.5
Text 2	5.0	1.4	2.2
Text 3	0.0	6.5	1.7
Text 4	0.0	4.0	0.0
Text 5	3.6	13.3	0.0
Text 6	1.9	3.5	5.8
Text 7	0.9	6.6	6.0
Text 8	1.0	3.9	0.0
Text 9	2.1	5.1	2.2
Text 10	4.2	2.5	2.8
Text 11	0.0	5.7	1.8
Average frequency	1.8	5.2	2.2

The average frequencies presented in Table 4 are difficult to interpret, since the two English subcorpora have several texts with no marked themes functioning as complements in them. This naturally means that the average frequencies give a distorted picture of the distribution in the corpora. The statistics show, however, that the average frequencies may be trusted in this case in that they indicate that marked themes functioning as Complements are more frequent in Norwegian than in English: we once again find a significant contrastive difference with very large effect size between L1Nor and L1Eng ($p=0.0038413$, $d \approx -1.52$, CI(95%) $\approx$ [-2.6, -0.45]) and between L1Nor and L2Eng ($p=0.0242967$, $d \approx 1.3$, CI(95%) $\approx$ [0.26, 2.33]). The low frequencies naturally result in plots that occur towards the bottom of Figure 3:
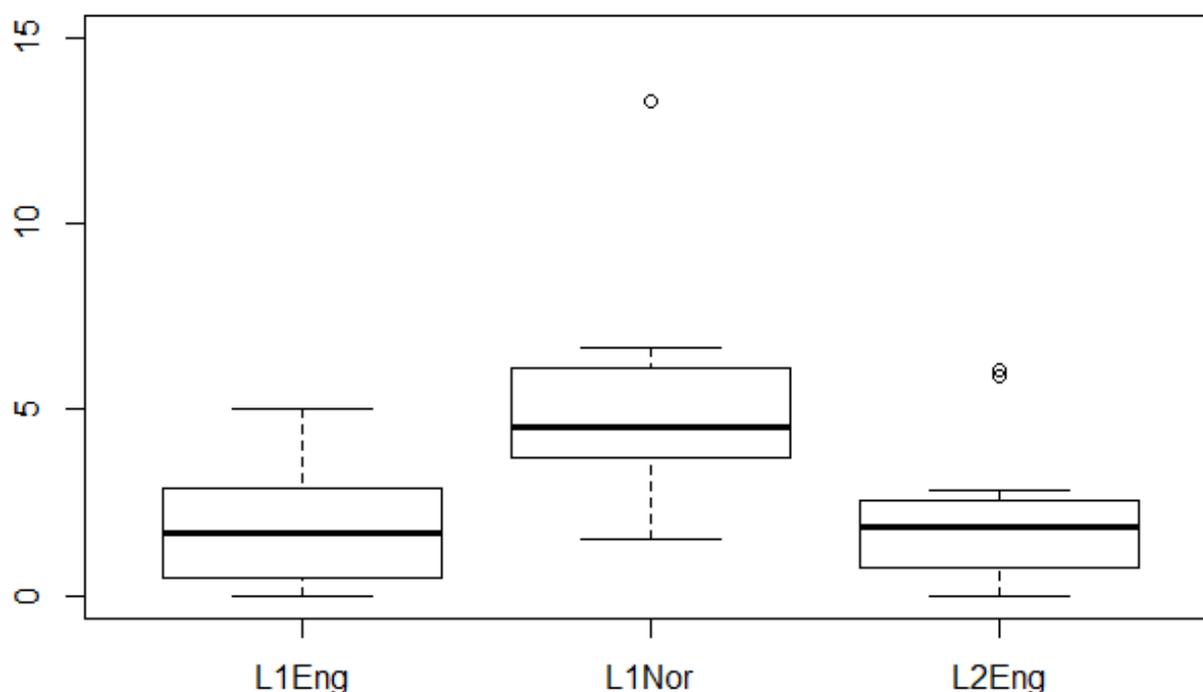

Figure 3. Distribution of marked themes functioning as Complement per 100 marked themes in each subcorpus

There is more corpus-internal variation in the two L1 subcorpora than in the L2 corpus, but L1Nor has one outlier with a lot more marked themes functioning as Complement than the other Norwegian texts. L2Eng has two outliers with slightly higher frequencies, but these are not enough to pull the average frequency up to a figure more closely resembling that found in L1Nor. Thus, the cross-linguistic difference for marked themes functioning as Complements goes in the opposite direction to that found for Adjuncts: Complements are more frequent in Norwegian than in English, which is in line with the results presented by Hasselgård (1998, 2004). It also seems clear that, as regards both functions of marked themes found in the present material, the L2 writers have not been influenced by transfer from Norwegian, and have managed to avoid a foreign flavor in their English texts.

5.3 The realizations of marked themes

In this section the focus is on the realization of marked themes, starting with marked themes realized by dependent clauses, for which the frequencies per 100 marked themes are presented in Table 5.

Table 5. The frequency of marked themes realized by dependent clauses per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	27. L1Nor 3	L2Eng
Text 1	31.7	16.4	45.5
Text 2	46.3	23.7	30.0
Text 3	40.4	12.1	33.3
Text 4	28.2	46.7	50.0
Text 5	45.0	30.6	57.4
Text 6	42.6	26.7	49.0
Text 7	45.5	39.2	21.2
Text 8	59.4	29.3	35.6
Text 9	48.4	15.4	31.1
Text 10	45.1	21.4	17.1
Text 11	40.0	26.3	51.9
Average frequency	43.0	27.3	38.4

Based on the average frequencies presented in Table 5 it seems that the L1 English texts contain more marked themes realized by dependent clauses than do the L2 English texts, and that transfer from Norwegian may play a role in lowering the frequency in L2Eng, since the Norwegian texts contain by far the lowest frequency. However, the difference between the two English corpora is not significant (p=0.5849766), so the only real difference is between the Norwegian subcorpus and the two corpora containing texts written in English. The difference between the Norwegian subcorpus and L2Eng is large (p=0.0341214, d ≈ -1.02, CI(95%) ≈ [-2.02, -0.02]), and the difference between L1Nor and the L1 English texts is very large (p=0.0028936, d ≈ 1.78, CI(95%) ≈ [0.66, 2.9]). It is still possible that transfer may affect the L2 English writers, just not to the extent that it results in a significant difference between their texts and the L1 English texts. The plots in Figure 4 show a very different distribution among the L2 writers than among the writers in L1Eng.

There is much more variation in L2Eng than in L1Eng, and quite a lot of overlap between the plots for the Norwegian subcorpus and the non-native English subcorpus. It is therefore possible that a larger sample would have resulted in greater similarity between the L2 English texts and the Norwegian texts.
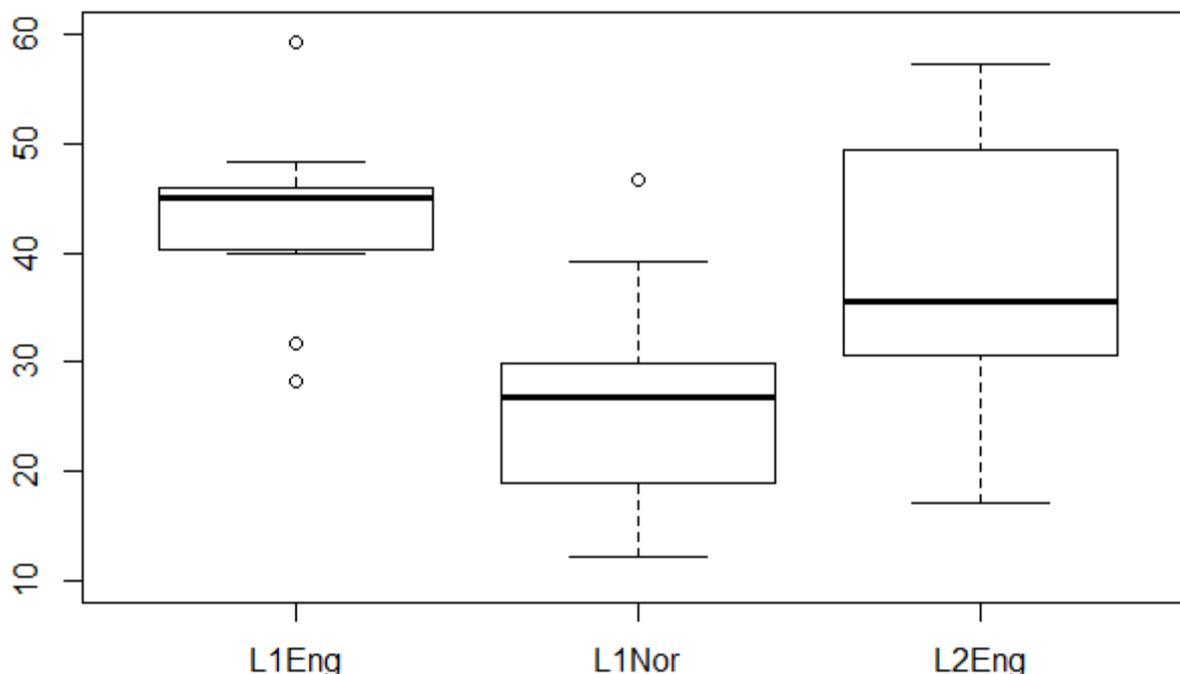
Figure 4. Distribution of marked themes realized by dependent clauses per 100 marked themes in each subcorpus

One potential explanation for the lower frequency of marked themes realized by dependent clauses in L1Nor is that the writers of English may have used a great deal of non-finite *-ing* clauses, which do not have an direct counterpart in Norwegian. Future studies of the kind reported on here would do well to include this variable, but for present purposes we refer to Hasselgård (2014), who did not find more initial adjuncts realized by non-finite clauses in English than Norwegian.

The second structure to be included here is prepositional phrases functioning as marked themes. Table 6 contains the normalized frequencies for these.

Table 6. The frequency of marked themes realized by prepositional phrases per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	55.0	50.0	50.0
Text 2	38.7	67.1	61.1
Text 3	53.1	53.9	64.9
Text 4	59.1	68.6	50.0
Text 5	48.6	36.6	39.3
Text 6	52.4	55.2	41.1
Text 7	52.4	66.6	72.7
Text 8	36.4	43.1	57.6
Text 9	39.5	55.1	62.2
Text 10	47.8	71.7	65.7
Text 11	58.8	61.4	44.4
Average frequency	49.3	57.2	55.3

The differences between the average frequencies in Table 6 are relatively small, both between the two English subcorpora and cross-linguistically, and the differences that do occur are in

fact non-significant for all comparisons (L1Eng vs. L2Eng: p=0.3522646; L1Eng vs. L1Nor: p=0.1749415; L2Eng vs. L1Nor: p=0.9015830). Hasselgård (2014: 85) reports twice as many initial prepositional phrases in Norwegian as in English, but since the present material does not exhibit the same contrastive difference, we conclude that text-type differences may influence the use of prepositional phrases in the sense that a the cross-linguistic difference found for news and fiction does not appear to exist in the case of academic articles in the field of didactics.

The final results reported on in this subsection are those concerning the marked themes realized by adverb phrases. The normalized frequencies are given in Table 7.

Table 7. The frequency of marked themes realized by adverb phrases per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	11.7	18.2	3.0
Text 2	7.5	10.4	4.4
Text 3	4.3	10.5	0.0
Text 4	8.5	15.2	0.0
Text 5	0.9	6.7	1.6
Text 6	5.0	10.6	2.0
Text 7	2.0	3.3	0.0
Text 8	3.1	13.7	5.1
Text 9	9.9	10.3	2.2
Text 10	2.8	5.1	11.4
Text 11	1.1	11.4	0.0
Average frequency	5.1	10.5	2.7

On the basis of the average frequencies in Table 7 it would seem that the two English corpora are fairly similar as regards marked themes realized by adverb phrases, although there is a slightly higher frequency in the L1 English texts than in the L2 texts. What is more striking, however, is the fact that the Norwegian texts contain twice as many adverb phrases functioning as marked themes than do the L1 English texts, and nearly four times as many as can be found in the L2 English material. It should come as no surprise, therefore, that the only significant differences here are the cross-linguistic ones. The difference between L1Nor and L1Eng is very large (p=0.0243690, d ≈ -1.3, CI(95%) ≈ [-2.3, -0.2]), and the difference between L1Nor and L2Eng is huge (p=0.0243690, d ≈ 2.1, CI(95%) ≈ [0.92, 3.2]). As there are significant differences between the subcorpora with regard to this feature, we include the boxplots in Figure 5.
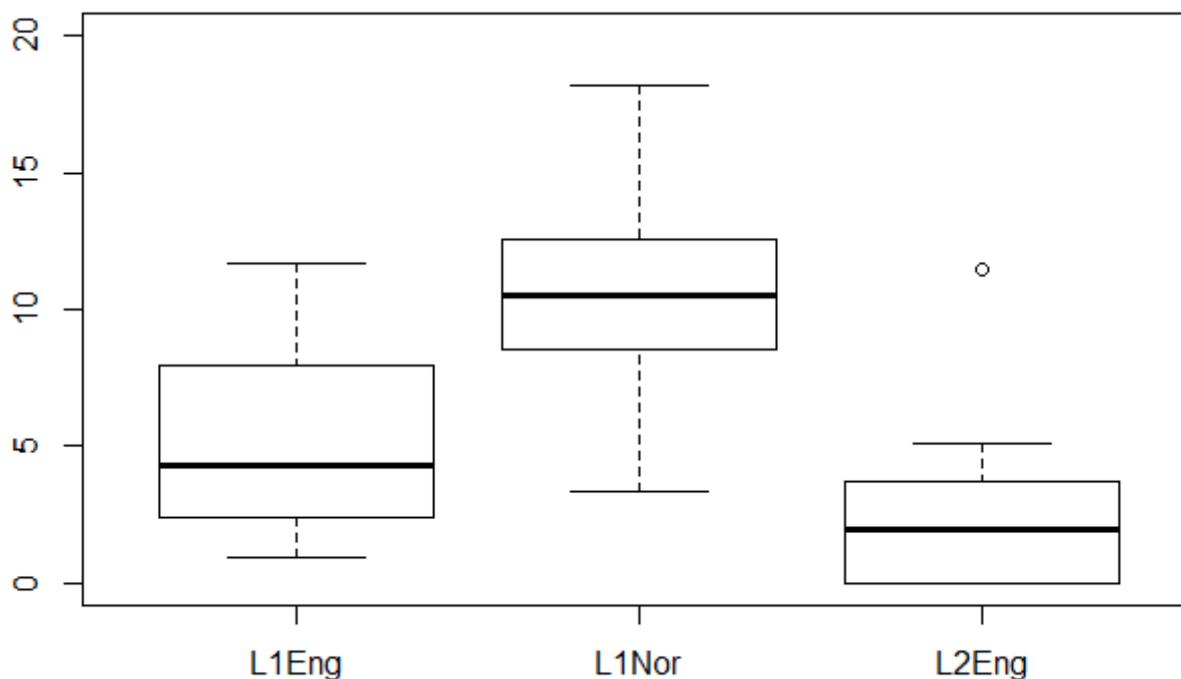
Figure 5. Distribution of marked themes realized by adverb phrases per 100 marked themes in each subcorpus.

The boxplots in Figure 5 show that there is more variation within the L1Nor subcorpus than within the two other corpora, and this might imply a need to be cautious regarding the firmness of conclusions drawn about the existence of cross-linguistic differences in the use of adverb phrases: there is quite a lot of overlap between the lower range of the Norwegian texts and the L1 English texts, for instance. However, it is also clear that the upper range of the Norwegian frequencies is well above all of the L1Eng values, so on the basis of the present material, at least, there is a clear contrastive difference. Despite this, the L2 writers have managed to follow English conventions in their adverb-phrase usage of the kind discussed here. The contrastive difference matches that found in Hasselgård (2014), but it should be mentioned that the frequencies there are very low overall.

5.4 The meanings expressed by marked themes

As described in Section 3.4, we include findings regarding seven meanings expressed by marked themes in our discussion here. The meanings where significant differences were identified between the subcorpora will be discussed first. These include 'place', 'instrument', 'concession', and 'purpose'. Next, we will present a brief overview of those meanings where no significant differences were found between the subcorpora, i.e. 'time', 'condition', and 'reason'. We end this subsection with a comparison of the present findings and the results from previous studies regarding the meaning expressed by marked themes.

5.4.1 *Meanings where there is a significant difference*
The first meaning where there is a significant difference between two or more of the subcorpora is 'place'. Table 8 presents the normalized frequencies for this meaning.

Table 8. The frequency of marked themes expressing 'place' per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	38.3	38.6	15.2
Text 2	6.3	35.8	33.3
Text 3	14.9	34.2	21.1
Text 4	5.6	32.3	7.7
Text 5	14.4	10.0	9.8
Text 6	15.8	29.4	15.7
Text 7	25.7	43.3	39.4
Text 8	9.4	25.5	30.5
Text 9	8.8	22.4	15.6
Text 10	7.0	33.3	25.7
Text 11	35.6	38.6	22.2
Average frequency	16.5	31.2	21.5

The only significant difference among the three subcorpora as regards the frequency of adverbials expressing 'place' is between the two L1 corpora, where the effect size is very large ($p=0.0050644$, $d \approx -1.4$, $CI(95\%) \approx [-2.4, -0.3]$). Figure 6 illustrates the difference, and additionally shows a much greater degree of corpus-internal variation in L1Eng than in L1Nor, with the L2 writers of English in between the other two groups.
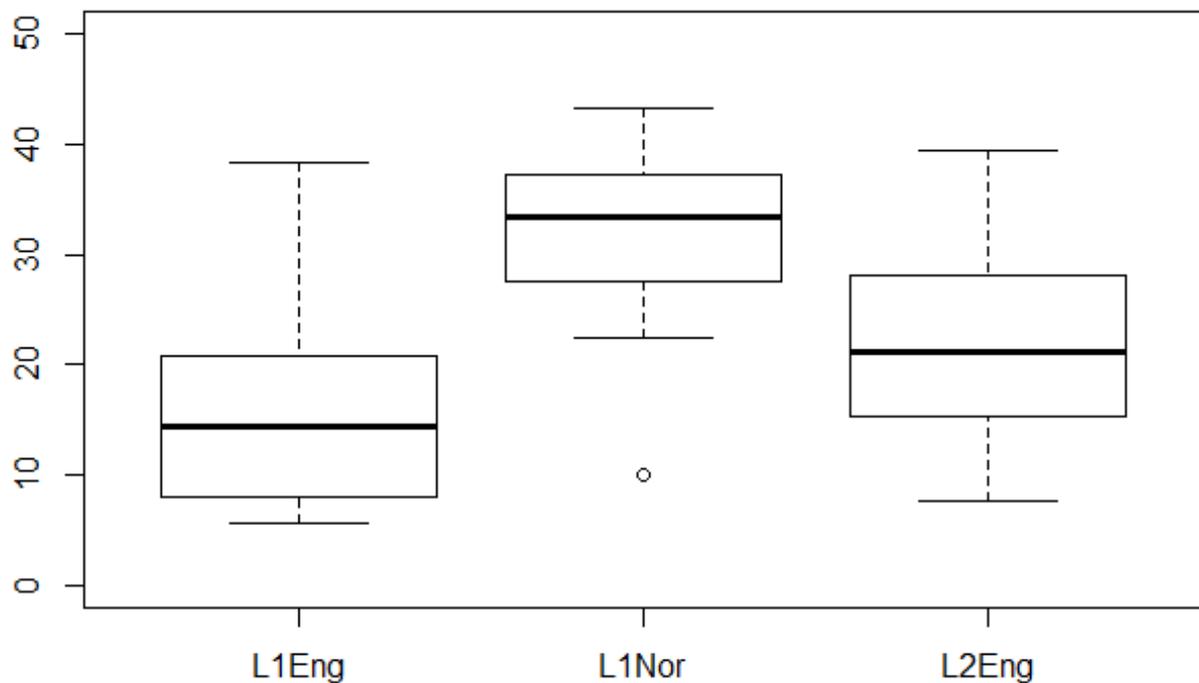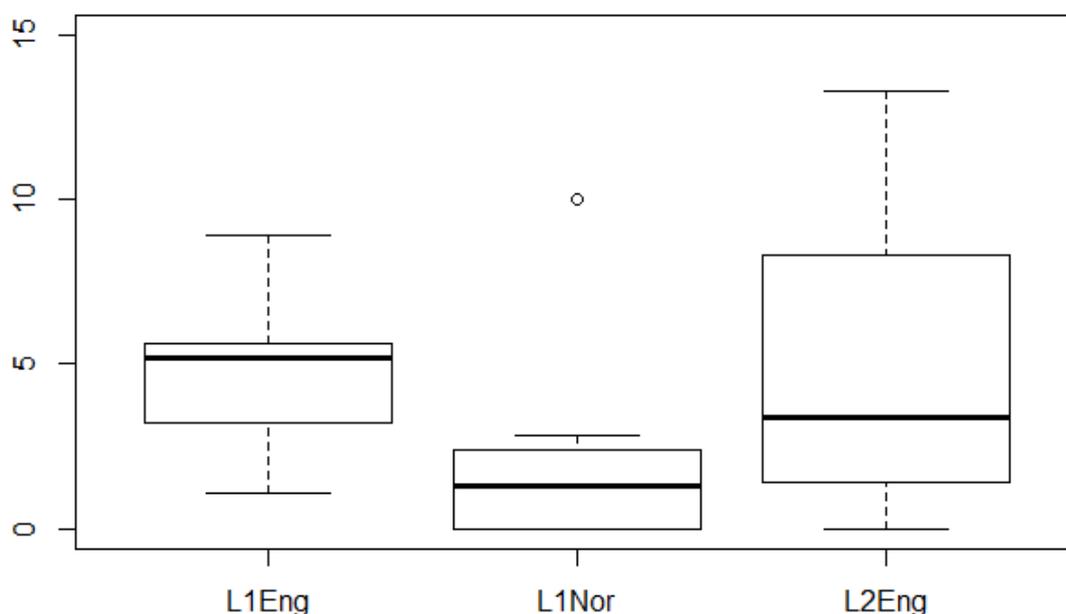

Figure 6. Distribution of marked themes expressing 'place' per 100 marked themes in each subcorpus.

Next we will examine the frequencies for marked themes expressing 'instrument', which are given in Table 9.

Table 9. The frequency of marked themes expressing 'instrument' per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	1.7	2.3	3.0
Text 2	3.8	1.5	11.1
Text 3	8.5	1.3	10.5
Text 4	5.6	0.0	0.0
Text 5	2.7	10.0	0.0
Text 6	8.9	1.2	3.9
Text 7	5.0	0.0	6.1
Text 8	5.2	0.0	3.4
Text 9	1.1	0.0	13.3
Text 10	5.6	2.6	2.9
Text 11	5.6	2.9	0.0
Average frequency	4.9	2.0	4.9

Although the two English subcorpora have exactly the same average frequency, it is only the difference between L1Eng and L1Nor that is significant, with a very large effect size ($p=0.0286952$, $d \approx 1.3$, $CI(95\%) \approx [0.28, 2.3]$). The reason for this is probably easier to understand if we look at the distribution within each subcorpus, since, as illustrated in Figure 7, the texts in L1Eng are much more consistently placed around the average of 4.9, whereas L2Eng has several texts with no marked themes expressing 'instrument' and several texts that score higher than in any of the two other subcorpora. The Norwegian texts, on the contrary, include one outlier with a relatively high frequency of 'instrument' meanings, and also four texts with no such marked themes, and in general extremely low frequencies.

Figure 7. Distribution of marked themes expressing 'instrument' per 100 marked themes in each subcorpus.

Marked themes expressing 'concession' are slightly more frequent than those expressing 'instrument', as shown by the figures in Table 10.

Table 10. The frequency of marked themes expressing 'concession' per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	6.7	9.1	12.1
Text 2	12.5	7.5	4.4
Text 3	12.8	1.3	3.5
Text 4	7.0	2.0	30.8
Text 5	18.0	3.3	18.0
Text 6	9.9	7.1	5.9
Text 7	11.9	0.0	9.1
Text 8	27.1	15.7	15.3
Text 9	16.5	5.2	13.3
Text 10	15.5	10.3	11.4
Text 11	17.8	5.7	24.1
Average frequency	14.1	6.1	13.4

The two English subcorpora are practically identical in terms of how many times the writers use marked themes expressing 'concession'. This does not necessarily mean that the corpus-internal distribution is identical, as was discussed above with regard to 'instrument', and indeed Figure 8 shows that there are distribution differences for 'concession' as well.
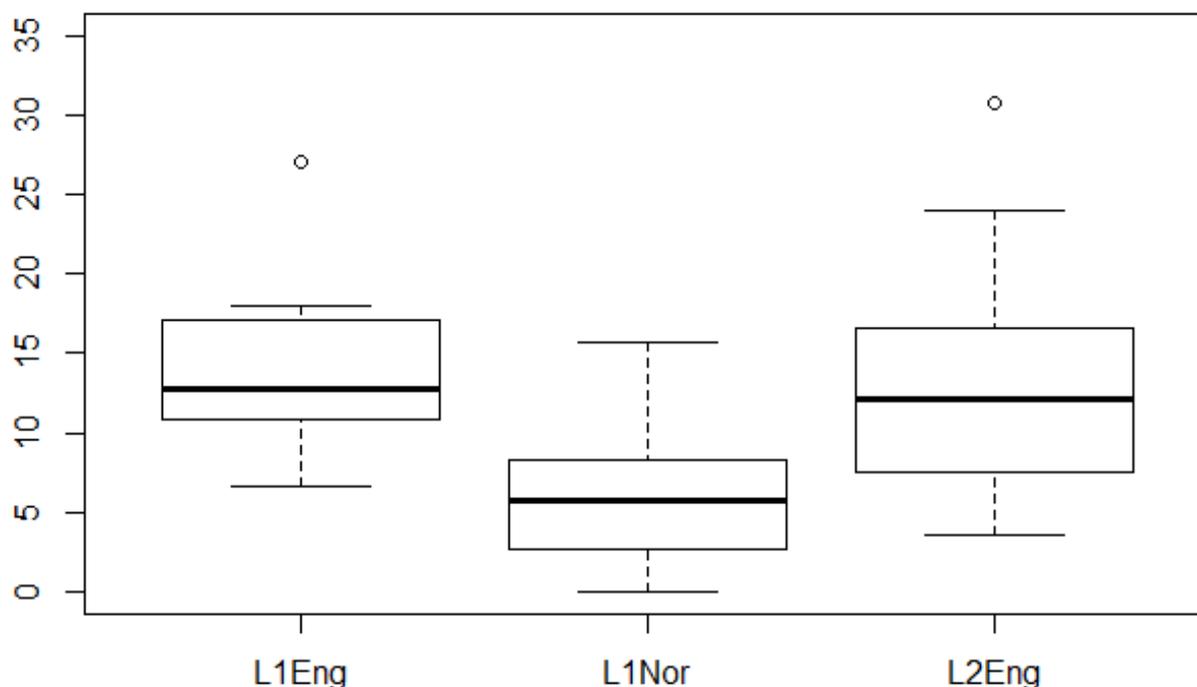

Figure 8. Distribution of marked themes expressing 'concession' per 100 marked themes in each subcorpus.

There is more variation within the L2 English corpus than within the L1 English corpus, but a sufficient number of the texts are centered around the mean for there not to be a significant difference between these two subcorpora. However, each of them in turn are significantly different from the L1 Norwegian corpus, despite some overlap in the boxplots (L1Eng vs. L1Nor: $p=0.0170723$, $d \approx 1.5$, CI(95%) $\approx$ [0.46, 2.6]; L2Eng vs. L1Nor: $p=0.0310533$, $d \approx -1$, CI(95%) $\approx$ [-2.09, -0.08]).

Finally, we will look at the frequencies of marked themes expressing 'purpose', which can be found in Table 11.

Table 11. The frequency of marked themes expressing 'purpose' per 100 marked themes in each text and the average frequency per corpus.

	L1Eng	L1Nor	L2Eng
Text 1	5.0	0.0	0.0
Text 2	15.0	1.5	1.1
Text 3	4.3	3.9	17.5
Text 4	1.4	0.0	11.5
Text 5	2.7	0.0	3.3
Text 6	1.0	3.5	3.9
Text 7	4.0	0.0	0.0
Text 8	7.3	0.0	8.5
Text 9	5.5	1.7	6.7
Text 10	5.6	5.1	0.0
Text 11	7.8	1.4	1.9
Average frequency	5.4	1.6	4.9

The difference between L1Eng and L2Eng is not significant (p=0.6515050), but the difference between L1Eng and the Norwegian texts is (p=0.0292506, d ≈ 1.3, CI(95%) ≈ [0.31, 2.4]), so once again we have identified a feature where the L2 writers have managed to comply with English discourse conventions despite a contrastive difference. The difference between the L2 texts and the Norwegian texts is not significant, so it is possible that there is a slight transfer effect, but if so the L2 writers have come sufficiently far in their development to approach native-speaker usage with regard to marked themes expressing 'purpose'. As is illustrated in Figure 9, however, there is a great deal of variation in the L2 subcorpus, and some texts may be approaching overuse of 'purpose'.
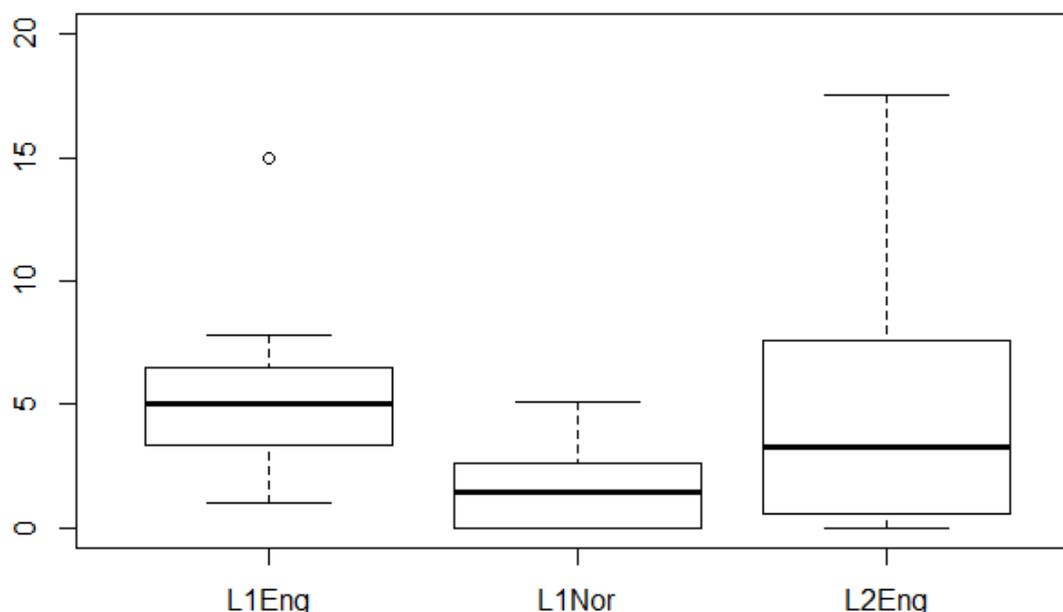

Figure 9. Distribution of marked themes expressing 'purpose' per 100 marked themes in each subcorpus.

Although there is some overlap between the plots for the two native-language corpora, there are generally lower frequencies in the Norwegian subcorpus, and also five texts where there are no marked themes expressing purpose.

5.4.2 *Meanings where there is no significant difference*

No significant differences were found for the three remaining meanings discussed here, namely 'time', 'condition', and 'reason'. We nonetheless include the frequencies for these meanings in the three subcorpora, since it is interesting to compare the present findings with relevant findings from previous studies. The frequencies for all three meanings are given in Table 12.

Table 12. The frequency of marked themes expressing 'time', 'condition', and 'reason' per 100 marked themes in each text and the average frequency per corpus.

	'Time'			'Condition'			'Reason'		
	L1Eng	*L1Nor*	*L2Eng*	*L1Eng*	*L1Nor*	*L2Eng*	*L1Eng*	*L1Nor*	*L2Eng*
Text 1	16.7	9.1	28.8	3.3	4.5	4.5	6.7	11.4	6.1
Text 2	15.0	20.9	20.0	1.3	6.0	1.1	11.3	0.0	6.7
Text 3	23.4	19.7	12.3	0.0	13.2	7.0	10.6	1.3	12.3
Text 4	31.0	39.4	7.7	5.6	2.0	0.0	9.9	0.0	23.1
Text 5	23.4	13.3	16.4	3.6	20.0	6.6	5.4	6.7	6.6
Text 6	30.7	7.1	9.8	2.0	11.8	5.9	8.9	5.9	15.7
Text 7	16.8	23.3	9.1	5.0	0.0	0.0	9.9	10.0	6.1
Text 8	6.3	13.7	22.0	5.2	11.8	3.4	8.3	3.9	5.1
Text 9	30.8	22.4	22.2	3.3	8.6	2.2	7.7	13.8	8.9
Text 10	23.9	17.9	25.7	0.0	0.0	0.0	9.9	7.7	2.9
Text 11	17.8	20.0	7.4	2.2	4.3	3.7	2.2	4.3	14.8
Average frequency	21.4	18.8	16.5	2.9	7.5	3.1	8.2	5.9	9.8

As Table 12 shows, 'time' is the most frequent of the three meanings, followed by 'reason' and 'condition'. If we look more closely at the frequencies for 'time', we can see that there are slight differences between the corpora: the highest frequency is found in L1Eng, the second highest in L1Nor, and the lowest in L2Eng. 'Condition' is more frequent in Norwegian than in English, and marked themes expressing 'reason' are most frequent in the two English subcorpora. However, these slight differences in the normalized frequencies were not sufficiently large to be significant.

5.4.3 *Meanings expressed by marked themes: comparison with previous research*

As discussed in Section 5.4.1, the meanings 'instrument', 'concession', and 'purpose' were more frequent in the L1 English texts than in the Norwegian texts, with 'concession' also being more frequent in L2 English. 'Place' was found to be more frequent in Norwegian than in English. In Section 5.4.2 it was reported that no significant contrastive difference was found for the meanings 'time', 'condition', and 'reason'. As discussed above it is difficult to relate these findings to the percentages reported in related previous studies, but we can add to the discussion the frequency ranking of the meanings discussed here. Thus, the most frequent meanings in L1Eng are 'time', 'place', and 'concession', in L2Eng they are 'place', 'time', and 'concession', and in L1Nor the most frequent is 'place' followed by 'time' and 'condition'. Table 13 summarizes the findings from previous studies by ranking the three most frequent meanings identified in each study.[5]

---

[5] Note that not all of these studies investigate marked themes specifically, and that some of them employ more inclusive definitions of marked themes than the framework used in the present study. The reader is referred to the individual studies for details.

Table 13. The three most frequent meanings expressed by marked themes in previous studies.

Study	Material	Meaning 1	Meaning 2	Meaning 3
Gosden 1992	Physics, chemistry, and biology	'place'	'contrastive/ concessive'	'time'
Martínez 2003	Method and discussion sections in biology	'purpose'	'time'	'place'
Heng & Ebrahimi 2012	Abstracts in applied linguistics and economics	'contrast'	'time'	'condition'
Hasselgård 2014	English and Norwegian fiction and news	E & N news & fiction: 'time'	E & N news: 'contingency' E & N fiction: 'space'	E & N news: 'space' E & N fiction: 'contingency'
Ebrahimi 2016	Method sections in applied linguistics (AL), psychology (P), and chemistry (C)	AL: 'purpose', 'condition' P: 'purpose' C: 'time'	AL/P: 'location' C: 'purpose'	AL: 'contrast' P: 'time' C: 'condition', 'location'
Present study	L1 Norwegian, L1 English, and L2 English articles in didactics	L1Nor: 'place' L1Eng: 'time' L2Eng: 'place'	L1Nor: 'time' L1Eng: 'place' L2Eng: 'time'	L1Nor: 'condition' L1Eng: 'concession' L2Eng: 'concession'

The three most frequent meanings in each of the present subcorpora all appear among the meanings listed in Table 13, but it seems clear that L1, text type and/or academic discipline influences the internal ranking of meanings expressed by marked themes. Larger datasets are required in order to determine whether articles within didactics vary significantly from other disciplines in this regard, but it does seem to be the case that academic disciplines across the board have in common a fixed set of meanings expressed by marked themes, which may serve as text-structuring strategies by means of which the reader can be guided through the text.

6.   Conclusion

Two research questions formed the starting point for this study. The first asked to what extent there were contrastive differences between L1 writers of Norwegian and English in the frequency, functions and realizations of marked themes, and in the meanings expressed by marked themes. The results showed that there was no difference in the frequency of marked themes, but fewer marked themes functioning as Adjuncts and more functioning as Complements in Norwegian than in English. In terms of the realization of marked themes there were fewer dependent clauses in Norwegian than in English, but more adverb phrases, and as regards the meanings expressed by marked themes it was found that 'place' was more frequent in Norwegian than English, but that 'instrument', 'concession', and 'purpose' were more frequent in English.

The second research question focused on the extent to which L2 writers of English had been able to adapt to English discourse conventions, and whether any evidence of transfer from Norwegian could be found. The only difference between the L1 and L2 writers that could be found was that the L2 writers underused marked themes. This difference could not be related to transfer.

Previous research has shown that text type, academic discipline, and language may influence thematic structure. The present study provides additional evidence that discipline and text type play a role, since the results differ from some previous studies (e.g. Hasselgård,

1998 and Shaw, 2004) in not finding a higher proportion of marked themes in Norwegian (or a Scandinavian language, in the case of Shaw, 2004) than in English. We consider this the main contribution of the present paper, in addition to the empirical evidence collated about thematic structure in didactics articles which can serve as the basis for evidence-based teaching of academic writing.

References

Babaii, E., Atai, M. R. and Shoja, L. 2016. A Comparison of Thematic Choices and Thematic Progression Patterns in the Research Articles of Well-established and Emerging Disciplines. *Iranian Journal of Applied Linguistics* (19)2: 33–60.

Carter-Thomas, S. and Rowley-Jolivet, E. 2008. *If*-conditionals in medical discourse: From Theory to Disciplinary Practice. *Journal of English for Academic Purposes* (7): 191–205.

Davies, F. 1997. Marked Theme as a Heuristic for Analysing Text-Type, Text and Genre. In *Applied Languages: Theory and Practice in ESP*, J. Pique and D. J. Viera (eds), 45–79. Valencia: Universitat de Valencia.

Ebrahimi, S. F. 2016. Across Disciplinary Study of Marked Theme in Method Sections. *The Journal of Teaching English for Specific and Academic Purposes* (4)3: 689–699.

Ebrahimi, S. F., Chan, S. H. and Tan, B. H. 2014. Cross-Disciplinary and Cross-Linguistic Manifestations of Theme in Academic Writing. *Pertanika Journal of Social Sciences & Humanities* (22)3: 695–707.

Faarlund, J. T., Lie, S. and Vannebo, K. I. 1997. *Norsk Referansegrammatikk.* Oslo: Universitetsforlaget.

Gilquin, G. 2000/2001. The Integrated Contrastive Model. Spicing up your data. *Languages in Contrast* 3(1): 95–123.

Gilquin, G. 2008. Combining Contrastive and Interlanguage Analysis to Apprehend Transfer: Detection, Explanation, Evaluation. In *Linking Up Contrastive and Learner Corpus Research*, G. Gilquin, S. Papp and M. B. Díez-Bedmar (eds), 3–33. Amsterdam – New York: Rodopi.

Gosden, H. 1992. Discourse Functions of Marked Theme in Scientific Research Articles. *English for Specific Purposes* (11): 207–224.

Granger, S. 1996. From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora. In *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, K. Aijmer, B. Altenberg and M. Johansson (eds), 37–52. Lund: Lund University Press.

Gray, B. 2015. On the Complexity of Academic Writing. Disciplinary Variation and Structural Complexity. In *Corpus-Based Research in Applied Linguistics: Studies in Honor of Doug Biber*, V. Cortes and E. Csomay (eds), 49–77. Amsterdam: John Benjamins.

Halliday, M. A. K. 2004. *An Introduction to Functional Grammar. 3rd edition, revised by C. M. I. M. Matthiessen.* London: Arnold.

Hasselgård, H. 1998. Thematic Structure in Translation between English and Norwegian. In *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*, S. Johansson and S. Oksefjell (eds), 145–167. Amsterdam: Rodopi.

Hasselgård, H. 2004. Thematic Choice in English and Norwegian. *Functions of Language* 11(2): 187–212.

Hasselgård, H. 2005. Theme in Norwegian. In *Semiotics from the North. Nordic approaches to Systemic Functional Linguistics*, K. L. Berge and E. Maagerø (eds), 35–47. Oslo: Novus Press.

Hasselgård, H. 2014. Discourse-Structuring Functions of Initial Adverbials in English and Norwegian News and Fiction. *Languages in Contrast* (14)1: 73–92.

Heng, C. S. and Ebrahimi, S. F. 2012. Marked Themes as Context Frames in Research Article Abstracts. *GEMA Online Journal of Language Studies* (12)4: 1147–1164.

Hunt, K. W. 1965. Grammatical structures written at three grade levels. NCTE Research Report No. 3. Champaign, Illinois: National Council of Teachers of English. Available at

`https://files.eric.ed.gov/fulltext/ED113735.pdf`. [last accessed February 27 2018]

Hyland, K. 2000. *Disciplinary discourses: Social Interaction in Academic Writing.* Harlow, Essex: Pearson.

Jensen, B. U. 2017. *Leksikosyntaktiske trekk og skriveverktøy. En kvantitativ undersøkelse av tekster skrevet for hånd og på tastatur av elever i VG1.* PhD thesis, University of Bergen.

Lado, Robert. 1957 [1971]. *Linguistics across Cultures: Applied Linguistics for Language Teachers.* Ann Arbor MI: University of Michigan Press.

Martínez, I. A. 2003. Aspects of Theme in the Method and Discussion Sections of Biology Journal Articles in English. *Journal of English for Academic Purposes* (2): 103–123.

Moyano, E. I. 2016. Theme in English and Spanish. Different Means of Realization for the Same Textual Function. *English Text Construction* (9)1: 190–220.

North, S. 2005a. Different Values, Different Skills? A Comparison of Essay Writing by Students from Arts and Science Backgrounds. *Studies in Higher Education* 30(5): 517–533.

North, S. 2005b. Disciplinary Variation in the Use of Theme in Undergraduate Essays. *Applied Linguistics* 26(3): 431–452.

Peña, E. A. and Slate, E. H. 2014. gvlma: Global Validation of Linear Models Assumptions (R package version 1.0.0.2). Available at `https://cran.rproject.org/web/packages/gvlma/gvlma.pdf`. [last accessed February 27 2018]

RStudio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Available at `http://www.rstudio.com/` [last accessed February 27 2018]

Rørvik, S. 2013. *Texture in learner language.* PhD thesis, University of Oslo.

Sawilowsky, S. S. 2009. New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods* (8)2: 597–599.

Shaw, P. 2004. Sentence Openings in Academic Economics Articles in English and Danish. *Nordic Journal of English Studies* 3(2): 67–84.

Torchiano, M. 2017. effsize: Efficient Effect Size Computation (R package version 0.7.1). Available at `https://cran.r-project.org/web/packages/effsize/effsize.pdf`. [last accessed February 27 2018]

Valipour, V., Nader, A. and Asl, H. D. 2017. The Generic Structures and Lexico-Grammaticality in English Academic Research Papers. *Southern African Linguistics and Applied Language Studies* (35)2: 169–182.

Corresponding author's address

Sylvi Rørvik
Department of Humanities, Faculty of Education
Inland Norway University of Applied Sciences
PO box 400
2418 Elverum
Norway

sylvi.rorvik@inn.no

Exploring cross-linguistic congruence: The case of two stance frames in English and Norwegian

Signe Oksefjell Ebeling

University of Oslo (Norway)

Abstract: This paper explores the degree of congruence between two closely related languages – English and Norwegian – in a case study of the two stance frames *it BE * that* and *det VÆRE * at*. It is first established that the open slot in the frames is most typically occupied by an adjective, thus steering this investigation towards a more detailed comparison of *it BE ADJ that* and *det VÆRE ADJ at*. The adjective determines the evaluative orientation of the frames, and, following Lemke (1998), the present study operates with seven categories to establish the attitudes and evaluations present in the frames. The study, which draws on material from a bidirectional translation corpus – the English-Norwegian Parallel Corpus+, reveals that the degree of congruence between the frames to some extent seems to depend on the attitude/evaluation expressed. Furthermore, the study reveals that there is 45% non-congruence between the languages. The translational correspondence network of the English and Norwegian patterns is broad, showing that there are a number of linguistic resources in the two languages that can be used to express attitudinal/evaluative stance. It is concluded that, while English and Norwegian are shown to have similar means of expressing attitudinal meanings with the frames, the two languages have their preferred ways of doing so both in terms of individual adjectives and attitudinal/evaluative class.

Keywords: congruence, non-congruence, English-Norwegian, extraposition, stance frame, projecting clause, evaluation, bidirectional translation corpus

1. Introduction

A recent trend in corpus-based contrastive studies is to investigate collocational and colligational frameworks, with the aim to "discover recurrent patterns in the lexical and semantic make-up of such sequences" (Hasselgård 2016: 55). The present paper seeks to do exactly this in an investigation of the cross-linguistically similar patterns *it BE * that / det VÆRE * at* in English and Norwegian.[1] Parts of speech typically occurring in the open slot in both languages include ADJ, N and ADV (see examples 1-3, respectively), giving rise to patterns of different grammatical status, notably extraposition (1) and clefts (2)-(3).

---

[1] Frame, framework and pattern are used interchangeably in this study to refer to *it BE * that / det VÆRE * at*.

(1)   It is *fortunate* that I am the one who took your call." [MoAl1E] [2]

(2)   It was *grief* that filled her, transmuting, but slowly, to rage. [DL2][3]

(3)   It was *then* that Fibich wept. [AB1]

Preliminary observations of the patterns in the English-Norwegian Parallel Corpus+ suggest that ADJ is the main colligate in both languages, and that the adjectives typically express evaluation or judgement (cf. Francis 1993), or attitudinal meaning in Herriman's (2000; 2013) terms. In a similar vein, and with reference to Francis (1993), Hunston & Sinclair (1999: 84) refer to this pattern as one used for expressing evaluation, and in a more recent paper, Hunston & Su (2017: 3) contend that meanings associated with the ***it* v-link ADJ that** pattern "relate to the domain variously termed 'stance', 'attitude', or 'evaluation'".

Following an overview of the distribution of colligates in the two patterns in both languages, the focus will be narrowed down to instances containing an ADJ in the slot. The study is set within an analytical framework concerned with linguistic resources for attitudinal/evaluative meaning (Lemke 1998). Interestingly, the correspondences of these seemingly identical patterns in English and Norwegian are fairly equally divided between (semi-)congruent (4) and non-congruent (5) renderings in source vs. translated texts.

(4)   *It is true that* I sought out Trent, Hopkinson and Hurst ... [MiWa1E]
      Det er sant at jeg oppsøkte Trent, Hopkinson og Hurst ... [MiWa1TN]

(5)   "*It's evident that* the artist who designed this label had never seen the church ... [JW1T]
      "*...selvfølgelig* har kunstneren som tegnet denne etiketten aldri sett kirken. [JW1]
      Lit.: ... *obviously* has the artist who drew this label never seen the church

As the two languages have similar patterns to express similar meanings, non-congruence is really an unexpected alternative. Therefore, one of the aims of this investigation is to discover what triggers congruence and non-congruence in each case, based on the lexical and semantic make-up of the cotext (collocates) and context of the frames, including their attitudinal meaning. It is also of importance to identify stance patterns (words, phrases, expressions) of non-congruence in translation in going from English into Norwegian and vice versa. In a bottom-up approach, a bidirectional translation corpus – the English-Norwegian Parallel Corpus+ – is investigated to establish the potential of the corpus in uncovering semantic, cross-linguistic networks of attitudinal meaning. Drawing on insights from Herriman (2000, 2013), in particular, it is hypothesised that modal adverbs and modal verbs will feature in such networks.

The paper is organised as follows. Section 2 gives an overview of the material used for this study as well as an outline of the extraction process before surveying the overall trends in the two languages regarding which element occurs in the open slot. Section 3 offers a survey of some previous studies of extraposition with ADJ, as well as outlining the framework used for categorising the attitudinal meanings expressed by the patterns. In Section 4, the material is analysed in terms of congruence and according to attitudinal class. A discussion and

---

[2] The code in brackets is the corpus text identifier and stands for author [MoAl = Monica Ali], text number and language [1E]. Identifiers with a capital T in them indicate that the example is from a translation.

[3] Some text identifiers do not specify language, but absence of T means that the text is an original/source text, in this case a text originally written in English, by Doris Lessing [DL]. See further Ebeling & Ebeling (2013) for an overview of texts and text codes in the corpus.

summary of the main findings is offered in Section 5, while Section 6 offers a conclusion and suggests avenues for further research.

2.   Material and method

2.1   The English-Norwegian Parallel Corpus+

The material for this study is extracted from the English-Norwegian Parallel Corpus+ (ENPC+), which is a bidirectional translation corpus of contemporary fiction (1980-2012), containing texts originally written in English and Norwegian as well as their respective translations into Norwegian and English. The ENPC+ is made up of four sub-corpora: English originals (EO), English translations from Norwegian (ET), Norwegian originals (NO), and Norwegian translations from English (NT). The sub-corpora are similar in size and content, each containing 39 texts and approximately 1.3 million words, with the exception of ET which contains approx. 1.4 million words.[4] The ENPC+ is not lemmatised nor part-of-speech tagged; thus, the extraction process had to be carefully thought out, and is described in some detail below.

2.2   Extraction and delimitation of data

To some degree, the frames in the study have flexible syntax in the sense that they enter into (positive and negative) declarative and interrogative sentences, triggering different word order in both English and Norwegian. The study takes this into account and includes all of these syntactic variants with all possible inflected forms of BE and VÆRE. Multiple searches for the frames had to be carried out. Combinations with simple present and past tense forms resulted in 14 different search strings for each of the English sub-corpora and 12 for the Norwegian sub-corpora, and can be summarized as follows:

- it is/was/'s/is n't/was n't (not) * that
- is/was/is n't/was n't it (not) * that
- det er/var (ikke/ikkje) * at
- er/var det (ikke/ikkje) * at

Additionally, to get at verb forms other than the simple present and simple past, the following strings were used, although it has to be noted that they yielded very few results:

- it * be* * that
- it be* * that
- it * n't be* * that
- it * not be* * that
- det * (ikke/ikkje) vært/vore * at
- det (ikke/ikkje) vært/vore * at

The searches do not allow more than one word in the open slot in the frames, apart from the negator *not/n't/ikke/ikkje*. Thus, compared to Larsson (2016a), who studies variability in the patters, the search criteria applied here are much stricter.

---

[4] See e.g. Ebeling & Ebeling (2013) for a more detailed overview of the structure and contents of the ENPC+.

As shown in Table 1, these searches returned 967 hits altogether, distributed fairly evenly across the four sub-corpora, although the translated texts (ET and NT) display a higher frequency of the frames than the original texts. Table 1 further shows the actual colligates that occur in the open slot of the frames. While there appear to be some differences between the languages in terms of preferred colligate, the top two are the same in the original texts, namely ADJ and ADV. The preference for an adjective in the slot is more prominent in the Norwegian material, whereas there is more of a division of labour between ADJ and ADV in English. Other observations that can be made on the basis of Table 1 include the apparent difference between the two languages in the use of V-*ed*, and the difference between original and translated texts in the use of a noun. However, for the purpose of this study, I have chosen to focus on the most frequent pattern (ADJ), for the sake of homogeneity and delimitation of the study.

Table 1. Overview of colligate in pattern in English and Norwegian original and translated texts.

Colligate	EO	ET	NO	NT	Total	
ADJ	78 (37.7%)	151 (50.3%)	145 (73.2%)	165 (63%)	539	
ADV	80 (38.6%)	66 (22%)	33 (16.7%)	29 (11.1%)	208	
V-*ed*	22	33	5	10	70	
NOUN	4	16	5	39	64	
PRON	11	28	8	15	62	
PREP	9	4			13	
Other[5]	3	2	2	4	11	
Total	207	300	198	262	967	
E/N total	507		460			

The *it* BE ADJ *that* / *det* VÆRE ADJ *at* patterns are homogeneous in the sense that, with the exception of one case in EO and three cases in ET,[6] they represent *that*-extraposition with anticipatory *it* / *det* in both languages (as in example (1) above). Before we move on to the contrastive analysis of the 535 instances of the patterns (i.e. 539 minus the four instances mentioned above), some relevant background will be outlined.

3. Previous studies of the patterns

The patterns examined touch on a number of broad topics in linguistics, notably dummy subject constructions, adjectival complementation and stance. It is beyond the scope of this study to offer a survey of these areas in full,[7] but Section 3.1 outlines some relevant publications on stance structures, notably projection clauses that resemble the ones being investigated here. Section 3.1.1 narrows the focus to previous English-Norwegian cross-linguistic studies of similar structures.

---

[5] CONJ, V-*ing*, NO SOURCE, False hit.

[6] In the four cases that were excluded from the study, *that* does not introduce a *that*-clause; it is instead a demonstrative determiner, as in *It is safest that way*. [ToEg1TE].

[7] So-called dummy subject constructions have received a lot of attention, both from an English and a Norwegian perspective (e.g. Collins 1991, Herriman 2000, Kaltenböck 2003, Leira 1970; 1992, Søfteland 2014), as well as from a cross-linguistic perspective (e.g. Ebeling 2000, Gundel 2002, Chocholoušová 2007; 2008, Herriman 2013). Moreover, the anticipatory *it* pattern has been extensively investigated across disciplines and in learner vs. expert data; see Larsson (2016a; 2016b) and references therein.

3.1 Expressing stance by means of *that*-clause extraposition

The English and Norwegian frames can be said to be projection clauses belonging to one branch of the projection framework illustrated in Figure 1. According to Halliday and Matthiessen (2004: 443), "[t]hrough projection, one clause is set up as a representation of the linguistic 'content' of another", i.e. the projecting clause projects the following proposition. Projecting clauses can thus be seen to carry stance, in the sense that they express the speaker's/writer's attitude towards the projected proposition.
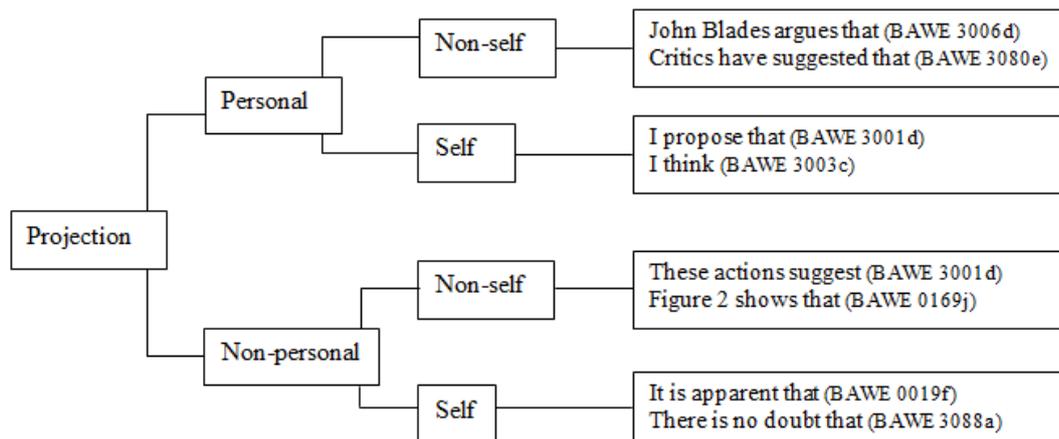

Figure 1. Projection framework (from Ebeling & Wickens 2012: 27, inspired by Hunston 1993; Wickens 2001).

From Figure 1 it becomes clear that the frames investigated in this study are found in the final category of projection, namely that of non-personal self, exemplified by *It is apparent that*. The projection clause serves the purpose of stepping back (non-personal) to disguise the speaker's/author's (self) involvement (cf. Hyland 2005: 176). This is in line with Herriman's (2000: 212) statement that extraposition "allows the writer to omit the source of the attitudinal meaning and to give it an appearance of objectivity and generality". The framework presented in Figure 1 was set up in the context of analysing undergraduate writing from different disciplines and of how projection can reflect disciplinary differences in the way in which students engage with the literature and the object of study (Ebeling & Wickens 2012). In the present paper the framework serves the function of placing the frames in a larger context of projection and stance.

The overarching purpose of the selected frames, as established above, is in agreement with Herriman's (2000) view of extrapositon in general as "one of many linguistic means of expressing attitudinal meaning" (p. 204). Or, as Lemke puts it, this is one of the "[l]exicogrammatical resources [that] enable us to construct attitudinal stances" (1998: 33). Within the framework of Systemic Functional Linguistics, Herriman (2000) discusses and analyses the functional properties of extraposition across different text types in relation to the three metafunctions of language: experiential, interpersonal and textual.

Herriman's (2000) study has a broader scope than the present one in that it is not restricted to extraposition in the form of relational processes (BE/VÆRE) followed by an ADJ and a *that*-clause; thus, not all her findings and observations are directly relevant to the present investigation. However, she draws attention to other ways of expressing attitudinal meaning and how they differ from extraposition. A case in point, and of relevance to this study, is attitudinal meaning expressed by means of modal adverbs or modal verbs such as *probably* and *might*. Herriman, with reference to Halliday (1994), states that "when attitudinal meaning is realised" in this way "it cannot be questioned, negated, or focused in a

pseudocleft […] and is therefore neither made explicit nor negotiable" (2000: 211), in contrast to attitudinal meaning expressed through extraposition.

Also working within the systemic-functional tradition, Lemke (1998), with reference to Francis (1995),[8] suggests that:

> … if we consider occurrences of sentences or clauses of the form:
>
> It is … that …
>
> where *that* introduces an embedded (rank-shifted) noun clause, and the extraposed *it* is followed by an adjective, then the adjectives which occur in this frame fall into a small number of semantic classes, all of which are in some basic sense evaluative epithets. (Lemke 1998: 35-36)

Lemke proposes seven semantic dimensions to reflect evaluative orientations in the *it is ... that* frame and beyond; in fact, the dimensions are meant to account for attitudinal meaning in general. As pointed out by Lemke himself there are parallels between his dimensions and other semantic representations of attitudinal/evaluative meaning, e.g. Francis's categories, which have been coupled with Lemke's dimensions in Table 2.

Table 2. Evaluative orientations in the *it is … that* pattern (adopted from Lemke 1998: 37ff).

Lemke's (1998) semantic dimensions[9]	Francis's (1995) parameters of evaluation[10]
DESIRABILITY / INCLINATION It is simply wonderful that John is coming. It is really horrible that …	Value and appropriacy
WARRANTABILITY / PROBABILITY It is quite possible that … It is very doubtful that …	Modality
NORMATIVITY / APPROPRIATENESS It is quite necessary that … It is entirely appropriate that …	Value and appropriacy
USUALITY / EXPECTABILITY It is quite normal that … It is highly surprising that …	Predictability
IMPORTANCE / SIGNIFICANCE It is very important that … It is really quite trivial that …	Importance
COMPREHENSIBILITY / OBVIOUSNESS It is perfectly understandable that … It is quite mysterious that …	Obviousness
HUMOUROUSNESS / SERIOUSNESS It is just hilarious that … It is ironic that … It is very serious that …	

Well-aware of the fact that other semantic classification schemes of the pattern under discussion and related patterns have been proposed (e.g. Herriman 2000, Groom 2005 (with

---

[8] I have not managed to get hold of this publication by Francis: Francis, G. (1995). Corpus-driven grammar and its relevance to the learning of English in a cross-cultural situation. In Pakir A., (ed.), *English in Education: Multicultural Perspectives*. Singapore: Unipress.

[9] It is interesting to note that most of Lemke's examples in Table 2 include a premodifier in the ADJP. As mentioned above, my search strings do not allow for an extra (premodifying) element; however, a quick search in the ENPC+ shows that this is not a frequent expansion of the pattern in the material at hand.

[10] It is unclear how the two remaining categories in Francis (1995) – Ability and Rationality – fit in with Lemke's dimensions; rationality is said to be mixed, and ability is said to be actional.

reference to Francis *et al*. 1998, Larsson 2016b, Hunston & Su 2017),[11] I will apply Lemke's semantic framework in the classification of the attitudes and evaluations present in the English and Norwegian frames (see Section 4.2), as they seem sufficient and well-suited for the analysis of this specific pattern (with ADJ).

An interesting view on the patterns is proposed by Hunston & Su (2017), who relate them to the concept of a local grammar, i.e. "a grammar of a discourse function […] closely related to performative speech acts" (Hunston & Su 2017: 5). The local grammar, or discourse function, relevant in the current context is that of evaluation, "expressed by adjectives occurring with complementation patterns" (*ibid*.: 7), i.e. the local grammar of evaluation is performed by the two frames in English and Norwegian. Thus, the patterns studied here are form-meaning pairings in the sense of pattern grammar (*ibid*.), and in the formulation of a local grammar of evaluation in the **it v-link ADJ that** pattern, the following meaning elements are proposed (see Hunston & Su 2017: 16-17):

- Hinge (signals that an evaluation is made), e.g. **it is** awful **that** it should end like this
- Evaluation (the evaluative meaning that is made), e.g. it is **awful** that it should end like this
- Target (the entity that is evaluated), it is awful that **it should end like this**

In the sense of Hunston & Su, then, the current study is mainly interested in what type of evaluation element is present in the local grammar of the English and Norwegian patterns, and whether this element influences cross-linguistic congruence.

3.1.1 *Previous cross-linguistic studies*

In two closely related studies, Chocholoušová (2007, 2008) investigates Norwegian and English dummy subjects and their translations into English and German and into Norwegian and German, respectively. Chocholoušová's studies are both broader and narrower in scope than the current one, as well as having a different cross-linguistic focus. First, their scope is broader in the sense that a range of 'dummy' structures are considered and that German translations are part of both studies. On the other hand, her studies are narrower in that they are restricted to dummy subjects in sentence initial position and that the data set is much smaller than in the current investigation.[12] The studies particularly centre around the degree of congruence of dummy subjects in translation, i.e. to what extent the dummy subjects are retained, and to what extent other solutions are chosen. As mentioned in Section 1, the current study, however, considers one particular kind of dummy subject structure, namely extraposition with ADJ + *that*-clause, focusing on the attitude expressed and the cross-linguistic behaviour of the whole stance structure.

In agreement with other studies (e.g. Ebeling 2000; Gundel 2002), Chocholoušová finds that "constructions with dummy subjects are much more frequently used in Norwegian than in English and German, and appear in a greater variety of construction types" (Chocholoušová 2008: 1). Dummy subjects as a group are said to have three basic functions in that they 1) syntactically act as slot-fillers on the level of grammar; 2) topologically shift

---

[11] Hunston & Su's (2017: 3) framework, in particular, is very similar to, and highly compatible with, Lemke's, in that they operate with the following eight groups of adjectives that enter into the **it v-link ADJ that** pattern: 'likelihood', 'obviousness', 'desireability', 'undesirability', 'importance and necessity', 'interest and surprise', 'relevance', and 'other'.

[12] Both studies use the English-Norwegian-German part of the Oslo Multilingual Corpus (http://www.hf.uio.no/ilos/english/services/omc/). While the 2007 study draws on a sub-corpus of 22 Norwegian original texts, amounting to around 290,000 words, the 2008 study draws on 33 English original texts, amounting to around 430,000 words. In comparison, the ENPC+, on which the present study is based, contains around 1.3 million words in both directions of translation (see Section 2.1).

rhematic expressions away from sentence initial position; 3) semantically avoid thematically weak and peripheral elements to be used as sentence subjects (cf. Chocholoušová 2008: 96). While the frames being studied undoubtedly perform these functions, they will be investigated here from the perspective of attitude and stance, and seen in relation to other attitudinal expressions which exist in the two languages.

Of particular relevance to the current study are Chocholoušová's findings regarding the level of congruent translations of extraposed clauses (see Section 4.1). In the translations from Norwegian into English, congruent translations of dummy *det* + extraposed clause were found in 52.7% of the cases (2007: 49), while they were found to be more frequent when going from English into Norwegian (76.7%) (2008: 44). However, it is not clear whether these percentages are evenly distributed across the different types of extraposed clause, i.e. *to*-infinitive clauses, *that*-clauses, *-ing* clauses, and conditional clauses. Nor is it clear whether a translation of an infinitive clause as a *that*-clause is seen as congruent or not.

In cases where the dummy subject is not retained in the translations, full subjects in combination with evidential adverbs such as *selvsagt* 'obviously' or *clearly*, as in example (6), are commonly inserted (2007: 51; 2008: 45).

(6) Det var *tydelig* at hun hadde grått. [EG1]
    Lit.: It was evident that she had cried
    … she had *clearly* been crying. [EG1T]

Chocholoušová (2008: 80) notes that in all the dummy constructions she investigated "congruent translations are preferred in a great majority of cases; and if no structural constraints apply, sentence subjects tend to be preserved in the translation". Compared to Chocholoušová's studies, the current investigation offers a narrower, but more detailed look at the cross-linguistic mechanisms at work in constructions with *it/det* BE/VÆRE ADJ + extraposed *that/at*-clause.

The cross-linguistic study that perhaps bears the most resemblance to the present one is Herriman's (2013) on "The extraposition of clausal subjects in English and Swedish". Using a sample from the English-Swedish Parallel Corpus comprising eight source texts in each direction of translation, Herriman includes subject *that/att*-clauses and *to-/*infinitival clauses in her study, and starts out by measuring the proportion of extraposed:nonextraposed subject clauses in the two languages. However, more pertinent to the present study is her overview of congruent and non-congruent correspondences of extraposed *that/att*-clauses. Herriman finds that around 72% of the English *that*-clauses have a congruent Swedish translation, while only 53.6% of the Swedish *att*-clauses have a congruent English translation. She examines the non-congruent correspondences in more detail and classifies them into five types "depending on how their constituents correspond to the matrix predicate and subordinate clause of extraposition" (Herriman 2013: 245): clause-external evaluations (7), clause-internal evaluations (8), nominal correspondences, zero correspondences and free correspondences.[13]

(7) But I'm sorry she's dead. (PDJ3)
    men det var tråkigt att höra att hon är död. (PDJ3T)
    'But it was sad to hear that she is dead.'

(8) Det var vanligt att dom på söndagarna drack kaffe hos Elna. (SC1)
    'It was usual that they …'
    Eriksson and Oman usually had coffee at Elna's on Sunday. (SC1T)

---

[13] The examples given in Herriman (2013: 247) of the non-congruent categories nominal, zero and free correspondences do not include an adjective, and have therefore been left out here.

In both directions of translation (Swedish to English and English to Swedish), it is the clause-internal evaluations that are most frequently found, and more prominently so in the Swedish-into-English translations. In other words, a larger number of Swedish extrapositions seem to correspond to simple clauses (with clause internal evaluations) in English than vice versa. Herriman suggests several reasons for this, and the one that is accompanied by examples of the type of extraposition examined here has to do with the placement of adverbials in the two languages. Swedish more readily accepts adverbials in initial position followed by extraposition, as in example (9) from Herriman's study (p. 251) where the translator has opted for a single clause in which *offices* "has been made subject by tough movement" (*ibid.*).

(9)   I Drottningholm var det svårt att finna ytterligare kontorsutrymmen… (EGE1)
       'In Drottningholm was it difficult to find more office space'
       Offices were hard to find in Drottningholm… (EGE1T)

Herriman sums up this part of her study in the following way:

> … though extrapositions are translationally equivalent in English and Swedish, they are often used in Swedish when English may have a simple clause. This is due to a number of differences between the languages. These are formal, e.g. the absence of a Swedish equivalent of the English gerund *-ing* form; syntactic, e.g. different word order constraints in English and Swedish, semantic, e.g. the preference for adverbial functions for inanimate semantic roles in Swedish, and pragmatic, e.g. a tendency for Swedish to follow the information principle more strictly than English. (2013: 253)

Although focusing on slightly different translation mechanisms and categories of non-congruence, the current analysis will be undertaken with Herriman's findings and reasons for non-congruence in mind when identifying stance patterns (words, phrases, expressions) of non-congruence between English and Norwegian.

Finally, based on both Chocholoušová's (2007, 2008) and Herriman's (2013) findings, we can hypothesise that the English frame will have more congruent translations into Norwegian than vice versa. We can also predict that Norwegian translations will have more non-congruent sources than English translations.

4.   Analysis

4.1   Congruence

One of the aims of this study is to investigate to what extent "identical" structures in the two languages correspond to each other in translation. Unlike Herriman, we are interested in all four directions of correspondence. This means that in addition to recording what happens when going from English originals to Norwegian (EO→NT) and from Norwegian originals to English (NO→ET), the elements in the source texts that give rise to the patterns in the translated texts will also be considered, i.e. searches for the patterns are made in the NT and ET sub-corpora to establish their sources in the English and Norwegian original texts (NT←EO and ET←NO, respectively). This will be measured in terms of congruence. A congruent translation or source is understood as an instance where the patterns correspond to each other, i.e. they "formally correspond to each other and are explicitly expressed" (Ebeling 2015: 37), as in example (10), where the Norwegian pattern has a congruent English

translation (NO→ET), and in example (11), where the Norwegian pattern has a congruent English source (NT←EO).

(10) *Det er ikke viktig at* jeg er sjalu på min datters far mens hun står bak ham ... [PeRy1N]
It is not important that I am jealous of my daughter's father as she stands behind him ... [PeRy1TE]

(11) *Det er mulig at* jeg kjente Booth." [AnCl1TN]
It is possible that I knew him." [AnCl1E]

The category "semi-congruent" is less strict in that it does not require full formal correspondence; however the pattern is still recognizable in the other language but the slight formal difference lies in syntactic flexibility, the use of a different verb or verb form, or the presence of an adverb (typically a particle) (in Norwegian). These differences are illustrated, respectively, in (12), where the English translation has a Ø-*that* clause, in (13), where the Norwegian translation of *will be* is the simple present tense *er* 'is', and in (14), where the particle *jo* 'of course' has been added in the Norwegian translation.

(12) Det er mulig *at* jeg er født sjarmløs. [EHA1]
It's possible [Ø] I was born without charm. [EHA1T]

(13) It *will be* clear that I do not wish to enter upon marriage burdened with debt, ... [RDA1]
Det *er* innlysende at jeg ikke ønsker å inntre i ektestanden tynget av gjeld, … [RDA1T]

(14) And it was true that this ordinary businessman in his Ford Cortina,... [FW1]
Og det var *jo* sant at denne alminnelige forretningsmannen i sin Ford Cortina , ... [FW1T]
Lit.: And it was of course true that …

The non-congruent category refers to instances where there is an overt translation or source which is formally further from the patterns than in cases involving syntactic flexibility, a change of verb (form), or presence of a particle. An example of a non-congruent correspondence is given in (15), where existential *det* in combination with an indefinite NP is used in the Norwegian translation.

(15) "It's possible that the victim tried to stop the opening from happening at all. [AnCl1E]
"*Det* er *en mulighet for at offeret prøvde å forpurre hele utstillingsåpningen.* [AnCl1TN]
Lit.: There is a possibility for that the victim …

As can be seen in Table 3 below, there is more non-congruence than full congruence in the material overall (45.6% vs. 35.9%).[14] However, it can also be noted that non-congruence is mostly found in the English translations from Norwegian (NO→ET; 49%) and in Norwegian translations from English originals (NT←EO; 59.4%). In other words, non-congruence seems

---

[14] If we collapse the categories congruent and semi-congruent, congruence is predominant in going from original to translated texts, which in accordance with both Chocholoušová's (2007, 2008) and Herriman's (2013) studies. In fact, it is only in the direction NT←EO that this is not the case.

to prevail when starting the search in the Norwegian texts.[15] It is hard to determine why this should be the case, but in what follows I will attempt to give a detailed account of what happens cross-linguistically when comparing English and Norwegian on the basis of translation data.

Table 3. Degree of congruence in all directions of correspondence between English and Norwegian.

Congruence	EO → NT	NO → ET	NT ← EO	ET ← NO	Total
Congruent	43 (55.8%)	53 (36.6%)	43 (26.1%)	53 (35.8%)	192 (35.9%)
Semi-congruent	13(16.9%)	21 (14.4%)	24 (14.5%)	41 (27.7%)	99 (18.5%)
Non-congruent	21 (27.3%)	**71 (49%)**	**98 (59.4%)**	54 (36.5%)	244 (45.6%)
Total	77	145	165	148	535

The next step in the analysis will be to classify the patterns according to their attitudinal/evaluative meanings (see Table 2; cf. Lemke, 1998), in order to find out to what extent the meanings expressed have an impact on congruence.

4.2   Attitudinal meanings

The most frequently occurring adjectives in the slots vary slightly across the different sub-corpora. The top three in the two English sub-corpora (EO and ET) are *true*, *clear* and *possible*, while for NO they are *tydelig* 'clear', a form of *god* 'good' and *rart* 'strange', with *mulig* 'possible' in fourth place. Finally the top three in NT are *tydelig* 'clear', *mulig* 'possible' and *sant* 'true'. *True*, *clear* and *possible* were also found to be among the five most frequent adjectives in Larsson's (2016a: 71) material, i.e. the Louvain Corpus of Research Articles.

These adjectives do in fact reflect the main tendencies in the sub-corpora with regard to attitudinal class: *clear*/*tydelig* represent category F (Comprehensibility/Obviousness), which is the favoured category in three of the four sub-corpora. *Possible*/*mulig* represent category B (Warrantability/Probability), also one of the most prominent categories, while *true*/*sant* represent category C (Normativity/Appropriateness), which is a favoured category in the two English sub-corpora and NT. *God* 'good' and *rart* 'strange' are particularly frequent in NO and represent categories A (Desirability/Inclination) and D (Usuality/Expectability), respectively. Table 4 gives an overview of the distribution of attitudinal meanings in the material. For convenience, all but one of the categories (G) are exemplified by congruent correspondences from the ENPC+.

Focusing on the original texts first, it can be seen that, proportionally, both English and Norwegian favour category F (Comprehensibility/Obviousness). However, while English has category B (Warrantability/Probability) in second place and category C (Normativity/Appropriateness) in third place, Norwegian has category D (Usuality/Expectability) in second place and category A (Desirability/Inclination) in third place. The latter two are marginal categories in the English original texts in the ENPC+. In Norwegian there is a slightly more even distribution across the less favoured categories.

---

[15] In Section 3.1.1, a similar trend was noted by Herriman (2013), who finds more congruent correspondences of extraposed *that*-clauses in the English-Swedish direction of translation than vice versa.

Table 4. Attitudinal meanings expressed by the patterns in the four sub-corpora, including examples.[16]

	Meanings	EO	ET	NO	NT	Example from ENPC+
A	DESIRABILITY / INCLINATION	4 (5.2%)	14 (9.5%)	28 (19.3%)	27 (16.4%)	It was terrible that her mother was so young. [MoAl1E] Det var grusomt at moren var så ung. [MoAl1TN]
B	WARRANTABILITY / PROBABILITY	21 (27.3%)	49 (33.1%)	21 (14.5%)	38 (23%)	"It's possible that Jon Moreno took his own life," Sejer said. [KaFo1TE] - Det er mulig at Jon Moreno tok sitt eget liv, sa Sejer. [KaFo1N]
C	NORMATIVITY / APPROPRIATENESS	16 (20.8%)	28 (18.9%)	18 (12.4%)	21 (12.7%)	It was not right that a young woman from a good family was so uninhibited. [HW2T] Det var ikke riktig at ei ung jente av god familie var så uten hemninger. [HW2]
D	USUALITY / EXPECTABILITY	4 (5.2%)	18 (12.2%)	29 (20%)	17 (10.3%)	In fact, it's strange that it doesn't happen more often." [AnHo1TE] Egentlig er det rart at det ikke skjer oftere. [AnHo1N]
E	IMPORTANCE / SIGNIFICANCE	9 (11.7%)	10 (6.8%)	11 (7.6%)	7 (4.2%)	… it is convenient that one witness to Cill's murder is dead ... [MiWa1E] … det er beleilig at et av vitnene til drapet på Cill er død ... [MiWa1TN]
F	COMPREHENSIBILITY / OBVIOUSNESS	22 (28.6%)	29 (19.6%)	37 (25.5%)	55 (33.3%)	"And then, of course," he said, when it was obvious that I wasn't going to answer, "we come to last night. [TaFr1E] "Og så," sa han da det var tydelig at jeg ikke kom til å svare, "kom vi til i går kveld. [TaFr1TN]
G	HUMOUROUSNESS / SERIOUSNESS	1		1		... reflecting that it was ironic that it was he who should meet death by chance ... [DF1] ... tenkte på ironien i at det var han som møtte et slikt tilfeldig endelikt ... Lit: 'thought of the irony in that it was…' [DF1T]
		77	148	145	165	

4.3 Attitudinal meaning and congruence in EO→NT and NO→ET

On the basis of the different preferences mentioned for English and Norwegian, it may be hypothesised that non-congruence is mainly found within the Usuality/Expectability (D) and Desirability/Inclination (A) categories when going from Norwegian into English. In the other direction of translation (EO→NT), but perhaps to a lesser extent, more instances of non-

---

[16] A note on dispersion: Even in the attitudinal classes with fewest attested examples, e.g. A/EO, D/EO and E/NT, the instances are dispersed across different texts. In the case of A/EO, the four instances are found in two different texts, while in the case of D/EO, the four instances are found in four different texts.

congruence in the Warrantability/Probability (B) and Normativity/Appropriateness (C) categories may be expected. The charts in Figures 2 and 3 show this to be partly true.
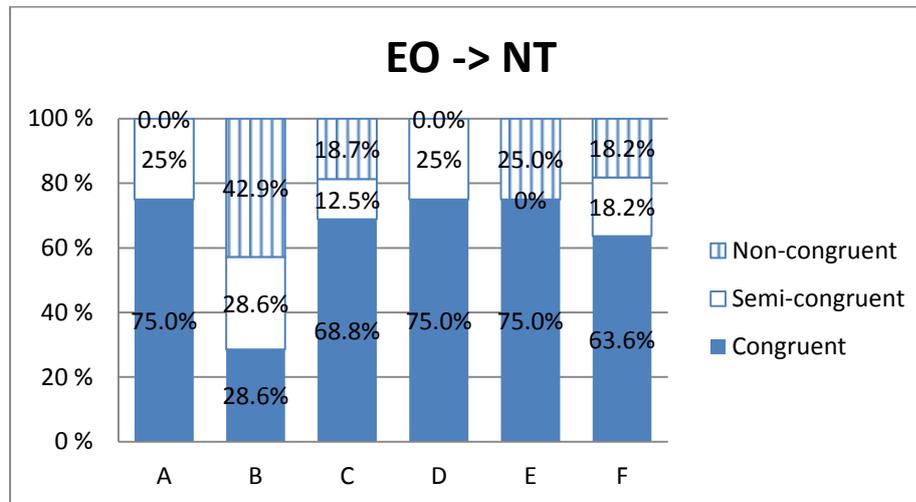

Figure 2. Proportions of congruent, semi-congruent and non-congruent correspondences according to semantic category from English into Norwegian (excluding category G, as it only has one occurrence).[17]

In going from English originals to Norwegian translations the highest proportion of non-congruent correspondences is indeed found in category B, while category C does not behave as expected in this respect; curiously, as shown in Figure 3, category B is one of the categories in which non-congruence is most prominent in going from Norwegian into English as well. I will return to some of these unexpected results below.
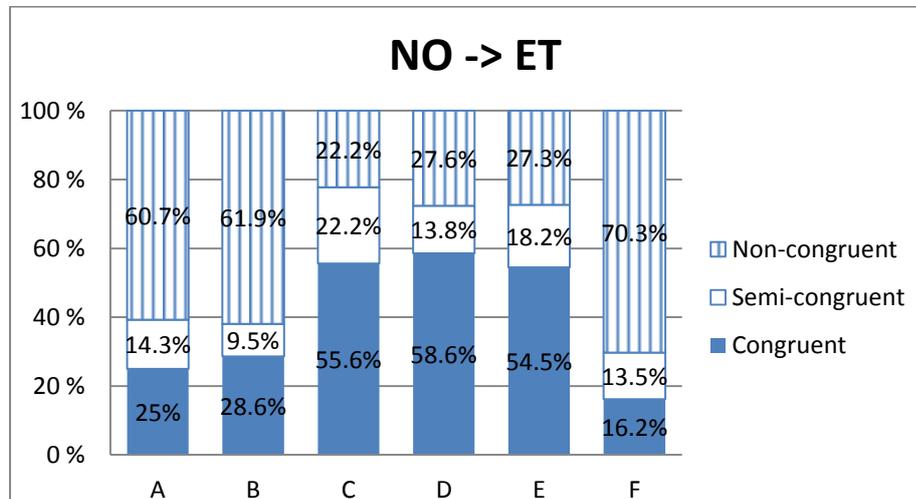

Figure 3. Proportions of congruent, semi-congruent and non-congruent correspondences according to semantic category from Norwegian into English (excluding category G, as it only has one occurrence).

There is generally a higher proportion of non-congruence in all categories in NO→ET as compared to EO→NT. As was the case in the EO→NT direction of translation, there is a high proportion of non-congruence in only one of the two categories in which it was predicted, viz. category A. Category D, on the other hand, did not yield the expected proportion of non-congruence. In the NO→ET direction of translation Category F (Comprehensibility/Obviousness) contains the largest proportion of non-congruence,

---

[17] A note on dispersion for Figures 2-5: Instances of congruent, semi-congruent and non-congruent in all attitudinal categories are distributed across several different texts, as attested by the examples cited below.

reaching around 70%. This is contrary to expectations, since the English original texts also make frequent use of this category, and one would expect this to be reflected in the translations.

In the following, I will take a closer look at what happens in the non-congruent cases in category B in both directions of translation and in category D in the NO→ET direction. Moreover, category F will be dealt with in some detail, as it was not expected that non-congruence would feature as prominently as it does, particularly in the NO→ET direction.

4.3.1 *Non-congruence in category B (Warrantability/Probability) in EO→NT and NO→ET*
Category B is highly interesting in terms of congruence; proportionally it is a more important category for EO than NO, but the percentage of non-congruence for this category is in fact higher (approx. 62%) when going from Norwegian to English, than when going from English to Norwegian (approx. 43%).

Examining the non-congruent cases in both directions of translation, similar patterns of translation can be observed, featuring modal auxiliaries and/or adverbs (+passive), as shown in examples (16), (17) and (18). In (16) the Norwegian pattern is rendered as a single main clause with the modal *may*, while the Norwegian pattern in (17) is translated into the adverb *possibly*. Finally in (18), the English pattern is translated into a similar pattern in Norwegian, but with different elements in it, i.e. the modal *kan*, the modal particle/adverb *jo* and the so-called *s*-passive form *tenkes* 'be thought'.

(16) Da *er det ikke sikkert at* dere trenger noen profil. [AnHo1N]
Lit.: Then is it not certain that you need any profile
Then you *may* not even need a profile. [AnHo1TE]

(17) På den annen side var hun gravid i tredje måned, og *det var mulig at* det kunne ha en slags forbindelse med mordet ... [FC1]
Lit.: … and it was possible that …
On the other hand, she was three months' pregnant, and *possibly* that had some connection with the murder ... [FC1T]

(18) ... "and I suppose *it's possible that* he knew about the market-square CCTV ... [PeRo1E]
... "og *det kan jo tenkes* at han visste om overvåkingskameraene på markedsplassen ... [PeRo1TN]
Lit.: and it could of course be thought that he knew …

Modals and adverbs are the two main types of non-congruent correspondence patterns in category B. This suggests, not surprisingly perhaps, that the two languages have similar linguistic means at their disposal to express Warrantability/Probability, of which the *it* BE ADJ *that* / *det* VÆRE ADJ *at* is but one of three/four possibilities (see also Herriman 2000; 2013).

4.3.2 *Non-congruence in category D (Usuality/Expectability) in NO→ET*
Of the 29 instances in the Usuality/Expectability category in NO, eight have a non-congruent correspondence in the English translations. From the numbers in Table 4, more non-congruence would have been expected, as this seems to be a marginal category for the *it* BE ADJ *that* pattern in EO. The relatively low number of non-congruence could be related to the fact that there does not seem to be a clear-cut non-congruent equivalent, i.e. it is not obvious which English expression(s) to choose instead. The non-congruent correspondences include ADJ + (*for* x) *to*-clause (three occurrences), Noun + *that*-clause (two occurrences), and one

occurrence each of a main clause, a main clause with a modal, and an exclamation. Examples of the first and last of these are given in (19) and (20), respectively.

(19)   - *Det er uvanlig at* noen føler behov for hjelp når de er maniske. [AnHo1N]
       Lit.: It is unusual that someone feels …
       "*It's unusual for people to* feel the need for help when they're in a manic phase. [AnHo1TE]

(20)   *Var det ikke urettferdig at* livet en gang tok slutt? [JG1]
       Lit.: Wasn't it unfair that life one time took end
       How unfair that life had to end! [JG1T]

4.3.3 *Non-congruence in category F (Comprehensibility/Obviousness) in NO→ET*
Non-congruence in the Comprehensibility/Obviousness category is particularly prominent in the NO→ET translation direction (26 out of 37 instances). In the other direction (EO→NT) of correspondence only four out of the 22 instances are non-congruent, which is more in line with what would be expected. I will therefore focus on how this non-congruence manifests itself in the English translations.

Typically, and in 15 out of the 26 cases, English has an adverb in the translation of the Norwegian pattern, as shown in example (21).

(21)   Trass i arsenalet av skumle redskaper, *er det tydelig at* de er varsomme med treskulpturen. [ToEg1N]
       Lit.: … is it obvious that …
       Despite the arsenal of alarming instruments, they have *obviously* treated it with great care. [ToEg1TE]

The only other observable tendency is to have a noun and a *that*-clause, as in example (22), but as four out of the five instances come from the same text, this may not be assumed to be a general trend.

(22)   "*Er det tilfeldig at* Cistercienserordenen anla Lysekloster i 1146, ... [ToEg1N]
       "And *is it a coincidence that* the Cistercian Order built Lyse Abbey in 1146, ... [ToEg1TE]

In the remaining four instances, the translators have opted for four different clause types: *wh*-clause, main clause, existential clause and a *for* x *to*-clause.

In other words, in going from Norwegian into English, non-congruence in this category partly resembles category B in the preference for using adverb(ial)s as correspondences of the *det* VÆRE ADJ *at* pattern. It is therefore tempting to conclude that English adverbials have a wider range of attitudinal meaning and use than Norwegian adverbials. In English they seem, in many cases, to be preferred choices and to perform similar functions to both categories B and F.

4.4   Attitudinal meaning and congruence in ET←NO and NT←EO

When looking at the preferred categories in the translated material (see the unshaded columns in Table 4), some discrepancies between EO and ET and between NO and NT can be noted, although the internal ranking of the different categories show some similar tendencies. All sub-corpora, apart from ET, show a preference for category F, while category B is ranked in

first or second position in all but NO. EO and ET have category C in third place, while NO and NT have A. For the less frequently used categories there are some minor differences, but category E is fairly stable, being ranked 6th in all but EO.

Although some differences between translated and non-translated English (EO vs. ET) and Norwegian (NO vs. NT) can be observed, it is not obvious that these are due to translation effects (at least not induced by the patterns under study). For example, category B is proportionally more prominent in NT than in NO. It is also a category that is prominent in EO and it is thus tempting to conclude that this must be because of source language influence (i.e. influence of English on the Norwegian translation). However, the number of instances in category B is higher in NT (33) than in EO (21), pointing to the fact that there are more complex relationships at play. A closer look is therefore warranted in order to establish which structures in the source languages give rise to the *it* BE ADJ *that*/*det* VÆRE ADJ *at* patterns in the English and Norwegian translations.

In Table 3, it was shown that the proportion of non-congruent correspondences in the sources is around 59% in the direction NT←EO, while in the direction ET←NO it is around 36%. The next step is to find out what is happening across the attitudinal categories in terms of congruence, when starting from the translated texts. An overview is given in Figures 4 and 5.
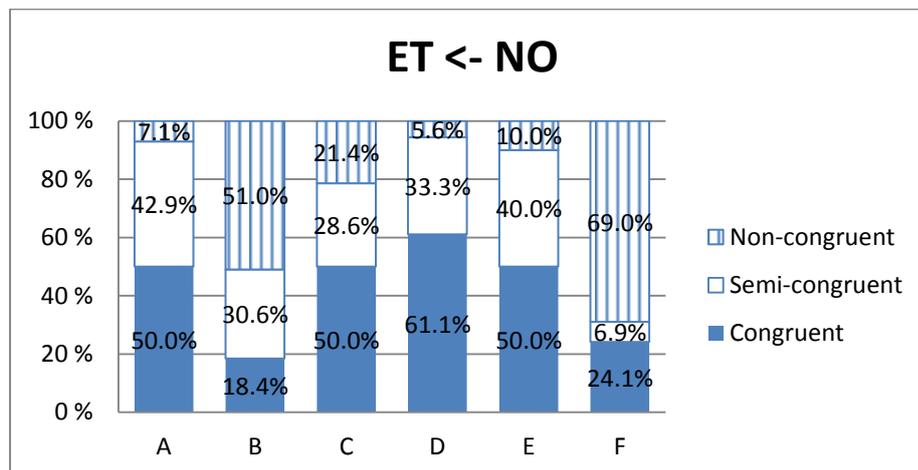

Figure 4. Proportions of congruent, semi-congruent and non-congruent correspondences according to semantic category in English translations from Norwegian sources.
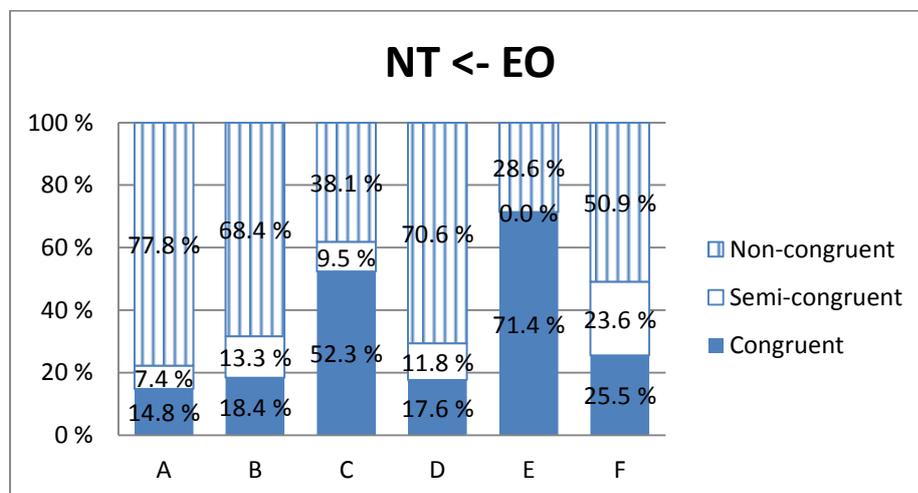

Figure 5. Proportions of congruent, semi-congruent and non-congruent correspondences according to semantic category in Norwegian translations from English sources.

It can be observed from Figures 4 and 5 that in all categories but F, NT from EO proportionally produces more non-congruence than ET from NO. It is also interesting to note that non-congruence also features markedly in category B in both directions, which was also the case when starting from original texts (see Figures 2 and 3 above). Categories A and E are also interesting in the NT←EO direction. I will start by looking at categories B and F in both directions to see whether similar mechanisms are at play between translation and source text.

4.4.1 *Non-congruence in category B (Warrantability/Probability) in ET←NO and NT←EO*
The two most prominent sources of non-congruence between NT and EO, with eight and six instances, respectively (out of 26, see Table A in the Appendix), are main clause with a modal auxiliary and existential *there* sentences, as shown in examples (23) and (24), respectively.

    (23) *"Det er mulig at* sjefen min vil ha et ord med deg om disse to menneskene som kom på besøk." [PeRo2TN]
           Lit.: It is possible that …
           "My boss *might* want to have a word with you about those two people who came to visit." [PeRo2E]

    (24) Han ville gi oss skylden for hele greia, og *det var utenkelig at* han ville la det ligge. [TaFr1TN]
           Lit.: … it was unthinkable that …
           ... he would blame the whole thing on us, and *there wasn't a chance in hell* he would let it lie. [TaFr1E]

In example (23), the subject of the *at*-clause in the translation is the subject of the overall structure in the English source text and *might* arguably gives rise to the Norwegian pattern, viz. *det er mulig at* 'it is possible that'. Similarly in (24), the Norwegian translation *det var utenkelig at* 'it was unthinkable that' captures the meaning of the English existential clause *there wasn't a chance in hell*. There are no other non-congruent correspondence types that stand out in particular, perhaps with the exception of what I have called *to*-infinitive, with three occurrences. In example (25), the subject of the *at*-clause in the translation – *Grace* – comes from the subject of the overall structure in English followed by an ADJ and a *to*-infinitive clause. This is a kind of alteration that, according to Quirk *et al*. (1985: 1228), is found with certain adjectives, including (*un*)*likely*; the adjectives "occur with a corresponding construction with extraposition of a *that*-clause: Jill is *likely* to attend. ~ It is *likely* that Jill will attend" (*ibid*.).

    (25) Ikke desto mindre *er det usannsynlig at* Grace ville åpne døren etter mørkets frembrudd ... [MiWa1TN]
           Lit.: … is it unlikely that Grace would open…
           Nevertheless, *Grace was unlikely to open* her door after dark ... [MiWa1E]

When looking at the non-congruent sources of the English translations (ET←NO), there are two other main triggering factors at work, namely a similar pattern with the Norwegian *s*-passive (26) (see also example (18) above), with eight occurrences, and an adverb (27), with seven occurrences.

(26) *Was it conceivable that* the rumour was true ... [BHH1TE]
Kunne det tenkes at ryktet talte sant ... [BHH1N]
Lit.: Could it be thought that …

(27) "*It's possible that* all the changes in climate were what caused it to disintegrate. [ToEg1TE]
"*Sannsynligvis* ble alle de klimatiske forholdene for mye for den. [ToEg1N]
Lit.: Probably became all the climatic conditions …

The findings for non-congruent correspondences in category B are in fact in line with what was found when going from EO into NT and from NO into ET, with the exception of existential sentences. Existential sentences were, however, found in non-congruent cases of category F.

4.4.2 *Non-congruence in category F (Comprehensibility/Obviousness) in ET←NO and NT←EO*

Although the Norwegian sources of the non-congruent cases in category F are varied, there is one pattern with a phrasal verb instead of VÆRE + ADJ that gives rise to *it* BE ADJ *that* in seven out of the 20 cases (see Table A in the Appendix): *det går frem at* 'it emerges that' (Lit.: it goes forth that) or *det skinte gjennom at* 'it shone through that'. Moreover, an adverb is the source of the pattern in the English translations in five cases. The remaining eight cases range from instances of complete rewriting in the translation (28) to one instance of using the non-extraposed version of the pattern with an initial *at*-clause (29).

(28) Because *it was clear that* neither furnishings nor ornaments came from chain stores. [EG2T]
For ingen skulle fortelle ham at møblene var fra Ikea eller at nipsgjenstandene kom fra en eller annen basar på Grønland! [EG2]
Lit.: For no one should tell him that the furniture was from IKEA or that the ornaments came from one or other bazaar in Grønland

(29) It is understandable *that they couldn't find us*. [ToEg1TE]
At de ikke har funnet oss, er forståelig. [ToEg1N]
Lit.: That they not have found us, is understandable

When starting from Norwegian translations in category F, there are 29 cases of non-congruent correspondence, of which 19 have an English adverb as source (see Table A in the Appendix). The remaining 10 instances show a variety of correspondences, including modal verbs, Ø correspondence, the passive voice and two instances of the structure *it was/seemed ADJ to X that*, which has been deemed (formally) non-congruent, although it resembles the patterns studied here.

4.4.3 *Non-congruence in categories A (Desirability/Inclination) and D (Usuality / Expectability) in NT←EO*

When searching in the Norwegian translations for *det* VÆRE ADJ *at*, it is found, as shown in Figure 5 above, that there is a non-congruent English source in 77.8% of the cases for category A and in 70.6% for category D.

The non-congruent sources in category A vary across many of the types already discussed, including modal auxiliaries, *it* BE ADJ *for* X *to* and *to*-infinitive clauses. There are also a couple of correspondence types that, albeit marginal, do not occur elsewhere in the

material, namely two instances of ADJ+Ø-*that*-clause, as in (30), and the use of the verb *mind*, as in example (31).

(30) Ingen grunn til at de må stå opp allerede - *det er bedre at* de får sove videre. [StGa1TN]
Lit.: … it is better that they get sleep on
There's no reason for them to get up yet. *Better* they remain asleep. [StGa1E]

(31) "*Er det greit at* vi snakker her?" spurte Stuart. [AnCl1TN]
Lit.: Is it fine that we talk here
"*Do you mind* chatting here?" Stuart said. [AnCl1E]

Category D is comparatively small, with only 17 occurrences altogether (see Table 4), but the fact that 12 out of these have non-congruent English sources makes it interesting in the current context. Again there does not seem to be one major correspondence type accounting for the high proportion of non-congruent cases. The *it* BE ADJ *for* X *to* and *it* SEEM ADJ *to* x *that* constructions account for two instances each. In addition, there are two instances of Ø correspondence, one case with an adverb and one with a main clause capturing the content of *it* BE ADJ. There is one case of ADJ + *thing*, as shown in example (32), and finally, there are two cases of other clause types (*wh-* and *-ing* clause) which may be seen as variants of the *it* BE ADJ *that* pattern.

(32) På mange måter *var det rart at* Lexie hadde klart seg så lenge som hun hadde. [TaFr1TN]
In a lot of ways, *the amazing thing was* that Lexie had made it this far. [TaFr1E]

5.   Discussion and summary of findings

Above, only the most frequent non-congruent correspondences were discussed in detail. They do, however, seem to give a relatively accurate picture of what resources are available in the two languages (in the material at hand) to express attitudinal meanings of the kinds represented by the *it* BE ADJ *that*/*det* VÆRE ADJ *at* patterns.

The overall tendencies are quite clear when looking at the 244 non-congruent cases together, regardless of direction of correspondence and attitudinal class. As shown in Table 6, ADV is the most common non-congruent correspondence overall. However, the tendency is stronger for an English ADV to correspond to the Norwegian pattern than vice versa (see also Table A in the Appendix).

Table 6. Distribution of all non-congruent correspondence types in the material.

Non-congruent correspondence type	English↔Norwegian[18]	Norwegian↔English	No.
ADV	22	45	67
modal	5	25	30
main clause	4	21	25
inf.-clause	8	7	15
it BE ADJ *for* x *to* (in N↔E only)	--	16	16
NOUN	3	11	14

---

[18] In the two columns in the middle of Table 6, the directions of correspondence have been collapsed into a two-way correspondence, starting from English (English↔Norwegian includes both EO→NT and ET←NO) and Norwegian (Norwegian↔English includes both NO→ET and NT←EO).

s-passive (in E↔N only)	10	--	10
if-clause	2	7	9
Ø	6	8	14
existential (in N↔E only)	--	7	7
phrasal verb (in E↔N only)	7	--	7
other[19]	8	22	30
Total	75	169	244

The most general types that recur in all directions of correspondence are the top four on the list given in Table 6. Interestingly, and as seen above, some of them are preferred choices in a particular attitudinal class, e.g. ADV in category F, modal in category B, the Norwegian *s*-passive in category B and English existentials also in category B.[20] The first of these choices (ADV) may have to do with the ease with which adverbial obviousness markers can be used for category F, and in the case of the three typical means of expressing attitudinal meaning in category B, the modals are self-evident. As for the *s*-passive and why it should feature in this category, this most likely has to do with the fact that the *s*-passive is typically accompanied by a modal verb (conveying the writer's stance), as *kan* 'can' in (33).

(33)   "And now you think *it's conceivable that* he didn't disappear after all?" he said. [JoNe1TE]
"Og nå mener du at *det kan tenkes at* han ikke er så forsvunnet likevel?" sa han. [JoNe1N]
Lit.: And now mean you that it can be thought that …

This leaves us with the question why existentials give rise to the Norwegian *det* VÆRE ADJ *at* pattern. Quite clearly it is connected with the fact that the noun in these existential constructions expresses an attitude, as shown in example (34). Although a congruent Norwegian correspondence could in fact have been used here: *det er en mulighet for at*, the translator opted for the arguably lighter option with an adjective, *mulig* 'possible'.

(34)   *Det er mulig at* hun er sammen med en av dem nå. [MiWa1TN]
There's a *possibility* she may be with one of them now. [MiWa1E]

The remaining correspondence types are not tied to a specific attitudinal class to the same extent as the ones mentioned, suggesting that they are more general patterns of expressing a variety of attitudinal meanings.

Although it may be concluded that the different ways of expressing attitudinal meaning work well as each other's correspondences in translation, it should be noted that they may not be considered full equivalents in all contexts and at all levels of interpretation. As pointed out by Herriman (2000: 211):

> … when attitudinal meaning is realised by other linguistic means such as modal adverbs and verbs, e.g. *probably* and *might*, it cannot be questioned, negated, or focussed in a pseudocleft clause […] and is therefore neither made explicit nor negotiable (Halliday 1994: 354f). It would seem, then, that one very significant effect of realising attitudinal meaning as a finite clause is the fact that this enables speakers/writers to negotiate their attitudinal meaning with their addressees in a way which cannot be done by single lexical items. […] Extraposition thus provides a variety of ways of negotiating attitudinal meaning.

---

[19] See Table A in the Appendix for items included in the "other" group.
[20] There is one occurrence of a Norwegian existential, also in category B (see example 15).

The fact that a non-congruent correspondence of the patterns is chosen in between 27% and 59% of the cases in the ENPC+ may suggest that translators are not too concerned with, or indeed aware of, such differences between attitudinal expressions.

6.   Conclusion and further research

Taking the patterns *it* BE * *that*/*det* VÆRE * *at* as its starting point, this study first set out to outline the preferred elements in the open slot in English and Norwegian. ADJ turned out to be the predominant colligate, steering the study towards its main focus on *it*/*det* extraposition with ADJ + *that*-clause.

One of the purposes of this investigation was to demonstrate the potential of a bidirectional translation corpus in exploring the level of congruence between similar-looking patterns with similar functions in two languages. The structure of the ENPC+ facilitates contrastive studies going from originals to translation in two languages. An added value, offering an even more complete picture in terms of congruence between patterns across languages, is the possibility of looking up the patterns in the translations to find out what the source of the patterns was in the originals. The primary focus was on the unexpected cross-linguistic alternative, namely non-congruence, as this is a good way of establishing linguistic networks of attitudinal meaning across languages.

Overall, 45% non-congruence in the use of the patterns investigated was recorded between the languages. Not only is there a marked difference between the directions of correspondence in this respect, but also between attitudinal classes. The correspondence network, or paradigm, of the English and Norwegian patterns is fairly broad, pointing to the fact that there are a number of linguistic resources in the two languages that can be put to this use. Most notably, perhaps, were the quite stable trends in categories B and F in terms of choice of non-congruent correspondences.

The overall results also imply that there is a lot of similarity between the two languages in the use of these patterns (55% (semi-)congruence). The similarity is further attested in Table 6 and Table A in the Appendix, in which it is shown that similar alternative expressions are found in both languages. In other words, the current study has established that there are both similarities and differences in the "grammar of evaluation" in English and Norwegian.

The findings uncovered in this study suggest that, while English and Norwegian have similar means of expressing attitudinal meanings, the two languages have their preferred ways of doing so both in terms of individual adjectives and attitudinal class. In this context it is also important to stress that the tendencies reported here are only valid for the limited material studied, representing only one broad genre, namely contemporary fiction. A natural extension for future research would therefore be to investigate the various attitudinal expressions on the basis of large monolingual reference corpora of English and Norwegian, in order to establish with more certainty the cross-linguistic relationship between the different expressions. Moreover, a further exploration of potential differences between the different stance expressions within and across the languages would also be a welcome extension of this study.

Acknowledgements

References

Chocholoušová, B. 2007. *Norwegian det-constructions and their Translation Correspondences in English and German: A Contrastive Corpus Based Study of Dummy Subjects*. Unpublished Master's Diploma Thesis, Masaryk University.

Chocholoušová, B. 2008. *There and It in a Cross-linguistic Perspective. A Parallel Corpus Study of English dummy Subjects and Their Translation Equivalents in Norwegian and German*. Unpublished MA thesis, University of Oslo.

Collins, P. 1991. *Cleft and pseudo-cleft constructions in English*. London: Routledge.

Ebeling, J. 2000. *Presentative Constructions in English and Norwegian. A Corpus-based Contrastive Study*. Oslo: Acta Humaniora.

Ebeling, J. and Ebeling, S.O. 2013. *Patterns in Contrast*. Amsterdam: John Benjamins.

Ebeling, S.O. and Wickens, P. 2012. Interpersonal Themes and Author Stance in Student Writing. In *English Corpus Linguistics: Looking Back, Moving Forward; Papers from the 30th International Conference on English Language Research on Computerized Corpora (ICAME 30), Lancaster, UK, 27-31 May* 2009, S. Hoffmann, P. Rayson and G. Leech (eds), 23–40. Amsterdam: Rodopi.

Francis, G. 1993. A Corpus-driven Approach to Grammar – Principles, Methods and Examples. In *Text and Technology: In Honour of John Sinclair*, M. Baker, G. Francis and E. Tognini-Bonelli (eds), 137–156. Amsterdam: John Benjamins.

Francis, G., Manning, E. and Hunston, S. 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.

Groom, N. 2005. Pattern and Meaning Across Genres and Disciplines: An Exploratory Study. *Journal of English for Academic Purposes*, 4(3), 257–277.

Gundel, J. 2002. Information Structure and the Use of Cleft Sentences in English and Norwegian. In *Information structure in a cross-linguistic perspective*, H. Hasselgård, S. Johansson, B. Behrens, and C. Fabricius-Hansen (eds), 113–128. Amsterdam: Rodopi.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M.A.K. and Matthiessen, C.M.I.M. 2004 [3rd ed.]. *An Introduction to Functional Grammar*. London: Arnold.

Hasselgård, H. 2016. The Way of the World: The Colligational Framework "the N1 of the N2" and its Norwegian Correspondences. *Nordic Journal of English Studies* 15(3), 55–79, ed. by L. Nordrum, S.O. Ebeling and H. Hasselgård, *Languages in Contrast 20 Years on*.

Herriman, J. 2000. The Functions of Extraposition in English Texts. *Functions of Language* 7:2, 203–230.

Herriman, J. 2013. The Extraposition of Clausal Subjects in English and Swedish. In *Advances in Corpus-based Contrastive Linguistics. Studies in Honour of Stig Johansson*, K. Aijmer and B. Altenberg (eds), 231–259. Amsterdam: John Benjamins.

Hunston, S. 1993. Projecting a Sub-culture: The Construction of Shared Worlds by Projecting Clauses in Two Registers. In *Language and Culture*, D. Graddol, L. Thompson and M. Byram (eds), 98–112. Clevedon: BAAL and Multilingual Matters.

Hunston, S. and Sinclair, J. 1999. A Local Grammar of Evaluation. In *Evaluation in Text: Authorial Stance and the Construction of Discourse*, S. Hunston and G. Thompson (eds), 74–101. Oxford: Oxford University Press.

Hunston, S. and Su, H. 2017. Patterns, Constructions, and Local Grammar: A Case Study of 'Evaluation'. *Applied Linguistics* amx046, https://doi.org/10.1093/applin/amx046, 1–28.

Kaltenböck, G. 2003. On the Syntactic and Semantic Status of Anticipatory IT. *English Language and Linguistics* 7:2, 235–255.

Larsson, T. 2016a. The Introductory *it* Pattern: Variability Explored in Learner and Expert Writing. *Journal of English for Academic Purposes* 22, 64–79.

Larsson, T. 2016b. A Functional Classification of the Introductory *it* Pattern: Investigating Academic Writing by Non-native-speaker and Native-speaker Students. In thesis: *The Introductory* it *Pattern in Academic Writing by Non-native-speaker Students, Native-speaker Students and Published writers: A Corpus-based Study*, Uppsala University.

Leira, V. 1970. *Det*-setninger. *Mål og minne* 1-2. 48–71.

Leira, V. 1992. *Det*. *Norskrift* 75, 22–99.

Lemke, J.L. 1998. Resources for Attitudinal Meaning: Evaluative Orientations in Text Semantics. *Functions of Language* 5:1, 33–56.

Quirk, R, Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Søfteland, Å. 2014. *Utbrytingskonstruksjonen i norsk spontantale*. PhD dissertation, University of Oslo.

Wickens, P. 2001. *Computer-based Learning and Changing Legal Pedagogic Order of Discourse in UK Higher Education: A Comparative Critical Discourse Analysis of the TLTP Materials for Law*. PhD dissertation, University of Warwick.

Appendix

Table A. Distribution of all non-congruent correspondences according to attitudinal class.

	EO→NT		NO→ET		NT←EO		ET←NO	
A		0		17		21		1
			(modal +) if-clause to-inf. clause ADJ + *thing* modal main clause *the fact that …* BE ADJ ADJ + *that* *do good to*	5 2 2 2 2 1 1	*it* BE ADJ *for* x *to* modal main clause ADJ + *that* *to*-inf. NOUN mind *if*-clause ADJ + *thing* *it was just as well*	4 3 3 2 2 2 1 1 1	ADV	1
B		10		13		26		25
	(modal+adverb) *s*-passive *at* ADV modal existential Ø	4 3 1 1 1	ADV modal *it* (modal) BE ADJ *for* x *to* main clause	5 5 2 1	modal existential main clause ADV *to*-inf. clause *if*-clause NP NOUN	8 6 5 2 2 1 1 1	*s*-passive ADV main clause *om*-clause inf.-clause modal NOUN *bli*-passive Ø	8 5 3 2 2 2 1 1 1
C		3		4		8		6
	ADV Ø	2 1	ADV *it* BE ADJ *for* x *to* main clause ADJ + *that*	2 1 1 1	ADV Ø *it* BE ADJ *for* x *to* modal+*if* clause emphatic *do* main clause	2 2 1 1 1	ADV *det* VÆRE ADV *at* main clause	3 2 1
D		1		8		12		1
	ADJ x *to*-inf.	1	*it* BE ADJ *for* x *to* NOUN main clause *to*-inf. ADJ + *that* modal	2 2 1 1 1	main clause *it* BE ADJ *for* x *to* *it* SEEM ADJ *to* x *that* -*ing* clause *wh*-clause ADV	3 3 2 1 1	co-ordinated cl.	1

Signe Oksefjell Ebeling

				ADJ+ *thing*			
E	2		3		2		1
cleft	1	*it* BE ADJ *for* x	2	modal	2	*det* HOLDE *at* 'it HOLD that'	1
det VÆRE ADV *at*	1	*to*	1				
		Ø					
F	4		26		29		20
ADV	2	ADV	15	ADV	19	*det* PHRASAL VERB *at* 7	7
inf. clause + *at*	2	NOUN	6	modal	2	ADV 5	5
		wh-clause	1	*it* BE ADJ *to* x	2	inf.-clause 2	2
		main clause	1	*that*	1	main clause	1
		existential	1	*it* SEEM ADJ *to*	1	impersonal S +	1
		Ø	1	x *that*	2	VERB *that*	1
		other extraposed	1	Ø	1	initial *at*-clause	1
				passive	1	PREP + *at*	1
				SEEM *to have* NOUN		restructured from source Ø	
G	1						
NOUN	1						
Total		21		71		98		54

Author's address

Signe Oksefjell Ebeling
Department of Literature, Area Studies and European Languages
University of Oslo
PO box 1003
0315 Oslo
Norway

s.o.ebeling@ilos.uio.no

Sentence-initial indefinite subjects in English and Norwegian

Hilde Hasselgård

University of Oslo (Norway)

Abstract: The present study uses the English-Norwegian Parallel Corpus to investigate the frequency and use of indefinite noun phrases as subjects in English and Norwegian. Since subjects in both languages tend to appear clause-initially, indefinite subjects represent a deviation from the information principle. The clearest difference between the languages is the greater frequency of indefinite subject NPs in English. The lexicogrammatical features of the indefinite subjects and their immediate contexts are relatively similar in both languages. The indefinite subjects most commonly occur with intransitive verb phrases, and often in clauses with presentative or generic meaning. Translation correspondences of indefinite subjects show that the subject NP is retained in congruent form in the majority of cases, but more changes are made in translations from English into Norwegian than the other way round. This is taken to support the findings of the contrastive analysis and furthermore indicates that the light subject constraint is applied more strictly in Norwegian than in English.

Keywords: indefinite NP, subject, information structure, presentative, English, Norwegian

1.   Introduction

In languages such as English and Norwegian, subjects realized by indefinite noun phrases seem anomalous as regards information structure. Both languages tend to place the subject clause-initially (Hasselgård 2004) and are furthermore assumed to organize sentences in agreement with the information principle, that is, a distribution of given and new information in the clause that corresponds "to a gradual rise in information load" (Biber *et al*. 1999: 896). Thus, the beginning of the sentence – the subject position – is associated with given information. Since indefinite NPs are typically associated with new information, they should be disfavoured as subjects (Biber *et al*. 1999: 269), while definite noun phrase realization should be expected (Prince 1992). Yet, indefinite NP subjects are found in both languages, as shown by (1) and (2) from the English-Norwegian Parallel Corpus (ENPC).[1]

(1)   *En tyv* er ikke voldsom, men stillferdig. (KF1)
        A thief is not violent but quiet. (KF1T)

---

[1] Examples from the ENPC are given with the original first. Norwegian examples are followed by a literal translation if their wording diverges from that of their English translations/sources found in the corpus.

(2)   *A rat* crept out of the hole behind the dresser ... (GN1)
      En rotte kom ut av hullet bak kommoden ... (GN1T)

Previous contrastive studies of English and Norwegian have noted cross-linguistic differences in the use of indefinite subjects, particularly that Norwegian is less tolerant of them than English (Ebeling 2000: 191, Hasselgård 2004: 201, Johansson 2004: 41). However, these studies had other primary concerns (see Section 3). The present study zooms in on sentence-initial indefinite subjects in both languages. It first compares original texts in English and Norwegian to answer the following research questions:

- How frequent are indefinite subjects in English and Norwegian?

- What are the lexicogrammatical features of such subjects (e.g. +/- modification) and their verbs (e.g. +/- transitive)?

- What are the contexts for indefinite subjects – and are they the same in both languages?

The features of the subject NPs include their semantics, particularly referent type and specificity. Contexts include features of the verb phrase as well as other clause elements, particularly objects/complements and locative adverbials. For example, both English and Norwegian have indefinite subjects in contexts of generic reference, as in (1), and in clauses denoting the existence or appearance of a specific subject referent, as in (2).

The second part of the study probes further into cross-linguistic differences in the use of indefinite subject NPs by exploring their translations, the frequency with which changes are made, and the nature of such changes. Given that indefinite subjects are comparatively rare in both languages, the translation principle of normalization (e.g. Baker 1996: 183) might prompt translators to make changes to either the form of the subject (as in (3)) or the placement of the indefinite NP, as in (4).

(3)   *Cultured pearls* are in the vault. (DF1)
      Kunstperlene ligger i velvet. (DF1T)
      "The cultured pearls lie in the vault."

(4)   *En gammel kvinne* tok imot oss, vennlig, men uten smil. (JW1)
      "An old woman received us…"
      We were received by *an old lady*, in a friendly but unsmiling fashion. (JW1T)

Since indefinite subjects are assumed to be less frequent in Norwegian than in English, translations into Norwegian are expected to involve a change of the subject NP more often than translations into English, especially if the NP has specific reference, as is the case in (3) and (4). Information structure and semantics, especially the notions of existence/appearance are also expected to play a role. The overall aim is to learn more about information structuring in both languages, in particular the conditions in which apparent breaches of the given-to-new principle are tolerated, and whether (and how) the languages differ in this respect.

2. Indefinite noun phrases in English and Norwegian

English and Norwegian indefinite NPs are rather similar in form: there are indefinite articles for the singular only (English *a/an*, Norwegian 'bokmål' *en/ei/et* and Norwegian 'nynorsk' *ein/ei/eit*). Indefinite plurals and uncountable nouns occur with no article or with certain types of quantifying determiners (e.g. cardinal numbers, *some*; see Lyons 1999: 33 ff). The English indefinite articles vary according to the phonological context and the Norwegian ones according to grammatical gender (masculine, feminine, neuter), but they have the same functional properties and relatively similar conditions of use (for some exceptions, see Hasselgård *et al.* 2012: 120). It may be noted that definiteness of nouns is generally marked by suffixes rather than articles in Norwegian. The forms are shown in (5), in which the inflectional suffixes have been underlined.

   (5)   en hund – hund<u>en</u> – hund<u>er</u> – hund<u>ene</u>
         "a dog – the dog – dogs – the dogs"

While the indefinite article clearly marks an NP as indefinite in both languages, NPs occurring without a determiner or with determiners other than the articles may be harder to classify. Lyons (1999: 33) argues that e.g. cardinal numbers "do not encode [-Def]", and concludes that a noun phrase may be classified as indefinite simply through "the absence of any definite determiner" (*ibid.*). In cases of doubt, Abbott (2006) suggests locative existentials as a test for definiteness. That is, if it is possible to insert an NP in the formula "There is NP", with no special interpretation of the existential, then the NP is indefinite.[2] In my classification of definite and indefinite noun phrases I have mainly followed Lyons, with the exception of NPs with the determiner *all* or the corresponding Norwegian *alle*. According to Lyons (1999: 44) *all*, being a universal quantifier, is a definite marker when used as a determiner. However, in both English and Norwegian *all(e)* may or may not co-occur with the definite article/suffix, and so I have chosen to classify *all/alle* + indefinite plural as indefinite. NPs with *each/every,* on the other hand, have not been included, as they are generally associated with inclusiveness and hence definiteness (Lyons 1999: 32) and do not co-occur with (other) definiteness markers. Other notable exclusions are NPs introduced by *one of / en/ei/et av* + def. NP, which are 'containing inferables' (Prince 1981) and thus regarded as definite, and the indefinite pronouns *one/en* (as in *One might get £300,000… / En kunne få £300 000 …* (FF1)), which is homonymous with the numeral in both languages. An overview of the types of indefinite NPs included in the material is given in Table 1.

Table 1. Types of indefinite noun phrases in English and Norwegian.

Singular article	*a woman*	*en kvinne*
Singular, no determiner	*water*	*vann*
Singular quantifier	*one/no man*	*en/ingen mann*
Plural, no determiner	*families*	*familier*
Plural quantifier	*several witnesses* *a couple of books* *two republicans*	*flere vitner* *et par bøker* *to republikanere*
Comparative determiner	*other youngsters*	*andre ungdommer*

---

[2] Two of her examples are *There is a book in the shop window* and *There is the book in the shop window*, of which the latter requires the special interpretation of 'list existential' to be grammatical (Abbot 2006).

An indefinite noun phrase can have specific, non-specific or generic reference in both English and Norwegian, as explained by Biber *et al.* (1999: 260) and exemplified in (6)–(8):

(6)   SPECIFIC: *Mosquitoes and fireflies* had come in. (BO1)
       Moskitoer og ildfluer var kommet inn. (BO1T)

(7)   NON-SPECIFIC: *En drosjesjåfør* ville kjørt helt bort til Hansdals inngangsdør… (EG1)
       "A taxi driver would (have) driven right over to Hansdal's front door…"
       A taxi driver would have driven the old lady right to the door. (EG1T)

(8)   GENERIC: *A child* needs security. (ABR1)
       Et barn trenger trygghet. (ABR1T)

According to Hawkins,

> The speaker performs the following speech acts when using an indefinite article to achieve (specific) indefinite reference: He (a) introduces a referent (or referents) to the hearer; and (b) refers to a proper subset, i.e. not-all, of the potential referents of the referring expression. (1978: 186–187)

While non-specific reference involves reference to "any member of a total class" (*ibid.*: 215), a generic referent is considered "a typical representative of its class" (*ibid.*). These principles are assumed to apply to both English and Norwegian.

3.   Literature review

As noted above, the use of an indefinite subject in sentence-initial position typically involves a deviation from normal or unmarked information structure. Quirk *et al.*, for example, observe that "the organization of sentences … generally presumes that a sentence begins with a reference to 'given' information and proceeds to 'new' information" (1985: 1402). Although there is no syntactic rule against using an indefinite subject as NP, Quirk *et al.* argue that "… a certain awkwardness is sensed where the recipient is expected to interpret a theme as entirely new and unconnected with anything previously introduced" (*ibid.*). Quirk *et al.*'s views on word order and information structure are clearly influenced by the Prague School and its concept of Functional Sentence Perspective (FSP), in which information structure, or the distribution of communicative dynamism (CD), is seen as a crucial structuring principle for word order. According to Firbas (1992: 118), "the principle of FSP arranges the sentence elements in a Th[eme] – Tr[ansition] – Rh[eme] sequence. If asserting itself to the full extent, it … induces the sentence to display what has been termed 'the basic distribution of CD'", which implies a gradual rise (*ibid.*: 10).

Firbas (1992) devotes a section to *context-independent* subjects, i.e. subjects that do not link up with the preceding text or the situational context, and that are typically realized by indefinite noun phrases. Such subjects are likely to occur with (mostly intransitive) verbs that denote "appearance or existence on the scene" (1992: 60). In these sentences, the subject carries the highest degree of communicative dynamism (CD). However, as Dušková (1999: 249) argues, placing new information at the beginning of a sentence runs counter to the basic distribution of CD. Thus context-independent, rhematic subjects in subject position constitute a "deviation from basic distribution of CD" (*ibid.*: 254), and moreover, may represent an unresolved conflict between the basic distribution of CD and the grammatical word order principle (*ibid.*: 260). In a later paper, Dušková (2015) shows that sentence-initial rhematic

subjects in English, though rare, tend to occur in the pattern S-V-A and thus belong to the presentation scale (2015: 21–23).

Similar views on sentence openings and information structure are expressed by other functionally oriented linguists. For example Chafe (1994: 83–92) argues that subjects are governed by the *light subject constraint*: i.e., in conversation subjects normally refer to given or accessible information, and the new information (if any) comes later in the clause.[3] Although Halliday (1994) makes a distinction between theme (what comes first) and given information (what the speaker expects the hearer already knows), he concedes that the theme is typically selected from given information (1994: 299). A clause that starts with new information will typically have a marked information focus, provided it also contains some recoverable information (*ibid*.: 301).

Biber *et al*. (1999) investigate the distribution of definite and indefinite noun phrases (as marked by articles) across syntactic functions. They find that "indefinite subjects are much less common [than definite ones] but by no means rare" (1999: 269). This applies in all four registers studied (fiction, news, academic prose and conversation). The object function is much more characteristic of indefinite NPs, and to some extent prepositional complement (*ibid*.). This distribution is explained by the information principle: "When new information is introduced in subject position, it is marked as thematically important" (*ibid*.). Similarly, in a study of (English) subjects and information status, Prince (1992: 316) reports that only 10% of indefinite NPs in a single text had subject function as opposed to 38% of definite NPs. She suggests that definiteness is "a grammaticization of Hearer-status" (*ibid*.: 317), i.e. a marker that an entity is assumed to be identifiable by the hearer/reader. De Hoop & Krämer (2006) offer an alternative, processing-based explanation for the scarcity of indefinite subjects, arguing that they involve a "conflict of interpretation" because subjects in general (in the standard subject position) "favour a referential reading" while indefinites "favour a nonreferential reading" (2006: 119).

The Norwegian reference grammar (Faarlund *et al*. 1997) states that the subject usually expresses given information in Norwegian too: it does not normally convey new information (1997: 691). There are, however, contexts that make a "new" subject more likely, for example if it refers to an element within a recoverable group (i.e. what Prince 1981 refers to as "containing inferable") or if it is singled out for contrastive focus (Faarlund *et al*. 1997: 692). Furthermore, both generic reference and the presence of certain (plural) quantifiers also make an indefinite subject acceptable (*ibid*.: 690). In a paper on the typology of subjects, Faarlund questions the acceptability of the Norwegian sentence *Ein mann arbeider på vegen* ('a man works on the road') because the subject, though meeting the subjecthood requirement of agency, does not convey given information (1988: 199). A presentative with *det* ('there') is suggested as a fully acceptable alternative, namely *Det arbeider ein mann på vegen* ('there works a man on the road') (*ibid*: 200).

Indefinite subjects have not, to my knowledge, been the main topic of any cross-linguistic study of English and Norwegian, but they have come up in previous studies of presentative constructions (Ebeling 2000), thematic choice (Hasselgård 2004, 2005) and the extent to which syntactic subjects are preserved in translation (Johansson 2004).

A construction with an indefinite subject, an intransitive verb, and most typically a locative adverbial may have a presentative function even in the absence of the explicit presentative marker *there/det*, hence the term "bare presentative" (Ebeling 2000:157 ff). Ebeling discusses both subject-initial bare presentatives (as in example (3) above) and those

---

[3] The light subject constraint (Chafe 1994: 85) is also referred to as the "light starting point constraint". Chafe (1986) acknowledges that writing may not adhere as strictly to it as conversation does.

with locative inversion (*ibid*.: 160, 173).[4] An important finding is that English bare presentatives have a greater variety of lexical verbs than full presentatives with *there* (*ibid*.: 161). The difference in verb choice between the two construction types is somewhat smaller in Norwegian, whose full *det*-presentative construction is more flexible than the English one as regards its verbal collocates (*ibid*.: 176). Ebeling finds fewer examples of subject-initial bare presentatives in Norwegian than in English (*ibid*.: 179). He therefore suggests that "Norwegian is more susceptible to FSP than English" (*ibid*.). Locative-initial bare presentatives, on the other hand, are more frequent in Norwegian (*ibid*.: 186). These findings reveal "a greater tendency in Norwegian to avoid indefinite NPs in subject position than in English" (*ibid*.: 187).

Johansson (2004) studies the change or preservation of the subject in translation from English into Norwegian. He finds that the vast majority of subjects (around 90%) are preserved in translation (2004: 33, 49) probably because "the subject is essential in the building of both sentences and texts" (*ibid*.: 32). However, "changes may be induced by formal differences between the languages, or they may be due to differences in stylistic norms" (*ibid*.). Johansson notes that English indefinite subjects may be rendered by a construction with the anticipatory subject *det* ('it/there') in the Norwegian translation. This is related to "a lower tolerance for indefinite subjects in Norwegian than in English" (*ibid*.: 41).

English and Norwegian clause themes have been studied contrastively by Hasselgård (e.g. 2004, 2005) and Rørvik (2013).[5] (See also Rørvik and Monsen, this volume.) Although Norwegian is a V2 language while English is an SV language, there are great cross-linguistic similarities. However, an important difference is that Norwegian has more fronted non-subjects, particularly adverbials, as well as more anticipatory subjects realized by *det* ('it/there') (Hasselgård 2004: 192). The latter finding was "associated with a greater tendency in Norwegian than in English to avoid indefinite subject NPs and NPs conveying new information in thematic position" (*ibid*.: 208). It was found that translated texts in both languages show a great degree of similarity with their source texts, indicating that the differences between English and Norwegian originals pertain more to preferences than to syntactic constraints (*ibid*.: 190). The study indicated that "word order is freer Norwegian than in English, [but] *information structure* seems to be more flexible in English. I.e. English is more tolerant of new information in the Theme, while Norwegian has a stronger preference for 'light' Themes" (*ibid*.: 208). In a follow-up study it was argued that Norwegian has a greater preference than English for syntactically and informationally light themes (Hasselgård 2005: 46) due to differences in the functional load of themes between the languages: Norwegian themes are less often associated with contrast and more often with cohesion, and may therefore be less prominent (*ibid*.).

4. Material and method

The material for this study consists of declarative main clauses with indefinite subjects from the fiction part of the English-Norwegian Parallel Corpus (ENPC), i.e. 30 text extracts of 10,000–15,000 words in each language plus their translations (Johansson *et al*. 1999/2002, Section 1.2).[6] As the corpus is not parsed, subjects were retrieved (from the original texts) using a combination of lexical searches, positional criteria and PoS tags. Indefinite singular

---

[4] An example of locative inversion is *Behind the policemen was a middle-aged woman …* (DF1) > *Bak politimennene var en middelaldrende dame …* (Ebeling 2000: 159).

[5] In these studies, theme is defined as the first clause element with an experiential/referential function plus any preceding elements (Halliday and Matthiessen 2014: 105).

[6] The corpus was accessed from the PerlTCE interface (http://www.tekstlab.uio.no/cgi-bin/omc/PerlTCE.cgi).

subjects were retrieved by searching for indefinite articles and the numerals *one/én/ett* in first and second position in an s-unit.[7] Indefinite plurals and uncountable nouns were retrieved by searching for nouns in first position and in second position without a preceding definite determiner. Most of the resulting concordances had fairly (or very) poor precision, but were cleaned up manually to remove irrelevant hits, where the initial NP was either not indefinite or not a subject. The resulting concordances with their translations were transferred to a database and annotated for lexicogrammatical features and type of translation correspondence, as specified below.

The search method has (at least) two shortcomings. Most importantly, it misses indefinite subjects that are not sentence-initial. These include notional subjects in *there/det* presentatives, subjects occurring after initial adverbials of more than one word (most noticeable for English), subjects in post-verbal position (most noticeable for Norwegian, which is a V2 language) and subjects in conjoined clauses. There is no obvious way in which these subjects could have been retrieved reliably except manually. Since non-initial indefinite subjects are not quite so blatantly in breach of the information principle, this shortcoming, though not trivial, was deemed acceptable for the purposes of this study. The second shortcoming is that the PoS tagging, particularly for Norwegian, is not 100% reliable. Sifting through the concordance lines, I found examples of words wrongly tagged as nouns, which makes it likely that some nouns are wrongly tagged as something else. Short of reading through the whole corpus, there is little to be done about this. Note that Ebeling (2000: 158) used slightly different, and more comprehensive, search procedures to identify instances of bare presentatives, but still acknowledged less than full recall.

5.   Contrastive analysis

This section compares indefinite subjects in English and Norwegian original texts. In addition to frequency information, consideration is given to syntactic and semantic features of the subject NP, the verb phrase and the clause pattern in which the subject appears in order to investigate both the properties and the contexts of indefinite subjects.

5.1   Frequency and form of indefinite subjects in original texts

As expected, indefinite subjects are more frequent in English than in Norwegian original texts, as shown in Table 2. The raw frequencies are comparable since the amount of text in the two languages is similar. The last row of the table gives the frequency of indefinite subjects per 1000 s-units. This was considered the best available measure considering the opportunity of occurrence for a sentence-initial subject, even though not all s-units are complete sentences with a grammatical clause structure. Like the raw frequencies, the relative frequency of indefinite subjects per 1000 s-units is greater in English.[8]

---

[7] An s-unit is roughly equivalent to an orthographic sentence (Johansson *et al*. 1999/2002, section 2.3.4), i.e. a stretch of language starting with a capital letter and ending with a final punctuation mark.
[8] The (raw) frequency difference relative to the number of s-units is significant (LL=49.33, p<0.0001).

Table 2. Frequency of indefinite subjects in English and Norwegian (original texts).

Subject form	English	Norwegian
Sg N, indefinite article	273	268
Sg N, no determiner	112	55
Sg N, quantifier	34	13
Pl N, no determiner	267	161
Pl N, quantifier	38	49
Comparative determiner	8	2
TOTAL	**732**	**548**
Freq. per 1000 s-units	24.9	16.8

Singular noun phrases with indefinite articles are the most common realization of indefinite subject in both languages. Both singular and plural nouns with no determiner are much more frequent in English than in Norwegian indefinite subjects, and thus account for most of the frequency difference between the languages. Indefinite subjects are used in all the original corpus texts except one in Norwegian, with both the highest and the lowest number per text being higher in English (49 vs. 38 examples and 9 vs. 6). Thus, the material supports the hypothesis, derived from previous contrastive studies, that indefinite subjects are (even) less common in Norwegian than in English.

5.2 Lexicogrammatical features of the subject NP

In the analysis of indefinite subject noun phrases, the following features were noted:

- NP complexity, i.e. the absence/presence of premodifier(s) and postmodifier(s).

- NP semantics, i.e. whether the referent of the NP is human, (non-human) animate, concrete, or abstract.

- NP specificity, i.e. whether the reference of the NP is specific, non-specific, or generic.

An overview of noun phrase complexity is shown in Table 3. Simple (unmodified) noun phrases are most common, with a slightly higher proportion in English than in Norwegian. Premodification and postmodification are equally common in English, but in Norwegian premodified NPs are almost twice as common as postmodified ones.

Table 3. Complexity of NPs functioning as indefinite subjects.

	English		Norwegian	
	N	%	N	%
Simple NP	416	56.8	280	51.1
Premodifier + head	134	18.3	150	27.4
Head + postmodifier	137	18.7	82	15.0
Premodifier + Head + postmodifier	45	6.1	36	6.6
	732	99.9	548	100.1

Table 4 shows categories of referents of the indefinite subject NPs. Concrete (non-animate) referents are most common in both languages, followed by human referents. Abstract and (non-human) animate referents are comparatively rare. The proportions of each type of referent are fairly similar between the languages.

Table 4. Referents of indefinite subject NPs.

Subject NP referents	English		Norwegian	
	N	%	N	%
Human	238	32.5	212	38.7
Animate	51	7.0	26	4.7
Concrete	356	48.6	248	45.3
Abstract	87	11.9	62	11.3
	732	100	548	100

The analysis of noun phrase specificity also gave a similar distribution between the languages, as Table 5 shows. Specific reference is most common, followed by generic and non-specific. Example (9) thus shows a prototypical representative of an indefinite subject: a singular, unmodified noun with an indefinite article, a concrete referent and specific reference.

(9)   *En stemme* kom fra treet: (THA1)
       A voice came from the tree: (THA1T)

Table 5. Specificity of indefinite subject noun phrases.

NP specificity	English		Norwegian	
	N	%	N	%
Specific	418	57.1	317	57.8
Non-specific	137	18.7	102	18.6
Generic	177	24.2	129	23.5
	732	100	548	99.9

The proportion of generic reference in indefinite subjects seems high compared to Biber *et al*.'s (1999: 266) report that less than 2.5% of definite NPs in fiction had generic reference. (Figures for generic indefinite NPs are not given separately.) This indicates that indefinite subjects may be particularly prone to generic interpretation. Note that generic reference is possible with both singular and plural nouns in both languages, as shown in (10) and (11); see also Biber *et al*. (1999: 265).

(10)   *Scholars and artists* have no morals whatever about grants of money. (RDA1)
        Forskere og kunstnere har ingen som helst moralske skrupler overfor stipendier. (RDA1T)

(11)   Men *en soldat* så da ikke slik ut. (KAL1)
        But then *a soldier* did not look like that. (KAL1T)

With respect to the distinction between given and new information, Chafe (1994) argues that generic reference has a special status:

>   …sharing knowledge of generic referents is different from sharing knowledge of particular referents. Knowing a category, like the category that allows something to be called an elephant,

entails knowing something about a *typical* instance of that category, whereas the sharedness involved in identifiability depends on knowing a *particular* instance. (Chafe 1994: 103)

An indefinite NP with generic reference may thus not be considered to convey new information to the same extent as one that introduces a new specific referent. If this is the case, a generic sentence-initial subject will seem less objectionable as it does not so clearly violate the information principle.

5.3 Lexicogrammatical features of the verb phrase

The following features were noted in the analysis of verb phrases occurring with sentence-initial indefinite subjects:

-   **Voice**: active/passive[9]

-   **Transitivity**: intransitive, monotransitive, ditransitive, complex transitive, copular

-   **Semantics** (cf. Halliday 1994): material, mental, verbal, existence/ appearance, attributive, identifying, possessive

The classification of verb semantics is based on Halliday's process types (1994: 173), but modified to include a category of verbs of existence and appearance, mostly inspired by the presence of this category in FSP analyses of the presentation scale (Firbas 1992: 59, 67). The category overlaps with Halliday's existential, relational and material processes, thus reducing or replacing these categories in comparison with a purely systemic-functional analysis. Further, behavioural processes have been combined with material (as in Matthiessen 1995).

The distribution of active and passive voice is similar in both languages, with roughly 8.5% of clauses with indefinite subjects in the passive. This is slightly above the ratio of passives in fiction found by Biber *et al*. (1999: 476), who do not report exact figures, but present a diagram where the percentage of passives seems to be around five.

The two languages also have similar distributions of transitivity types. As shown in Table 6, approximately half of the clauses are intransitive in both languages, with mono-transitive verbs accounting for just over a quarter. Copular patterns take up about a fifth, while ditranstive and complex transitive verbs are infrequent in both languages. Note that transitivity type has been assigned from the verb in context, not from the lexeme, so that any clause not containing a grammatical object or predicative (including passives) have been classified as intransitive.

Table 6. Verb transitivity in clauses with indefinite subjects (original texts).

Verb transitivity (in context)	English		Norwegian	
	N	%	N	%
Intransitive	361	49.3	281	51.3
Monotransitive	193	26.4	154	28.1
Copular	156	21.3	98	17.9
Ditransitive	12	3.0	9	2.7
Complex transitive	10		6	
Total	732	100	548	100

---

[9] All verb phrases not marked for passive voice have been classified as active.

Intransitives are more common than expected from the general distribution of transitivity types reported by Oostdijk & de Haan (1993: 48). In their study of the distribution of clause patterns in main clauses, intransitives are only slightly more frequent than monotransitive and copular patterns, at 30%, 25% and 24%, respectively. It is thus a fair assumption that the intransitive clause pattern is a favourable context for indefinite subjects.

As regards the verb semantics, the most frequent process type is material, followed by attributive for English and existence/appearance for Norwegian (see Table 7).[10] The results are difficult to compare to other studies because of the modifications of the classification system. However, the proportional distribution of process types in clauses with indefinite subjects appears to be fairly similar to the general distribution of process types reported in Matthiessen (1999).

Table 7. Semantic types of verbs in clause with indefinite subjects (original texts)

Verb semantics	English		Norwegian	
	N	**%**	**N**	**%**
Material	378	51.6	261	47.6
Attributive	131	17.9	78	14.2
Ex/app	103	14.1	103	18.8
Mental	43	5.9	25	4.6
Verbal	30	4.1	40	7.3
Identifying	25	3.4	22	4.0
Possessive	22	3.0	19	3.5
Total	732	100	548	100

Based on the most frequent verb phrase properties, a prototypical sentence with an indefinite subject would thus be as in (12) in both English and Norwegian: it has active voice, an intransitive verb and refers to a material process.

(12) Et kaldt solgløtt *gnistret* i rutene på Deichmanske Bibliotek. (BV1)
      A bleak ray of sunshine *sparkled* in the windows of the Deichman Library. (BV1T)

5.4 Lexicogrammatical features of the clause

Besides the subject and the verb, I considered the presence of other constituents in clauses with indefinite subjects. In clauses with a direct object or predicative (combined here under the label 'complement', as in Halliday 1994: 80) the realization of the post-verbal element was analysed, for example to determine the extent to which the clause conveys given/identifiable information in a later position than the indefinite subject. As shown in Table 8, definite NPs (including pronouns) are the most common realization of complements in English, but not in Norwegian, where indefinite complements are preferred. Especially complements realized by indefinite NPs and adjectives are differently distributed across the languages. The more frequent adjective realization in English can be related to the higher frequency of attributive processes shown in Table 7.

---

[10] Material verbs denote processes of doing and causing (Halliday 1994: 109); attributive processes ascribe an attribute to the subject referent and are a subtype of relational processes (*ibid*.: 173).

Table 8. Form of complement in transitive and copular clause patterns

Complement	English			Norwegian		
	N	%		N	%	
Indef NP	107	28.2		96	35.4	
Def NP	79	20.8	124	52	19.2	82
Pronoun	45	11.8	32.6%	30	11.1	30.3%
Adjective	88	23.2		41	15.1	
Clause	41	10.8		41	15.1	
PP + adverb	20	5.3		11	4.1	
Total	380	100.1		271	100	

The higher frequency of indefinite complements in Norwegian is an interesting finding. But even for English, it appears that the proportion of indefinite NP complements in the present material is higher than in Biber *et al.*'s figures for fiction (1999: 269), which only includes object NPs with definite and indefinite articles. In terms of Functional Sentence Perspective, with communicative dynamism being a relative concept, a clause with indefinite clause complementation after an indefinite subject is less at odds with the basic distribution of CD than one with an indefinite subject and a definite complement (Firbas 1992: 8). It appears that an indefinite complement makes the indefinite subject more palatable since the information structure of the clause will appear as all new, as in (13). By contrast, the information in (14) proceeds from new to given: the door is inferable from previous mention of a building, while the brunette is mentioned for the first time.

(13) Bønder satte opp *uthus og innhegninger for hester og kveg* og lot en plog lage furer i jorden. (SH1)
Farmers put up *barns and corrals for horses and cattle* and ploughed long furrows in the earth. (SH1T)

(14) A fresh-faced brunette woman in her thirties, wearing a flowery apron, opened *the back door*. (MM1)
En dame i trettiårene med brunt hår og et spill levende ansikt åpnet *døren*. (MM1T)
"A woman in the thirties with brown hair and a most lively face opened the door"

The pattern of indefinite subject + indefinite complement also occurs in descriptive passages presenting a series of observations, as in (15), which occurs in a series of independent observations. In contrast to the brunette introduced in (14), the freemasons in (15) are not maintained as a topic of the ensuing discourse.

(15) A party of freemasons scrutinised a globe. (BC1)
En gruppe frimurere studerte en globus. (BC1T)

Finally, the pattern of new information in both subject and complement position includes a great number of generic sentences such as the one in (16). As noted above, generic noun phrases have a special status in information structure since they do not introduce actual discourse referents (Chafe 1994: 103). As example (16) illustrates, such sentences can be definition-like. This particular example also demonstrates the almost non-referential nature of generic reference, since the word *menasjeri/menagerie* also occurs in the previous sentence of the text, where it is used by an adult, while the current sentence explains the word for the benefit of a child.

(16) Et menasjeri var en samling av forskjellige dyr, ... (JG1)
      A menagerie was a collection of animals. (JG1T)

The second feature that was noted was the absence/presence of place adverbials. This adverbial type was singled out due to the close association between indefinite subjects and bare presentatives (Ebeling 2000) and between location and the presentation scale (Firbas 1992; Dušková 2015). Figure 1 shows the percentage of intransitive, monotransitive and copular clause patterns that also include a place adverbial. There were no place adverbials in ditransitive clauses and only one in a Norwegian complex transitive clause.
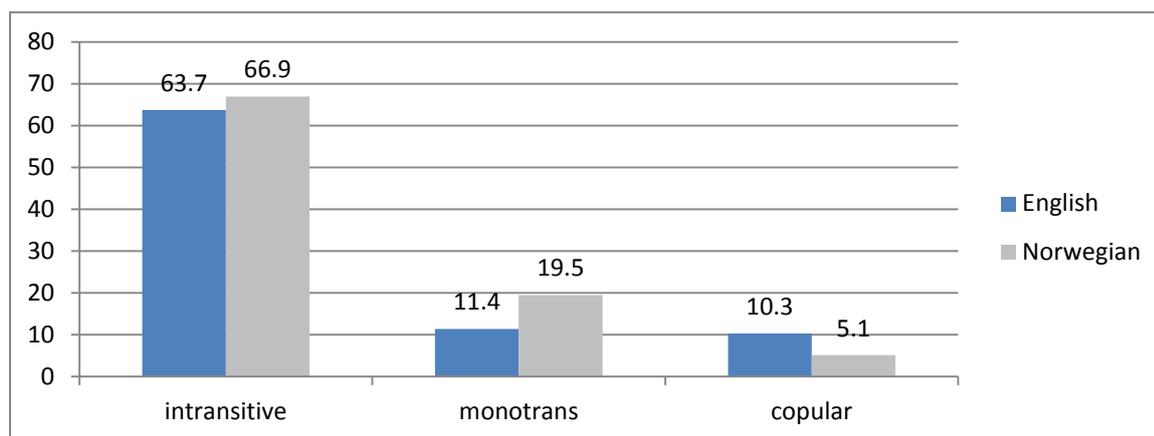

Figure 1. Place adverbials in clauses with indefinite subject (percentages)

Intransitive clauses are not only the most frequent environment for indefinite subjects but also clearly those most likely to occur with a place adverbial. However, it appears that intransitive clauses are on the whole most likely to occur with (all kinds of) adverbials: according to Oostdijk & de Haan (1993: 59) about 75% of intransitive non-embedded clauses contain one or more adverbials, as against about 50% of intensive clauses and 58% of monotransitive clauses. Considering that not all adverbials are place adverbials, the percentage of intransitive clauses containing a place adverbial shown in Figure 1 is still strikingly high.[11]

The co-occurrence of indefinite subject and place adverbial is frequent enough to constitute a pattern that tends to have some kind of presentative function, irrespective of process type. Such bare presentatives (Ebeling 2000: 157) typically involve intransitive verbs and subjects with specific reference. The pragmatic function is similar to that of full presentatives with *there/det*. Ebeling found bare presentatives to be more frequent in Norwegian than in English, a finding which is corroborated by the present study.[12] It may be noted that bare presentatives, as discussed by Ebeling (2000), may have an initial adjunct and a clause-final subject. This word order pattern has not been included in the present study; see Section 4. A typical example of the construction is given in (17), which contains a verb of existence/appearance.

---

[11] According to Hasselgård (2010: 34) and an unpublished study of Norwegian by the same author, close to 40% of adjunct adverbials are spatial in both languages.
[12] In Norwegian, bare presentatives account for 29.6% of the sentences with indefinite subjects (162 out of 548), and in English for 22% (161 out of 732).

(17) *En skikkelse* kommer til syne rundt hjørnet,… (LSC1)
      A figure comes into sight around the corner,…

In FSP (see e.g. Firbas 1992, Dušková 1999), sentences such as (17) represent the presentation scale. The subject has the dynamic role of Phenomenon with an Ex/App verb (and optionally an adverbial). Most such sentences could have been reformulated as full presentatives with *det/there* as an anticipatory subject, but as Ebeling (2000: 141) shows, English *there*-clauses are more restrictive than Norwegian *det*-clauses as regards the type of verb that occurs in them. The Norwegian original of (18) seems to have some sort of presentative function in spite of the fact that it does not contain a verb of existence/appearance. This type of passive construction is also noted by Dušková (1999: 255), who classifies it as presentational because the subject is taken to be rhematic. Example (19) is also presentational: although the verb denotes a material process, a full presentative would have been (marginally) acceptable in Norwegian (*Det plystret en fugl i hagen*), but not in English (**There chirruped a bird in the garden*). This suggests that the constructional meaning of indefinite subject + V + locative adverbial is more clearly presentative − and more readily paraphraseable by a full presentative − in Norwegian than in English.

(18) *En flokk griser* jages mot meg av folk med kjepper og stokker, … (SL1)
      "A herd (of) pigs is chased towards me by people with sticks and canes"
      A herd of pigs rushes towards me driven by men with whips and sticks, …

(19) *A bird* chirruped in the garden. (MM1)
      En fugl plystret i hagen. (MM1T)

5.5 Summary of the contrastive analysis

The contrastive analysis of indefinite subject NPs in comparable original data has shown that the clearest cross-linguistic difference concerns frequency. While indefinite subjects are more frequent in English, their lexicogrammatical features are relatively similar in both languages. However, the zero article is more common in English than in Norwegian, and premodified NPs are proportionally more common in Norwegian. The lexicogrammatical features of the verb phrases in clauses with indefinite subjects are also similar. Most VPs are intransitive with a material process verb in the active voice. However, verbs of existence/appearance are more common in Norwegian while attributive verbs are more common in English. The lexicogrammatical features of the clause pattern differ somewhat between the languages. Place adverbials are frequent in both, but NP complements are more likely to be indefinite in Norwegian than in English. It was found that generic reference and presentative function provide favourable conditions for indefinite subject NPs in both languages.

6.   The translation of indefinite subject NPs

This section discusses the translation of sentences with indefinite subjects from English into Norwegian and vice versa. Because indefinite subjects are comparatively rare in both languages, it is hypothesized that translators will make a number of changes to avoid the markedness of a rare construction and atypical information structure. And because indefinite subjects are less frequent in Norwegian than in English, Norwegian translators are expected to make more changes than English translators to avoid indefinite subject NPs in clause-initial position, due to the translation principle of normalization (Baker 1996).

6.1 The classification of translation correspondences

The starting point of the classification scheme is that of Johansson (2007: 23) into *congruent, divergent* and *zero* correspondences. Congruent correspondences are those that have the same types of linguistic form in both languages, and divergent correspondences are those where the realization types differ. With zero correspondence, either the source or the translation lacks the item under investigation. In the current material, zero correspondences were only found when whole s-units were omitted in the translation.

A more fine-grained classification system than Johansson's (2007) was needed for the present analysis, and so the categories of congruent and divergent were subdivided. Table 9 shows the correspondence types according to degrees of congruence. Congruent correspondences are those in which the subject NPs of the original and the translation correspond practically word for word with each other. Semi-congruent correspondences have the same phrase type, but with different internal structures. Divergent correspondences may involve correspondences between an indefinite NP and a definite one or a pronoun, or more substantial reorganizations of the clause, including the use of a full presentative (with the anticipatory subject *there/det*), a rearrangement of the clause content that involves a change of subject NP, and even more radical changes to clause/sentence structure, subsumed under the label of "rephrasing".

Table 9. Congruent and divergent correspondences.

		Original	**Translation**	**Gloss of Norwegian example**
Congruent: indefinite NP		*En tyv* er ikke voldsom, men stillferdig. (KF1)	*A thief* is not violent but quiet.	A thief is…
Semi-congruent: restructured indefinite NP		*Faces showing shock* looked out at our passing. (DF1)	*Tydelig sjokkerte ansikter* kikket ut på oss da vi passerte.	Clearly shocked faces…
Divergent correspondences	Definite NP	*Animals* don't eat me, and I don't eat them. (PDJ3)	*Dyrene* spiser ikke meg, og jeg spiser ikke dem.	The animals eat not me…
	Pronoun	Men *et slikt bytte* er like umulig: (JG1)	But *this* is equally unacceptable.	But a such reward is equally impossible
	Full presentative	*Puddles* had formed everywhere, …(SG1)	*Det var* sølepytter overalt, …	There were puddles everywhere
	Subject change	*En så viktig beslutning* må ikke avledes. (FC1)	*One* mustn't be side-tracked from *an important decision like that*.	A so important decision must not be sidetracked
	Rephrasing	*A raised voice* is remarked on, … (RR1)	Hvis noen hever stemmen, blir det bemerket, …	If anyone raises the voice, is it remarked

6.2 Changes made in translation

Table 10 shows the extent to which translations are congruent or divergent, and the types of changes made in case of divergence. The two directions of translation differ markedly from each other: fewer changes are made in translations from Norwegian into English than in the other translation direction. That is, congruent and semi-congruent correspondences are much more common in Eng→Nor than in Nor→Eng.

Table 10. Translation correspondences of indefinite subject NPs.

		Norwegian→English		English→ Norwegian	
		N	%	N	%
Congruent		407	76.4	402	55.3
Semi-congruent		55	10.3	73	10.0
Divergent	Definite NP	17	3.2	104	14.3
	Pronoun	8	1.5	14	1.9
	Full presentative	2	0.4	33	4.5
	Subject change	31	5.8	72	9.9
	Rephrased	13	2.4	29	4.0
	Total	533	100.1	727	100

The divergence that involves turning an indefinite subject into a definite one is clearly more frequent in Eng→Nor. The example of this type of change in Table 9 has generic reference. It appears that Norwegian definite plurals can have generic reference more readily than English ones, according to Johansson and Lysvåg (1987: 43), who continue: "If this tendency is carried over into English the result would be a noun phrase with more or less clear specific reference" (*ibid.*). However, the change from indefinite to definite form is also found when the subject has specific reference, as shown in (20).

(20) *Walls* had been pulled down to make this a room that accommodated nearly all the ground floor. (DL1)
Veggene var revet ned for å lage dette til et rom som tok opp nesten hele første etasje. (DL1T)
"The walls were torn down…"

The large room described in (20) has in fact been mentioned in a preceding sentence; hence the *walls* are inferable. This fact will have justified the definite form in the translation. But although many examples are of this type, others are harder to explain, such as (21), in which no experts have been mentioned or can be inferred from the context. It is tempting to assume that the change from indefinite to definite subject has been made chiefly to create an apparently smoother information structure.

(21) *Experts* restored the canvas by repairing the boot. (JH1)
Ekspertene restaurerte lerretet ved å reparere støvelen. (JH1T)
"The experts…"

The divergent correspondence that involves turning a bare presentative into a full presentative is less frequent than the indefinite-to-definite change, but still noticeably more common in translations from English into Norwegian. The example given in Table 9 is typical in that the verb is *be*, corresponding to the Norwegian *være*. However, the same type of change is found with other verbs, as shown in (22).

(22) *A queue* had formed in the area newly designated for waiting in... (RR1)
Det hadde dannet seg kø i det nye feltet som var avsatt for ventende kunder… (RR1T)
"There had formed itself queue…"

As noted above, the Norwegian presentative construction is more flexible than the English one as regards the range of lexical verbs that occur in it (Ebeling 2000: 260–261). The translator of (22) thus exploits a possibility that exists in Norwegian, but not in English, of relegating the subject to non-initial position while retaining a full, lexical verb.

Subject change is almost twice as common in translations from English into Norwegian as vice versa. Examples (23)–(25) show three recurrent types of change within this category: (i) another participant from the clause is promoted to subject; (ii) the impersonal pronoun *man* ('one') is introduced; (iii) a pronominal subject is supplied from the context.

(23)   A loud sharp barking suddenly disturbed *the silence*. (MM1)
       Stillheten ble plutselig forstyrret av skarp bjeffing. (MM1T)
       "The silence was suddenly disturbed by sharp barking"

(24)   A lot of inquiries can be done by phone… (SG1)
       Man kan gjøre en god del undersøkelser per telefon... (SG1T)
       "One can do a good deal (of) inquiries by phone, …"

(25)   Three men from town and another farmer named Hawkins helped. (JSM1)
       De fikk hjelp av tre karer fra byen og en gårdbruker som het Hawkins. (JSM1T)
       "They got help from three men from town…"

Subject change is the most frequent type of divergence in translations from Norwegian into English, and the examples are mainly of the same type as those in the other direction of translation. Thus, in (26) a pronominal subject is supplied from the context, and in (27) another participant is promoted to subject in the translation.

(26)   En arm lå over hoften *hans*. (KAL1)
       "An arm lay across the hip his"
       He felt an arm resting on his hip, ... (KAL1T)

(27)   En ny tanke slo ned i *henne*… (EG1)
       "A new thought struck down in her"
       She was suddenly struck by another thought… (EG1T)

6.3   Syntactic restructuring in translation

Beyond the changes in translation mentioned in the previous section, we find instances of syntactic restructuring in translation which removes the indefinite NP from sentence-initial position. That is, some translation correspondences involve a simple reordering of constituents. An example is given in (28), where the adjunct is moved from end to initial position in the translation. This type of change is not very frequent, but found more often in translation from English into Norwegian than in the opposite direction.[13] Interestingly, syntactic reordering in Nor→Eng translation seems to be related to another aspect of adverbial placement, i.e. the lower tolerance for long adverbials in medial position in English than in Norwegian (Johansson and Lysvåg 1987: 264; Hasselgård 2010: 107). Example (29) illustrates the movement of a long medial adverbial to initial position, thus delaying the indefinite subject.

---

[13] There are 11 instances in Nor→Eng translations and 23 in Eng→Nor translations, accounting for 2.1% and 3.2% of the material, respectively.

(28) *A bowl of hot, buttered, boiled potatoes* stood in the middle of the table. (MM1)
Midt på bordet sto *et fat med kokte poteter*. (MM1T)
"In the middle of the table stood a bowl of boiled potatoes"

(29) *En sosial handling* var *i lys av den oppståtte mistanken* blitt asosial, forbrytersk. (KA1)
"A social action had in light of the arisen suspicion become anti-social, criminal"
In light of the suspicion that had arisen, a social action had become anti-social, criminal. (KA1T)

Furthermore, a more thorough restructuring may take place, typically involving subject change (including the use of presentative *there/det*) as discussed in Section 6.2. These types of changes are also much more frequent in English-Norwegian translations. In the Norwegian translations we also find constructions with *det* that are not clearly presentative, but for example impersonal passives or clefts, as illustrated by (30) and (31).

(30) *A house in the city* could be bought for that much. (JH1)
For de pengene kunne det kjøpes et hus midt i byen. (JH1T)
"For that money could there be-bought a house in the middle of the city"

(31) *An intruder* had done this. (RR1)
Det var en tyv som hadde gjort dette. (RR1T)
"It was a thief who had done this"

Syntactic restructuring involving a change of voice is not very frequent in either direction of translation: it is found in only about 5% of the clauses. However, in translations into English, changes from active to passive and from passive to active are about equally common, while in translations into Norwegian, passive to active changes are twice as common as active to passive. This lends some support to Johansson's observation that "English appears to have a greater preference for passive perspectives" (2004: 49), although it should be recalled that the proportions of active and passive clauses were similar in both English and Norwegian originals (Section 5.3).

6.4 Factors affecting congruence in translations

As shown in Table 10, the large majority of indefinite subjects remain unchanged in translation, particularly going from Norwegian into English. Table 11 reports findings from a cross-tabulation of full congruence with the lexicogrammatical features of the subject NP, the verb phrase and the clause pattern. The aim of this exercise is to discover contexts that are particularly favourable to the preservation or change of an indefinite subject NP. A lexicogrammatical feature has been considered to promote either congruence or divergence if the percentage of congruent correspondences differs by at least 5 percentage points from the mean for each direction of translation. The features not mentioned in the table do not seem to pull the degree of congruence up or down from the mean percentage.

Table 11. Full congruence and lexicogrammatical features of the subject/context.

	English→Norwegian	Norwegian→English
Mean rate of congruence	54.9%	74.3%
NP complexity	Complex NP promotes congruence	Simple NP promotes congruence
NP specificity	Generic NP promotes congruence	
NP semantics	Animate and human reference promotes congruence.	Non-human animate promotes congruence.
	Abstract reference reduces congruence.	
VP transitivity	Transitive verb promotes congruence; intransitive reduces it.	Copular pattern reduces congruence.
Complement form	Clausal and indefinite NP complements promote congruence.	
		Adjectival complement reduces it.
Voice	Passive voice reduces congruence	

It is interesting that complex subject NPs should increase the chance of a congruent translation from English into Norwegian. Possibly, the presence of modifiers makes the NP referent more readily identifiable and therefore more acceptable as a subject. In the case of (32), the next sentence also makes anaphoric reference to the clause-final *pool*.

> (32) A statue that probably represented the pursuit of Daphne by Apollo was reflected in the dark waters of a shallow pool. (RR1)
> En statue som så ut til å forestille Apollons forfølgelse av Dafne, speilte seg i det mørke vannet i et grunt basseng. (RR1T)
> "A statue that seemed to represent Apollo's pursuit of Daphne…"

Abstract reference of the indefinite subject NP decreases congruence in both directions of translation. This may have to do with the hierarchy of subject selection proposed e.g. by Givón (1993: 93) by which the preferred subject roles are agent > dative > patient > others. That is, abstract referents may be less likely than (human) animate and concrete ones to have the role of agent. The hierarchy of subject selection may also impact on the relatively low degree of congruence with passive clauses noted in Table 11.

Subjects with a transitive verb followed by a complement promote congruence in translation from English to Norwegian, particularly if the complement is realized by a clause or an indefinite NP. Such sentences will adhere to either the information principle (new information last) or the end weight principle (long constituents last) (Biber *et al*. 1999: 896, 898), which is likely to make the indefinite subject more acceptable, as noted in Section 5.4.

6.5 Translation correspondences: main findings and further research

Translation correspondences of indefinite subjects show that although the subject NP is retained in congruent form in the majority of cases, there is a marked difference between the two directions of translation: more changes are made in translations from English into Norwegian than the other way round. This can be related to the lower overall frequency of indefinite subjects in Norwegian original texts, which in turn is taken to reflect a lower tolerance of such subjects. The most frequent change made in translation from English to Norwegian is replacing the indefinite NP with a definite one. This type of change occurs in the other translation direction too, but much less commonly. The most frequent change in Nor→Eng, and the second most frequent in Eng→Nor, is a restructuring which involves subject change. It may be noted that the changes made tend to bring the translated sentence

into better agreement with the information principle. Certain contexts that were found to be welcoming to indefinite subjects in the contrastive study, particularly generic reference, tend to encourage congruence in translation. Similarly, the presence of a post-verbal complement, particularly one that is either indefinite or long, is conducive to a congruent translation.

This study has only considered sentences with indefinite subjects in original texts and their translations. A future complementary study might look into indefinite subjects in translated texts and identify their sources. For example, translations may contain some sentence-initial subjects that do not appear in the original, as in (33), where the Norwegian original has a sentence-initial direct object (and the subject in post-verbal position due to the V2 constraint).

(33)   Kaméen tar sikkert en gullsmed med glede. (KF1)
       "The cameo takes surely a jeweler with pleasure"
       A jeweler will be glad to take the cameo. (KF1T)

The use of indefinite subjects is likely to be sensitive to genre. Thus, another avenue of further research would be to conduct a similar investigation using for example academic texts. Unfortunately, the ENPC does not contain sufficient material in any non-fiction genre to facilitate such an investigation.

7.   Concluding remarks

The present study set out to investigate sentence-initial indefinite NP subjects in English and Norwegian. Such subjects appear anomalous in the light of word order principles such as the basic distribution of communicative dynamism (Firbas 1992; Dušková 2015) and the light subject constraint (Chafe 1994) because of their association with new information. Most likely for this reason, indefinite subjects have been observed to be rare (Prince 1992, Biber *et al*. 1999). The indefinite subjects were first analysed cross-linguistically on the basis of original texts with respect to their frequencies, their lexicogrammatical features and the contexts in which they occur, particularly regarding features of the verb phrase and the presence of other constituents in the clause. The main hypothesis, based on previous contrastive studies of related issues, was that English would be more tolerant of sentence-initial indefinite subjects than Norwegian. This turned out to be the case. However, the lexicogrammatical features of both subject NPs and their verbs were relatively similar across the languages. This indicates that the differences between the languages are not systemic. That is, the language difference is stylistic rather than structural, although it can be argued, in the words of Johansson (2004: 49), that there is "no clear borderline between structural differences and stylistic preferences."

The study of translations gave additional evidence of the lower tolerance of indefinite subjects in Norwegian in that such subjects are changed more often in translation from English to Norwegian than vice versa. However, the high degree of congruence in both directions give further support to the idea that the differences observed are due to preferences rather than grammaticality. There are probably few cases – given the appropriate lexical resources in the target language – in which a congruent translation would be ungrammatical. When translators make structural changes in spite of the availability of a congruent correspondence, this must reflect language-specific preferences as to which syntactic patterns are perceived as natural and idiomatic.

The results of both parts of the investigation supported the hypotheses but also brought out more detailed information about the use of indefinite subjects. While Johansson (2004) found that the form of subjects in general is preserved in translation from English into Norwegian in about 90% of the cases, the present study showed that sentence-initial indefinite subjects are more prone to change, and moreover, that changes are made much more often in translation from English into Norwegian than from Norwegian into English (Table 10). The difference between translation directions is consistent with Ebeling's (2000) findings regarding the translation of bare presentatives: he argues that since English tolerates indefinite NPs in subject position in bare presentatives to a greater extent than Norwegian, the Norwegian S-V (+Locative) presentatives are expected to be translated by similar English patterns, which is also the case (*ibid.*: 191). Besides the bare presentatives, however, which are an important context for indefinite subjects (*ibid.* and Dušková 2015), generic sentences were found to be favourable to indefinite subjects in both languages. Similarly to sentences that also had indefinite post-verbal complements, these were considered less at odds with the information principle.

The study shows that the light subject constraint (Chafe 1994) is even more apparent in Norwegian than in English. In a cross-linguistic perspective it can be argued that the stricter application of the constraint in Norwegian works well for English in the sense that the Norwegian preferences will generally produce acceptable English sentences. In contrast, the greater tolerance of indefinite subjects in English works less well for Norwegian. Hence translators from English into Norwegian feel a need to change a number of sentences with indefinite subjects, whereas translators from Norwegian into English may find that such sentences in the source text can be rendered congruently because they already lie well within what is considered natural English usage.

References

Baker, M. 1996. Corpus-based translation studies: the challenges that lie ahead. In *Terminology, LSP, and Translation: Studies in Language Engineering in Honour of Juan C. Sager,* H. Somers (ed.), 175–186. Amsterdam: John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Chafe, W. 1986. Writing in the perspective of speaking. In *Studying Writing: Linguistic Approaches,* C.R. Cooper and S. Greenbaum (eds) , 12–39. Beverly Hills: Sage Publications.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Ebeling, J. 2000. *Presentative Constructions in English and Norwegian. A Corpus-based Contrastive Study*. Doctoral thesis. Oslo: Faculty of Arts, University of Oslo.

de Hoop, H and Krämer, I. 2006. Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition*, 13:2, 103–123.

Dušková, L. 1999. Basic distribution of Communicative Dynamism vs. Nonlinear indication of Functional Sentence Perspective. In *Prague Linguistic Circle Papers*, Vol 3, E. Hajičová, O. Leška, P. Sgall and Z. Skoumalová (eds), 249–261. Amsterdam: John Benjamins.

Dušková, L. 2015. *From Syntax to Text. The Janus Face of Functional Sentence Perspective*. Prague: Karolinum Press.

Faarlund, J.T. 1988. A typology of subjects. In *Studies in Syntactic Typology,* M. Hammond, E.A. Moravcsik and J. Wirth (eds), 193–208. Amsterdam: John Benjamins.

Faarlund, J.T., Lie, S. and Vannebo, K.I. 1997. *Norsk Referansegrammatikk.* Oslo: Universitetsforlaget.

Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

Givón, T. 1993. *English Grammar: A Function-based Introduction.* Amsterdam/Philadelphia: John Benjamins.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Arnold.

Hasselgård, H. 2004. Thematic choice in English and Norwegian. *Functions of Language* 11:2, 187–212.

Hasselgård, H. 2005. Theme in Norwegian. In *Semiotics from the North: Nordic Approaches to Systemic Functional Linguistics*, K.L. Berge, and E. Maagerø (eds), 35–48. Oslo: Novus.

Hasselgård, H. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press.

Hasselgård, H., Lysvåg, P. and Johansson, S. 2012. *English Grammar: Theory and Use.* Second edition. Oslo: Universitetsforlaget.

Hawkins, J.A. 1978. *Definiteness and Indefiniteness. A Study in Reference and Grammaticality Prediction*. London: Croom Helm.

Johansson, S. 2004. Why change the subject? On changes in subject selection in translation from English into Norwegian. *Target* 16(1), 29–52.

Johansson, S. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam/Philadelphia: John Benjamins.

Johansson, S., Ebeling, J. and Oksefjell, S. 1999/2002. English-Norwegian Parallel Corpus: Manual. http://www.hf.uio.no/ilos/english/services/omc/enpc/ENPCmanual.pdf.

Johansson, S. and Lysvåg, P. 1987. *Understanding English Grammar. Part II: A Closer View*. Oslo: Universitetsforlaget.

Lyons, C. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Matthiessen, C.M.I.M. 1995. *Lexicogrammatical Cartography: English Systems*. Tokyo: International Language Science Publishers.

Matthiessen, C.M.I.M. 1999. The system of transitivity: An exploratory study of text-based profiles. *Functions of Language* 6(1): 1–51.

Oostdijk, N. and de Haan, P. 1993. Clause patterns in Modern British English: A corpus-based (quantitative) study. *ICAME Journal* 18: 41–79.

Prince, E.F. 1981. Toward a taxonomy of given–new information. In *Radical Pragmatics,* P. Cole (ed.), 223–255. New York: Academic Press.

Prince, E.F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text,* W.C. Mann, and S.A. Thompson (eds.), 295–326. Amsterdam: John Benjamins.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rørvik, S. 2013. *Texture in Learner Language*. PhD thesis, University of Oslo.

Author's address

Hilde Hasselgård
Department of Literature, Area Studies and European Languages
University of Oslo
PO box 1003
0315 Oslo
Norway

hilde.hasselgard@ilos.uio.no

English supplementive *ing*-clauses and their German and Swedish correspondences

Jenny Ström Herold, Magnus Levin

Linnaeus University (Sweden)

Abstract: This paper investigates English supplementive *ing*-clauses (e.g., *Hitler exploded, demanding examples.*) in German and Swedish contrast. The material consists of popular non-fiction originals and their translations from the Linnaeus University English-German-Swedish corpus (LEGS) (version 0.1). The results show that coordination is the most frequent correspondence of supplementive *ing*-clauses in German and Swedish translations and originals. Like the supplementive *ing*-clause, a coordination is a compressed and semantically indeterminate structure. The other major correspondences include subordination, main clause and prepositional phrase. German translators more often use main clauses than Swedish translators, which seems to be related to an increasing German tendency for parataxis rather than hypotaxis. A number of German and Swedish instances involve different kinds of explicitation, including conjunctions and German pronominal adverbs.

Keywords: supplementive *ing*-clauses, free adjuncts, explicitation, the Linnaeus University English-German-Swedish corpus (LEGS), English/German/Swedish

1. Introduction

Nida (1964: 209) notes that "[t]he most acute problem in clause correspondence [in translation] occurs when a clause type that is important in the source language simply does not exist in the receptor language." A prime example of this kind of clause is the English supplementive *ing*-clause, which lacks productive equivalents in many languages. In the present study we define supplementive *ing*-clauses as zero-introduced subjectless non-finite, subordinate clauses in adverbial function (see, e.g., Quirk *et al.* 1985: 1123–1125; Biber *et al.* 1999: 820).[1] What makes supplementive *ing*-clauses particularly difficult to translate is that they "typically have an implicit and somewhat ill-defined relationship with the main clause" (Biber *et al.* 1999: 782–783).

The variation resulting from the lack of equivalent target-language structures along with the semantic indeterminacy of these clauses is illustrated in (1) and (2) from the

---

[1] For an overview of the variable terminology, free adjuncts being the most common variant, see Kortmann (1991: 18).

Linnaeus University English-German-Swedish corpus (LEGS). English source texts (EN ST) are followed by their German and Swedish translations.

(1)   In a bird the wing bones and leg bones are chopped through, *leaving the wings and feet attached to the skin*. (LEGS; EN ST)
      Bei Vögeln werden die Flügelknochen und Beinknochen durchschnitten, *Flügel und Füße bleiben mit dem Balg verbunden.*
      "wings and feet remain attached to the skin"
      Hos en fågel hugger man igenom vingbenen och benknotorna *och låter vingarna och fötterna sitta kvar vid skinnet.*
      "… you cut through … and let the wings and feet remain attached to the skin"

(2)   *Having run out of anti-tank ammunition*, soldiers of the exhausted and badly depleted 2nd Royal Norfolk Regiment were reduced to dashing out with hand-grenades to drop them into the tracks of the panzers. (LEGS; EN ST)
      Da ihnen die Panzerabwehrmunition ausging, konnten die Soldaten des erschöpften und stark dezimierten 2nd Royal Norfolk Regiment nur noch mit Handgranaten kämpfen, die sie in die Raupenketten der Panzer warfen.
      "since they ran out of ammunition"
      När den utmattade och svårt decimerade 2. bataljonen ur infanteriregementet "Royal Norfolk" fick slut på pansarvärnsgranater tvingades männen rusa ut med handgranater och släppa dem i stridsvagnarnas band.
      "when the exhausted and badly depleted 2nd battalion from the infantry regiment "Royal Norfolk" ran out of anti-tank ammunition"

In (1) the German and Swedish translators have chosen different target-language structures for the supplementive *ing*-clause. The German translator uses a main clause in which the subject 'wings and feet' is asyndetically linked to the first main clause. The Swedish translator retains one main clause by using a VP coordination. In (2) both translators have chosen the same structure, a subordinate clause, but the implicit link between the two clauses has been interpreted differently. The German version contains the causal conjunction *da* ('since') and the Swedish the temporal conjunction *när* ('when'). Thus, translations may involve more explicit alternatives (see, e.g., Blum-Kulka (2004 [1986]: 292) on explicitation). In examples such as (2), when translators opt for a dependent adverbial clause, it is even obligatory.

Supplementive *ing*-clauses have previously been studied from a contrastive perspective (e.g., Lindquist 1989; Blensenius 2006; Fischer 2013), but to date there has been no large-scale quantitative study. The present investigation of more than 1300 supplementive *ing*-clauses includes comparisons between English, German and Swedish, and addresses the following research questions:

1)   how frequent are supplementive *ing*-clauses and which sentence positions do they occupy in English originals and translations from German and Swedish,

2)   what German and Swedish target-language correspondences are used as translations of supplementive *ing*-clauses, and to what extent are Translation Universals (cf. Baker 1993; Chesterman 2004), such as explicitation, reflected in these choices,

3) how is the very same instance of an *ing*-clause rendered into German and Swedish, i.e. to what extent is there (non-)congruency between translations,[2]

4) what German and Swedish ST structures are rendered as supplementive *ing*-clauses.

2.   Background

2.1   English supplementive *ing*-clauses

The English supplementive *ing*-clause has received attention both in traditional grammars and monographs (Quirk *et al.* 1985: 1123–1126; Kortmann 1991; Biber *et al.* 1999: 782–783, 820, 829–833, 840, 907–908). These studies often focus on its grammatical properties and relation to similar constructions such as the absolute (Stump 1985; Kortmann 1991),[3] but also its semantic flexibility. The following examples illustrate the semantic diversity of the supplementive *ing*-clause, where (3) induces a temporal, (4) a causal and (5) a circumstantial reading, often referred to as accompanying circumstance (Kortmann 1991).

(3)   *Driving home after work*, I accidentally went through a red light. (Quirk *et al.* 1985: 1121)

(4)   John, *knowing that his wife was expecting a baby*, started to take a course on baby care. (Quirk *et al.* 1985: 1123)

(5)   "Oh all right, then," she said, *concealing her disappointment*. (Biber *et al.* 1999: 832)

As is evident from the examples above, the supplementive *ing*-clause can take three different positions within the complex sentence, initial (as in (3)), medial (as in (4)) and final position (as in (5)). There seems to be consensus that the final position is by far the most common position (see, e.g., Kortmann 1991: 9, 139; Biber *et al.* 1999: 830–833). As for the medial position, it can be difficult to distinguish the adverbial reading from the relative clause reading (Quirk *et al.* 1985: 1125; Kortmann 1991: 9). In writing, most supplementive *ing*-clauses are set off from the matrix clause by punctuation marks, such as commas or dashes. As pointed out by Stump (1985: 4), the punctuation criterion is not water-tight, as seen in (6), where the medial clause is not separated from the surrounding discourse by any punctuation marks.

(6)   Obama's much-heralded move in June 2014 *mandating emission reductions from power plants* was certainly the right direction, but the measures were (…) (LEGS; EN ST)

Thus, the material for this paper includes examples of both medial and final positionwith and without punctuation marks (as also, for instance, in Malá 2005), the requirement being that both authors independently perceived a looser semantic attachment to the matrix clause than seen with attributive relative clauses.

---

[2] In this paper congruency refers to a structural comparison between two target texts. Two translations are deemed to be congruent if they belong to the same category, such as coordination. This is in contrast to Johansson (2002–2003), who defines congruence as a relation between a source-text structure and a target-text structure.

[3] Absolutes include non-verbal instances, such as *Not a penny over, we had to leave the town* (Kortmann 1991: 10).

As already indicated in (2)–(5) above, supplementive *ing*-clauses can have a number of adverbial interpretations. Biber *et al.* (1999: 783) exemplify this with (7):

(7)   The result of the operation is placed in the accumulator, *destroying its previous contents.*

This *ing*-clause is ambiguous between three readings: the result reading, and two different temporal readings, i.e. simultaneity and posterity. According to Kortmann (1991), the interpretation process is mainly inference-driven, but verb semantics has also been considered an important factor (e.g., Stump 1985; Behrens 1998). Kortmann (1991) identifies as many as fourteen different semantic categories, ranging from different temporal interpretations to concessive, instrumental and result readings. In his view, the supplementive *ing*-clause is "an ideal problem-solving device for remaining obscure" (Kortmann 1991: 114) and different readings may coexist (ibid.: 112). Because of this semantic indeterminacy, this paper, in contrast to Kortmann, will not discuss semantic aspects to any great extent, but rather focus on German and Swedish correspondences from a structural perspective.

Supplementive *ing*-clauses also have a bearing on information structure. Due to their subordinate nature *ing*-clauses are generally considered backgrounded (Behrens and Fabricius-Hansen 2005: 9). Thus, according to Kortmann (1991: 113), "[i]t is at least this piece of information, i.e. the presentation of one proposition as backgrounded which always gets lost in paraphrases of free adjuncts/absolutes by means of coordinate clauses" [or main clauses [our addition]].

Previous studies (Kortmann 1991: 39; Biber *et al.* 1999: 821) indicate that supplementive *ing*-clauses are a highly genre-specific feature (see Table 1 below). They seem to be the most frequent in fiction, but so far no large-scale quantitative study has been performed on translated texts. The next section discusses the relevant previous contrastive work.

2.2   Supplementive *ing*-clauses from a contrastive perspective

One of the largest contrastive datasets on translations of supplementive *ing*-clauses is provided by Lindquist (1989: 120–128) on English adverbials in fiction translated into Swedish. His source-text material contains 93 adverbial *ing*-clauses. The four most frequent translation types (except for "deletion") are as follows: finite clause (45%), which includes the three subtypes VP coordination, new full finite clause and adverbial subclause introduced by a conjunction; infinitive (11%); non-finite *ing*-clause[4] (8%), and PP (8%).

Below are two of Lindquist's (1989: 126–127) examples from his category of finite clauses. The first of these, (8), is translated into a VP coordination and the second one, (9), into a subordinate clause, like the Swedish translations of (1) and (2) above. Lindquist notes that the translator of (8) has rendered the simultaneous event in the original as a temporally ambiguous structure in Swedish, while (9) is expressed as a causal relation. These translations support Cosme's (2008: 105) observation that a finite adverbial clause in general is more explicit than coordination.

(8)   […] said Mabs, watching through field glasses from the bedroom of Cadbury Farm.
        […] sa Mabs, och studerade dem i kikaren från sovrumsfönstret på Cadbury Farm.
        "and studied them through the field glasses"

---

[4] Lindquist's term; in Swedish formed with the *-ande/-ende* suffix.

(9)　Moral confusion excited him sexually […] giving him time to think […].
　　　Moralisk förvirring hetsade upp honom sexuellt […] eftersom det gav honom tid att tänka efter […].
　　　"since it gave him time to think"

Blensenius (2006) investigates supplementive *ing*-clauses in a corpus of English economics texts translated into Swedish. His results from non-fiction (2006: 33) seem to be in line with Lindquist's in that the majority of the Swedish translation equivalents are finite clauses. However, there are no quantifications of his rather limited data to support this claim.

Behrens (1998) also presents a qualitative study of the translation of English supplementive *ing*-clauses, in this case into Norwegian, which is closely related to Swedish. Behrens claims that semantics (the event structure of the *ing*-clause's verb phrase) plays a major role in the semantic resolution of the *ing*-clause, thus partly arguing against Kortmann's (1991) inference-driven approach. One notable finding is that Norwegian translators occasionally add explicit markers, "discourse particles" in Behrens' (1998: 259ff.) terminology, making the relevant interpretation overt (see Section 4.3.3).

Behrens and Fabricius-Hansen (2005: 5) notice a similar translation strategy in translations into German, as illustrated in (10). In this case the German translator has added the temporal connective *dabei* (lit. 'thereby'), thus spelling out the co-temporal relation holding between the two coordinated conjuncts.

(10)　He smiled slyly, *nodding*.
　　　Er lächelte verstohlen *und nickte dabei*.
　　　"and nodded thereby"

Fischer (2013) compares German and English sentence structure in a parallel corpus of fiction that includes originals and translations in both directions. He notes (ibid.: 169, 171) that the English texts have almost twice as many non-finite VPs – infinitives and participles – than the German texts, with present participles being as much as five times more common in English. Moreover, present participles are more frequent in German translations than in German originals. According to Fischer (ibid.: 171), this is probably a translation effect, resulting from the translator copying the source-text structure.

Finally, contrastive studies have been made with other languages than German and the Scandinavian languages. Cosme (2008) is a corpus-based contrastive study of clause-linking patterns in Dutch, English and French, focusing on the distribution of subordinating and coordinating structures – the latter interpreted in the broadest sense (also including juxtaposition of two independent main clauses) – in these languages. A finding relevant for the present study is that *ing*-clauses are often translated into Dutch as coordination, either as coordination of VPs or full finite clauses.

The previous contrastive work is thus rather limited and largely qualitative in nature. Nevertheless it seems that finite target-language structures predominate as correspondences in various Germanic languages. The present study, which draws on the most extensive dataset investigated to date, will indicate to what extent finite clauses are used as German and Swedish correspondences of supplementive *ing*-clauses in non-fiction.

3.　　Material and method

This study is based on the Linnaeus University English-German-Swedish corpus which is being compiled at Linnaeus University, Sweden, by the present authors. The corpus contains

recently published popular non-fiction books in English, German and Swedish with translations into the other two languages. Due to the low numbers of non-fiction books translated from Swedish and German and the few translators who produce a sizeable proportion of these translations, we settled for a large number of words from each text − at least 50,000 words or the whole book. Introductory chapters were excluded. No authors or translators are represented by more than one text each.

At the time of writing (version 0.1), LEGS comprises five English originals with translations into both German and Swedish, and three originals each from German and Swedish with their respective translations. The English originals amount to 272,000 words, the translations from German 184,000 and the translations from Swedish 150,000. Of the sixteen translations included in this study, six were translated by more than one translator. Although it is not clear exactly how many translators were involved in the sections selected for the corpus, there are certainly more than sixteen translators represented in the three subcorpora.

The sub-genres covered for each source language so far are largely comparable. For each source language there is one biographical text. Popular science and history are represented in both English and German, and English and Swedish both have texts concerned with political and societal issues.

Texts for inclusion were identified through searches for translated books with the respective source and target languages in the national Swedish library database Libris. Included are English originals published in the 2010s, while for the German and Swedish originals we include volumes from the whole of the 2000s.

The texts were scanned and manually corrected. The source texts were aligned semi-automatically with their respective target texts by a research assistant using the alignment function in the SDL Trados Studio translation software.[5] Laurence Anthony's parallel corpus software tool AntPConc[6] was used to search the aligned files using the search string *ing. This produced a large amount of noise, such as progressives and gerunds, that was weeded out manually. Both researchers agreed on which instances to include and how to classify the supplementive *ing*-clauses and their correspondences.

4. Results

4.1 Translation categories identified in the material

Four major categories of German and Swedish correspondences of supplementive *ing*-clauses were identified in LEGS: coordination, subordination, main clause and prepositional phrase (PP), and nine minor categories that were conflated into the Other category.

The main clause category comprises instances with full finite clauses. As exemplified in (11), these may involve new sentences separated by full stops, semicolons (as in (28) below), or, as in (1), commas, but sometimes also two coordinated full main clauses with subjects, as in (12). The coordination category instead includes VP conjunction where the subject is always omitted in the second conjunct, as in (13), and sometimes also the auxiliary. Thus, in contrast to the category main clause, coordination always entails some kind of reduction. Subordination covers adverbial clauses (as in (14)), sentential relative clauses and post-modifying relative clauses. The PP category consists of prepositional phrases introduced by simplex or complex prepositions, e.g. *med hjälp av* in (15).

---

[5] http://www.sdl.com/store/
[6] http://www.laurenceanthony.net/software.html

- Main clause

  (11) […] "Gobble, gobble, gobble, gobble," he said, *cracking up Smith and Hertzfeld.*
  (LEGS; EN ST)
  […] unterbrach ihn Jobs: "Bla, bla, bla." *Smith und Hertzfeld mussten lachen.*
  "Smith and Hertzfeld had to laugh"

  (12) Dumyat is the westernmost of the Ochils, *rising only 400 metres or so*, but […]
  (LEGS; EN ST)
  Dumyat är det västligaste berget i Ochilkedjan *och det är bara cirka 400 meter högt*
  men […]
  "and it is only circa 400 metres high but…"

- Coordination

  (13) Hitler exploded, *demanding examples.* (LEGS; EN ST)
  Hitler war außer sich *und wollte Beispiele genannt haben.*
  "Hitler was beside himself and wanted to hear examples"

- Subordination

  (14) Walking past the lineup of tables set up by the Heartland conference's sponsors, it's
  not terribly hard to see what's going on. (LEGS; ENG ST)
  När man går förbi de bord som ställts upp av Heartlandkonferensens sponsorer är det
  inte alls svårt att inse vad som är på gång.
  "when you walk past the tables that have been set up by the Heartland conference's
  sponsors, it is not at all …"

- Prepositional phrase (PP)

  (15) […] the Japanese crossed the Soochow Creek *using small metal assault boats* […].
  (LEGS; ENG ST)
  […] gick japanerna […] över Suzhoufloden *med hjälp av små landstigningsbåtar av
  metall* […]
  "with the help of small landing craft of metal"

The minor categories represent various translation solutions, as illustrated in (16)–(24) below.
Example (16) shows one of the rare instances where the *ing*-clause has been rendered as an
adjective phrase.

- Adjective phrase (adjP)

  (16) Weygand […] demanded more RAF fighter squadrons, *knowing that the British
  must refuse*. (LEGS; ENG ST)
  Weygand […] krävde fler jaktflygdivisioner från RAF, *väl medveten om att
  britterna skulle tvingas neka*.
  "well aware that the British would have to refuse"

The supplementive *ing*-clause in (17) is translated into an adverb phrase.

- Adverb phrase (advP)
  (17) These results underscore the importance of regulating attention to control and cool down stress, *beginning early in life*. (LEGS; ENG ST)
  Dessa resultat understryker vikten av att *redan tidigt i livet* styra uppmärksamheten till kontroll och nedkylning av stress.
  "already early in life"

The Swedish infinitives generally consist of prepositions followed by the infinitive marker *att*, e.g., *efter* ('after'), *för* ('in order to') and, as in (18), *på* ('on'). German infinitives mostly involve *um zu* ('in order to') (see further section 4.3.3 on explicitation).

- Infinitive clause
  (18) […] while the government wasted hundreds of millions (at least) *trying to clean up the unnecessary messes*. (LEGS; ENG ST)
  […] medan provinsregeringen slösade bort hundratals miljoner (minst) *på att försöka städa upp en aning i den onödiga röran*.
  "on to try to clean a bit in the unnecessary mess"

The noun phrase category involves examples such as (19) where the content of the *ing*-clause is rendered as a complex noun phrase.

- Noun phrase (NP)
  (19) *Reflecting its cheeky confidence*, Apple took out a full-page ad […]. (LEGS; ENG ST)
  Ett tecken på det fräcka självförtroendet var att man köpte en helsidesannons […].
  "a sign of the cheeky confidence [was that…]"

Participles, as in (20), cover both present and past participles.

- Participle
  (20) […] the Germans rushed the river in their heavy rubber assault boats, *paddling furiously*. (LEGS; ENG ST)
  […] überquerten die Deutschen den Fluss, *heftig paddelnd* in ihren schweren Gummibooten.
  "furiously paddling"

In the small category of verb phrases, we have included structures constituting parts of a matrix clause. As in (21), these only include a non-finite verb (*setzen*) and an optional adjunct (the participial adverb *zitternd* ('shivering')). Note that the German rendering itself contains a VP coordination where the second conjunct (*und damit Wärme erzeugen*) makes the causal relation explicit. Example (21) thus illustrates the complexity of many target-text structures found in the material, sometimes bordering on the rephrased category.

- Verb phrase (VP)
  - (21) With a stomach full of sugar she can start to fire up her flight muscles, *shivering them to produce heat*, and once she gets up to about 30°C, off she goes... (LEGS; ENG ST)

    Wenn ihr Magen voller Zucker ist, kann sie ihre Flugmuskulatur *zitternd in Bewegung setzen und damit Wärme erzeugen*, und wenn sie eine Temperatur von über 30°C erreicht hat, fliegt sie davon...

    "to set shivering in motion and thereby produce heat"

Rephrased instances, as in (22), contain much the same content as the supplementive *ing*-clause but in a syntactically and lexically altered form. The rephrased and omission (in (23)) categories form a continuum where the most extreme case, omission, contains no trace of the original *ing*-clause (marked by Ø in (23)). Addition (as in (24)) adds new information in the translation (see similar German-English examples in Fischer, 2013: 171) and can be seen as a mirror image of omission.

- Rephrased
  - (22) The whole village or neighbourhood, *paying homage to these martial values,* would usually turn out to bid farewell to a conscript departing to join the army. (LEGS; ENG ST)

    Soldatische Werte wurden so hoch gehalten, dass ein ganzes Dorf oder Wohnviertel einen Wehrpflichtigen verabschiedete, wenn er zur Armee ging.

    "martial values were so highly regarded that"

- Omission
  - (23) *Giving the country partial credit for the collapse of the Russian economy,* a New York Times Magazine piece in 2000 pronounced that "amid the recent proliferation of money-laundering centers that experts estimate has ballooned into a $5 trillion shadow economy, Nauru is Public Enemy #1." (LEGS; ENG ST)

    Ø New York Times Magazine förklarade i en artikel från år 2000 att "i den senaste tidens ökning av centraler för penningtvätt som enligt experter har växt till en skuggekonomi på fem biljoner dollar är Nauru allmänhetens fiende nr 1." […].

    "New York Times Magazine explained in an article from the year 2000 that […]"

- Addition
  - (24) Unter dem Mikroskop sieht ein Arzt dann ovale Eier. (LEGS; GE ST)

    "under the microscope a doctor then sees oval eggs"

    The doctor will examine the fruits of your labors under the microscope, *hunting for little oval eggs*.

The overview of examples shows that there is considerable variation regarding the construction types used as correspondences of supplementive *ing*-clauses. Nevertheless, the result section will mainly be focusing on the four major categories, coordination, subordination, main clause and prepositional phrase (PP).

Section 4.2.1 presents the counts of supplementive *ing*-clauses in originals and translations and compares these frequencies with previous studies, while Section 4.2.2 discusses some findings in relation to sentence position.

4.2   Quantitative overview

4.2.1 *The frequency of supplementive ing-clauses*
Figure 1 presents the frequencies of supplementive *ing*-clauses in the three subcorpora. This study is based on 709 English original examples translated both into German and Swedish, 456 German and 192 Swedish original structures translated into supplementive ing-clauses, in all 1357 *ing*-clauses and 1165 German and 901 Swedish correspondences.[7]
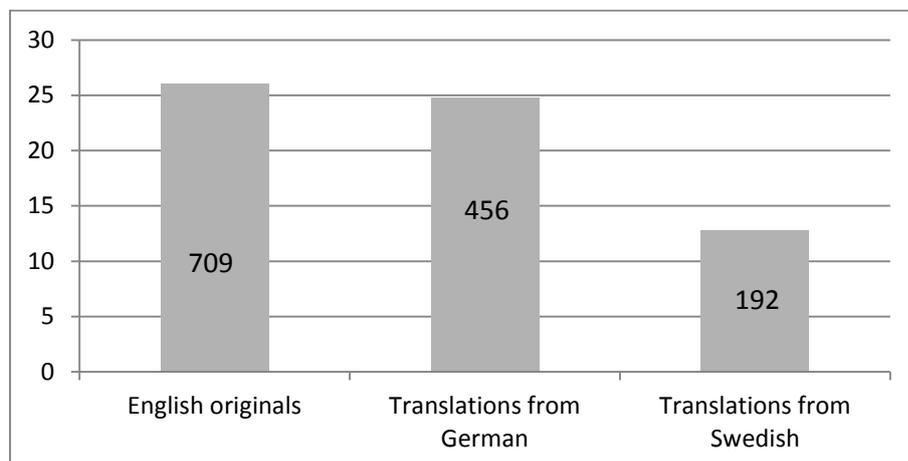

Figure 1. The frequencies of supplementive *ing*-clauses per 10,000 words.

The frequency of supplementive *ing*-clauses is lower in translations from Swedish than in English originals and translations from German but it is hard to draw firm conclusions because only three Swedish original texts are included in the study. It is noteworthy, however, that the two texts with the lowest token frequencies (4 and 10 *ing*-clauses/10,000 words) are based on Swedish originals. One of these largely contains short sentences and sentence fragments, making complex sentence structures less likely in translations. However, while there are notable differences across texts, no clear genre-specific differences emerge. For instance, the highest ratio for an individual text (47/10,000) occurs in one of the English popular science originals. This is three times higher than that of the lowest ratio in the English original biography (15/10,000). In contrast, the translation from Swedish with the highest frequency was the biography, and this contained more instances (22/10,000) than the English original biography.

In spite of the variation between individual texts, Table 1 shows that the frequencies in the three LEGS subcorpora, English originals, translations from German and translations from Swedish, are within the range of the non-fiction genres investigated in previous studies. While the semantically implicit *ing*-clauses are a typical feature of fiction and especially rare in unplanned conversation, they seem to be of intermediate frequency in various non-fiction genres. Among the non-fiction genres in Table 1, the English originals and translations from German produce a fair number of instances.

---

[7] We would like to thank Professor Jukka Tyrkkö, Linnaeus University, for assistance with statistical tests.

Table 1. Frequency comparisons with Kortmann (1991: 39)[8], Biber *et al.* (1999: 821) and Blensenius (2006).[9]

	Ing-clauses per 10,000 words
Fiction (Kortmann 1991)	60.9
Fiction (Biber *et al.* 1999)	c. 40
News (Kortmann 1991)	26.5
ENGLISH ORIGINALS (LEGS)	26.1
TRANSLATIONS FROM GERMAN (LEGS)	24.7
Science (Kortmann 1991)	16.6
TRANSLATIONS FROM SWEDISH (LEGS)	12.8
Economics text (Blensenius 2006)	11
Spoken language (Kortmann 1991)	10.1
Academic (Biber *et al.* 1999)	c. 10
News (Biber *et al.* 1999)	c.10
Conversation (Biber *et al.* 1999)	"almost non-existent"

In many ways, the LEGS material is similar to fiction and news reporting. For instance, the narrative parts of the biographies and history texts are comparable to fiction, while the popularized descriptions of scientific processes and phenomena seem closer to those found in newspapers rather than in academic texts. In view of these observations, it can be expected that the frequencies in LEGS would fall within the range of those found in the previous studies.

4.2.2 *Sentence position*
For the different positions within the complex sentence – initial, medial, and final –, we base our classification on Quirk *et al.* (1985: 490–501). As mentioned above, Biber *et al.* (1999: 830–833) found sentence-final position to be the unmarked choice for non-finite adverbial clauses and the medial position to be very rare.[10] Our results on the positions of supplementive *ing*-clauses in English originals and in translations from German and Swedish given in Figure 2 support these findings.

---

[8] The frequency information from Kortmann (1991) is based on a limited set of texts. The fiction data comprises a handful of texts, the news material was collected from one issue each of *The Guardian* and *International Herald Tribune*, and the science subcorpus consists of about 120 pages of linguistics texts from a single volume. Moreover, the texts do not appear to have been available in electronic format, which means that the most solid quantitative information can be found in Biber *et al.* (1999).

[9] Behrens and Solfjeld's (2014: 274) frequency (200/10,000 words) in English original fiction from the Oslo Multilingual Corpus is based on an extrapolated estimate and differs greatly from all the other studies.

[10] Behrens and Solfjeld's (2014: 274) estimates deviate greatly from the other findings in Table 1. Their results suggest that sentence-final position is almost 80 times more frequent than the sentence-initial one. In each case, the first 100 instances were classified and then the proportions were extrapolated by Behrens and Solfjeld.
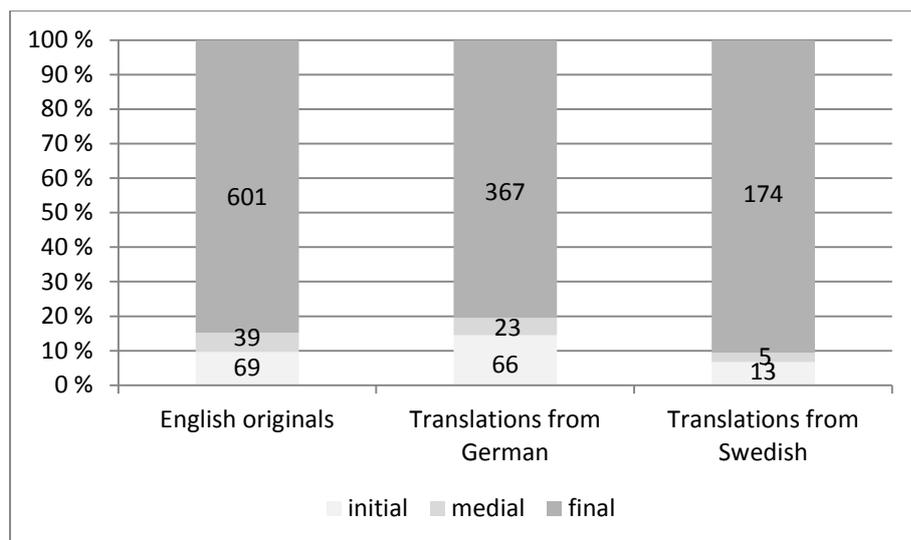
Figure 2. Positions of supplementive *ing*-clauses in English originals and in translations from German and Swedish.

Most of the translations are kept in the same sentence positions as their corresponding source-text structures.[11] For instance, of 601 sentence-final *ing*-clauses in English originals 77% (464) are translated into sentence-final correspondences in German and 90% (543) in Swedish, while slightly fewer of the 69 English sentence-initial clauses are kept in that position in translations (72% (50/69) in German; 78% (55/69) in Swedish).[12] The rarest position, the sentence-medial one, is an exception, however, as it has a lower level of "matching" position in German and Swedish translations.[13] It is likely that the marked and difficult-to-process (Biber *et al.* 1999: 830; Hasselgård 2010: 107–110) positions in the middle of a sentence is often avoided in translations due to a normalization strategy of unusual constructions (cf. Baker 1996: 183) even if the same position would be syntactically possible in the target languages.

Sentence-final *ing*-clauses are the most frequent in the translations from Swedish and the least frequent in the translations from German. The relative preference for sentence-initial position in translations from German mostly stems from sentence-initial German PPs and participles being retained in initial position as *ing*-clauses. It is nevertheless difficult to draw conclusions about translations from Swedish as there are too few instances.

Translations of *ing*-clauses in initial and medial positions are similar to each other in that both positions favour renderings into subordinate clauses in both German and Swedish, but there are notable differences in the preferred kinds of subordinate clause. There is a trend for sentence-initial *ing*-clauses to be translated into subordinate clauses: 43% (30/69) into German and 42% (29/69) into Swedish.[14] Of these, a large majority are translated into adverbial clauses (97% (29/30) in German and 83% (24/29) in Swedish). As exemplified in (25), initial *ing*-clauses can be rendered as temporal clauses introduced by *als/när* ('when')

---

[11] Behrens and Fabricius-Hansen (2005: 111) discuss such cases in terms of discourse structure and "downgrading effect", arguing that a final *ing*-clause keeps its backgrounded discourse status when translated into a sentence-initial conjunct.

[12] Not all correspondences could be classified according to specified sentence positions. This mainly applies to main clauses occurring in separate sentences, as in (11) above.

[13] 49% (19/39) are translated into medial German correspondences and 44% (17/39) into Swedish medial correspondences.

[14] This is significantly higher (according to a chi-square test) than the proportion of adverbial clauses in sentence-final position both in German (22% (133/601); p < 0.01) and in Swedish (25% (149/601); p < 0.01) translations from English.

(see further Section 4.3.3 on explicitation), the *ing*-clause here introducing a frame in which the activity of the main clause occurs (cf. Biber *et al.* 1999: 832).

(25) *Showing off the Homestead campus four decades later,* Jobs paused at the scene of the escapade and pointed.
Als Jobs 40 Jahre später über den Homestead-Campus schlenderte, blieb er stehen und deutete auf einen Balkon:
"when Jobs 40 years later strolled across the Homestead campus"
När Jobs visade mig Homesteads skolgård fyra årtionden senare stannade han till vid skådeplatsen för upptåget och pekade.
"when Jobs showed me Homestead's schoolyard four decades later"

In contrast to the initial position, medial *ing*-clauses are typically translated into relative clauses, which could perhaps be expected from their similarity with relative clauses, as discussed by Quirk *et al.* (1985: 1125) and Kortmann (1991: 9) above. 56% (22/39) of the medials were translated into subordinate clauses in German and 44% (17/39) into Swedish ones. Of these, most are post-modifying relative clauses (68% (15/22) in German and 71% (12/17) in Swedish).[15] The position immediately after the subject and a function close to that of a relative clause make post-modifying relative clauses readily available choices, as seen in (26).

(26) Most Germans, *having feared another bloodbath in Flanders and Champagne,* were overjoyed by the astonishing victory.
Die meisten Deutschen, *die ein weiteres Blutbad in Flandern und der Champagne befürchtet hatten,* waren angesichts des erstaunlichen Sieges überglücklich.
De flesta tyskar, *som hade fruktat ett nytt blodbad i Flandern och Champagne,* var överlyckliga över den häpnadsväckande segern.
"most Germans, who had feared another bloodbath…"

In conclusion, most supplementive *ing*-clauses occur in sentence-final position. This holds true for both originals and translations. Furthermore, there are some indications that translators avoid the marked medial position for adverbial clauses, but when the position is kept the clause is often rendered as a post-modifying relative clause. The next section presents an overview of the German and Swedish correspondences.

4.3   Quantitative overview of correspondences

4.3.1 *German and Swedish translations of English supplementive ing-clauses*
The German and Swedish translations of English source text *ing*-clauses are given in Table 2.

---

[15] Although only 5.5% of the supplementive *ing*-clauses occur in medial positions, they account for fair proportions of all post-modifying relative clauses found in both German (25%; 15/60) and Swedish (18%; 12/65) translations.

Table 2. German and Swedish translations of English supplementive *ing*-clauses.

	Translations into German		Translations into Swedish	
	N	%	N	%
coordination	246	34.7	287	40.5
subordination	185	26.1	197	27.8
main clause	167	23.6	90	12.7
PP	47	6.6	58	8.2
Other	*64*	*9.0*	77	10.9
infinitive	17	2.4	44	6.2
NP	7	1.0	14	2.0
participle	12	1.7	7	1.0
rephrased	13	1.8	2	0.3
omission	9	1.3	3	0.4
VP	4	0.6	3	0.4
adjP	2	0.3	1	0.1
advP	0	0	3	0.4
Total	**709**	**100**	**709**	**100**

The distributions across the two target languages are fairly similar. The four main categories coordination, subordination, main clause and PP follow in the same order and together account for around 90% of all translations. There is only one significant difference between the translations, a greater German preference for main clauses. This will be discussed below.

There is a striking difference between Lindquist's (1989: 121) Swedish fiction data and those from the LEGS corpus. Lindquist's finite clause category, which covers the three most frequent categories (coordination, subordination and main clause), accounts for only 45% in his fiction material, but in our non-fiction material these three cover more than 80%. In contrast, Lindquist's material contains more translations from the Other category (21%). Still, the content of Lindquist's Other category is more restricted than in the present study in that it does not include infinitives, or rephrased or omitted instances. However, Lindquist does not elaborate further on what his category contains, which rules out any further comparisons. The high degree of variation in Lindquist's material may indicate greater translator creativity in fiction than in non-fiction. The distributions of some of the minor categories in LEGS are fairly similar to Lindquist's: 8% PP for both Lindquist and the present study and 11% (Lindquist) vs. 6.2% (LEGS) for infinitives. The more frequent use of participles in Swedish fiction translations (8% as compared to 1%) is probably genre-related, since the short (one-word) examples of *ing*-clauses given by Lindquist (1989: 122) seem to be typical of fiction. The marginal use of participles (only 2% (42/2066) of all correspondences in the LEGS corpus) in the translation of supplementive *ing*-clauses shows that participles are no close German[16] or Swedish equivalents of these English constructions.

Figure 3 gives a visual representation of the findings, indicating the only significant difference in the greater German preference for main clauses.[17]

---

[16] The low German numbers are noteworthy in view of Fischer's (2013: 169, 171) finding that non-finite clauses are more common in German translations from English than in German originals.

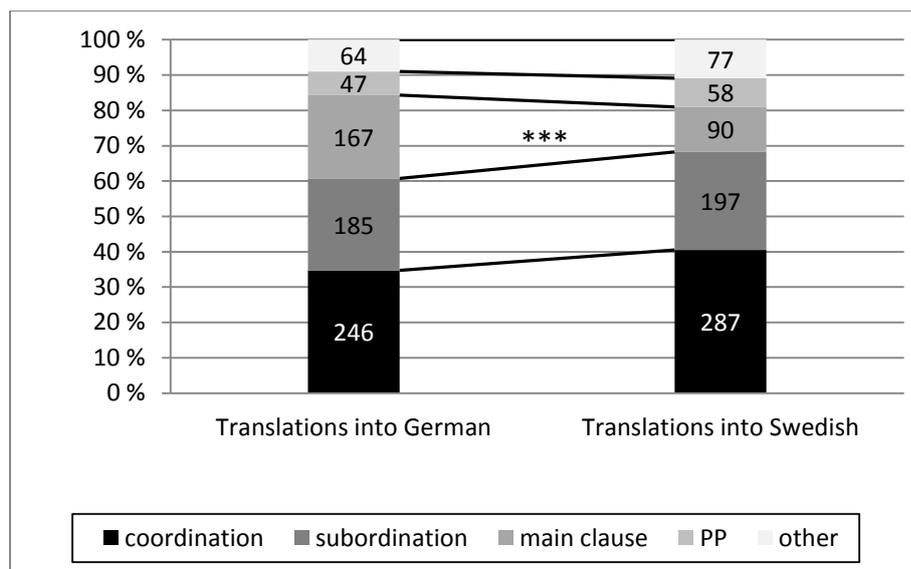[17] According to a chi square test and a post hoc test with a Bonferroni correction; df=4, X2=28.95, p=***.

Figure 3. German and Swedish translations of English supplementive *ing*-clauses.

German translations produce more main clauses than the Swedish ones do. This result is probably related to changing preferences in Present-Day German. Evidence of an increased German tendency of using parataxis rather than hypotaxis for causal clauses has been observed both for popular science by Becher (2011) and for business articles by Bisiada (2013), and both in original texts and translations. Becher (2011: 199) explains this trend towards parataxis with reference to readability. In subordinate clauses German uses verb-final position, which has been found to lead to processing difficulties for readers, and therefore the V2 position of parataxis is increasingly being used. Examples of supplementive *ing*-clauses translated into German main clauses are seen in (1) and (11) above and (28) and (29) below. In (1) and (29) the main clause corresponding to the *ing*-clause is separated by a comma, in (11) by a full stop and in (28) by a semicolon.

There is a notable consistency in the German and Swedish translations in that coordination is the most common choice in both target languages. There appear to be two main reasons for the correlation between the English supplementive *ing*-clause and the German and Swedish coordination. Coordination is a compact structure allowing the omission of the subject and sometimes the auxiliary, thereby matching the subject-less non-finite supplementive *ing*-clause. Moreover, coordination is also often semantically indeterminate in much the same way as the source structure.[18] This is illustrated below in (27) where the Swedish rendering, just as the English original, is ambiguous at least between a temporal reading (simultaneous or succession) and specification (cf. Kortmann 1991: 121). The description of Gore's utterance, offering his blessing, can either be interpreted as him expressing his private opinion and then referring to what energy experts are saying, or as Gore offering his blessing by the very act of declaring that the experts are united in their assessments.

---

[18] Dirdal (2017: 216) found that Norwegian novice translators use coordination more often than professional translators as correspondences of supplementive *ing*-clauses.

(27) On a visit to Toronto, Al Gore offered his highest blessing, proclaiming it "widely recognized now as the single best green energy [program] on the North American continent."
Vid ett besök i Toronto gav Al Gore programmet sin välsignelse och utropade det till "nu erkänt i vida kretsar som det allra bästa [programmet] för grön energi på den nordamerikanska kontinenten."
"and proclaimed it "now widely recognized…""

The alluvial flow diagram in Figure 4 illustrates the (non-)congruency of the German and Swedish translations.
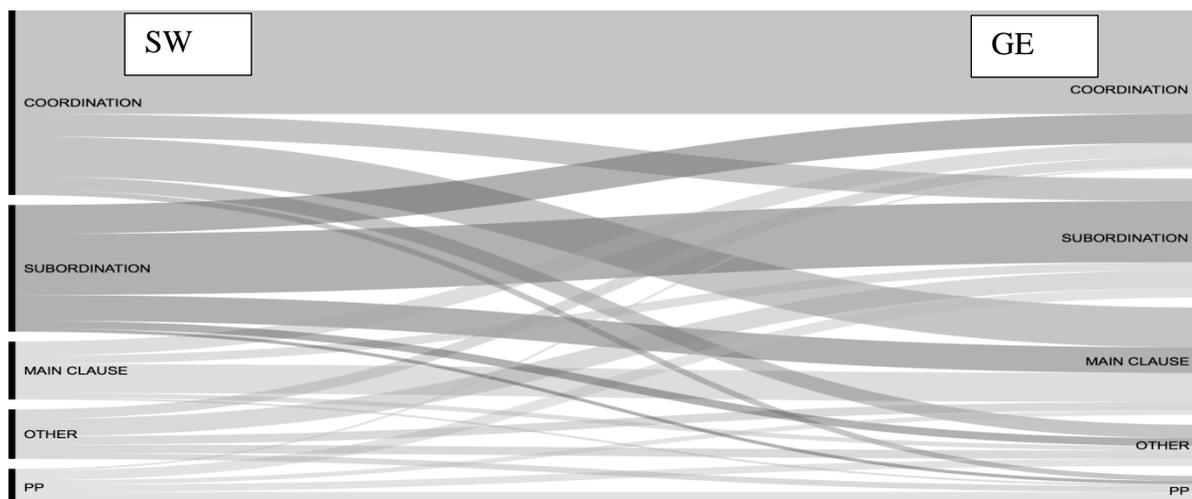

Figure 4. (Non-)congruency between German and Swedish translations of English supplementive *ing*-clauses.

There is a relatively high proportion of congruency between translations: 46.4% (329/709) of all supplementive *ing*-clauses are translated in the same way. Instances translated into coordination, subordination or PP in one target language are also fairly often translated into the same structure in the other.[19] This suggests that translators independently of their target languages often resort to similar choices when translating supplementive *ing*-clauses. Many of these choices involve clause building (Dirdal 2014) and more explicit structures. Dirdal (2014: 122) defines clause building as all changes moving towards independent main clauses. Thus, for instance, words or phrases rendered as clauses or non-finite clauses rendered as finite clauses are examples of clause building.

Some minor trends can be gleaned from the significantly greater preference for main clauses in German than in Swedish translations. From the 122 *ing*-clauses rendered as German main clauses while producing different Swedish translations (such as subordinate clauses) it appears that: (a) German translations more often than Swedish use semicolons to separate main clauses, (b) Swedish translators use certain subordinate clauses when German translators opt for main clauses, and (c) German sometimes uses main clauses linked by commas where the Swedish translators choose other strategies.

Firstly, German translators opt for main clauses separated by semicolons in 29 instances where the Swedish translators chose other options. These instances are distributed across translations of four out of five English original texts. A semicolon creates a clause

---

[19] Of the 709 instances, 161 (22.7%) are translated into coordination in both target languages, 95 (13.4%) are subordinate clauses, 45 (6.3%) are main clauses and 22 (3.1%) are translated into PPs in German and Swedish. In the Other category, 3 infinitives, 2 rephrased instances and 1 NP were translated in the same way.

boundary stronger than a comma but weaker than a full stop. This "intermediate" level of clause division, which is overall considerably more common in the German target texts,[20] would seem to facilitate the more frequent use of main clauses for translators not wishing to split up sentences. We here define sentence splitting more narrowly than Solfjeld (2008), restricting our definition to cases where supplementive *ing*-clauses correspond to separate main clauses, while Solfjeld (2008: 116) also includes different kinds of coordinations. The frequent German use of semicolons in the LEGS material may be a translation effect since semicolons are three times more common (15/10,000 words) in translations from English than in German originals. This issue, however, merits further study beyond the scope of this paper.

An example of a German translation with a main clause after a semicolon is given in (28). The Swedish translation illustrates the second tendency mentioned above, that of Swedish subordinate clauses when German translators choose main clauses. Sentential relative clauses (introduced in Swedish by the relative pronoun *vilket*; for similar Norwegian translation data, see Behrens 1998: 147) are more common in the Swedish than in the German translations (see below).[21] Sentential relatives are used 18 times in Swedish when the German translators make use of main clauses,[22] as illustrated in (28).

(28) The longer stress persists, the more those cognitive abilities are hurt and the more permanent the damage, *ultimately leading to mental as well as physical illness.*
Je länger der Stress anhält, umso stärker werden diese Fähigkeiten beeinträchtigt und umso dauerhafter ist die Schädigung; *dies führt letztlich ebenso zu psychischen wie zu körperlichen Erkrankungen.*
("this leads ultimately to…")
Ju längre stressen kvarstår, desto mer skadas de kognitiva förmågorna och desto mer permanent blir skadan, *vilket till sist leder till psykisk och fysisk ohälsa.*
("which ultimately leads to…")

Finally, there are 15 instances of main clauses linked by commas in German[23] among the examples of non-congruency. This is exemplified in (1) above and (29) below where the Swedish translators opt for coordination. In both examples, German and Swedish linguistic structure permit either translation method, but, as seen in Figure 3, there are different language-specific preferences which combine to create a significant difference between the target languages. The use of main clauses alters the discourse structure in that the subordinate clause is upgraded to a main clause and now forms an independent information unit (cf. Kortmann 1991: 113; Behrens and Fabricius-Hansen 2005: 111).

---

[20] Semicolons are three times more common in the German translations from English originals than in the Swedish ones. Notably, there are slightly more semicolons in the German translations than in the English originals.

[21] 54 supplementive *ing*-clauses are translated into sentential relative clauses in Swedish, and only 14 in German.

[22] In an additional 22 cases, the Swedish translators use other types of subordinate clauses, mostly adverbial ones.

[23] 58 German and 48 Swedish translations were linked asyndetically in the whole corpus.

(29) Unlike wasps or honeybees, most bumblebees don't even seem to mind very much if you poke around their nest, *stinging only as an absolute last resort.*
Anders als Wespen oder Honigbienen scheint es Hummeln nicht einmal sonderlich zu stören, wenn man in ihrem Nest herumstochert, *sie stechen wirklich nur im absoluten Notfall.*
"they sting really only…"
Till skillnad från getingar och bin verkar de flesta humlor inte ens bry sig särskilt mycket om ifall man rotar runt i boet, *och sticks bara i yttersta nödfall.*
"and sting only…"

The only notable difference in subordinate clauses in German and Swedish translations is, as mentioned above, that Swedish uses more sentential relative clauses than German.

Prepositional phrases are the final major category. Most involve similar prepositions in German and Swedish. These are three pairs of related prepositions *in/i* ('in'), *mit/med* ('with') and *mithilfe/med hjälp av* (lit. 'with the help of').[24] *In/i* mostly occurs in lexicalized complex prepositions, such as *im Gegensatz zu* ('in contrast to') or *i hopp om* ('in the hope of'). Swedish *med* has a slightly wider range of usage than German *mit*, for instance in some (semi-)lexicalized Swedish complex prepositions (e.g., *claiming it robbed the company of its right to* [...] translated as *med motiveringen att* ('with the motivation that') *det fråntog företaget dess rätt att* […].). *Mithilfe* and *med hjälp av* occur 19 times as correspondences of *ing*-clauses (one of which was found in a German source text and four in Swedish source texts). 18 of these *ing*-clauses are introduced by *using* (e.g., *using a beard of bristles on their mandibles* translated into *mithilfe der Borsten an den Mandibeln/med hjälp av skäggborst på käkarna*). This would seem to suggest that *using* is felt to be close to a preposition. Similar cases of preposition- and conjunction-like *ing*-forms in supplementive clauses are discussed by Visser (1972: 1218) (cited in Kortmann 1991: 191).

In summary, the translations into German and Swedish indicate fairly high degrees of correlations. Coordination is the most frequent alternative in both German and Swedish because of its indeterminate and compressed nature. The main difference between the target languages, i.e. the greater German use of main clauses, is probably a reflection of the ongoing German change towards parataxis identified by Becher (2011) and Bisiada (2013). So far the results have only concerned translations from English. Section 4.3.2 shows to what extent the translations into English produce similar findings.

4.3.2 *Comparisons with supplementive ing-clauses translated from German and Swedish*
Table 3 presents the German and Swedish source-text structures rendered as supplementive *ing*-clauses. As for the German and Swedish target-text structures in Table 2 above, coordination is by far the most common alternative. The order among the other alternatives, subordination, main clause, PP and Other, is slightly different and the frequencies are more equal than for the target-text structures. The Other category is slightly larger in the German and Swedish source texts than in the target texts.

---

[24] There are 12 *in*, 10 *mit* and 6 *mithilfe* in German (of 47 instances), and 13 *i*, 22 *med* and 8 *med hjälp av* in Swedish translations (of 58 instances).

Table 3. German and Swedish source-text structures translated into supplementive *ing*-clauses.

	German source-text structures		Swedish source-text structures	
	N	%	N	%
coordination	149	32.7	74	38.6
subordination	91	20.0	31	16.1
PP	69	15.1	32	16.7
main clause	69	15.1	24	12.5
Other	*78*	*17.1*	*31*	*16.1*
infinitive	17	3.7	10	5.2
NP	13	2.9	5	2.6
participle	14	3.1	9	4.7
rephrased	11	2.4	1	0.5
VP	12	2.6	0	0
addition	5	1.1	1	0.5
advP	4	0.9	2	1.0
adjP	2	0.4	3	1.6
Total	**456**	**100**	**192**	**100**

The differences between the German and Swedish source-text structures in Table 3 are not significant. As is evident from Figure 5 below, however, there are a number of significant differences between the *ing*-clause correspondences in the source texts and the target texts in both German and Swedish.
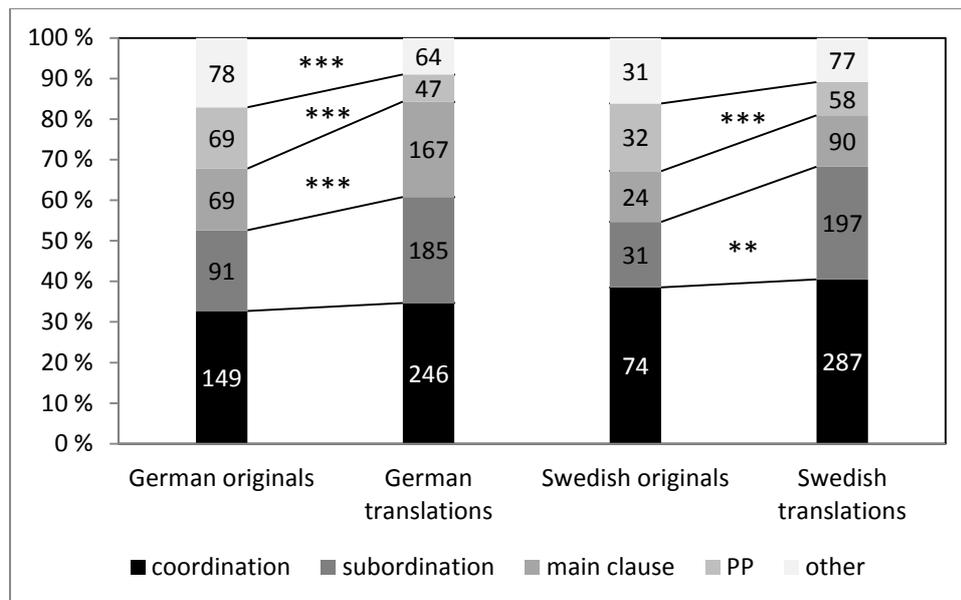

Figure 5. The correlations between the German and Swedish source-text and target-text correspondences of supplementive *ing*-clauses.

Three out of five categories produce significant differences between German source texts and target texts,[25] and two out five for the Swedish texts.[26] It is nevertheless noteworthy that the largest category, coordination,[27] does not produce any significant differences in either German or Swedish (see discussion above in Section 4.3.1). This means that it is equally likely for a translator into English to choose a supplementive *ing*-clause when translating a

---

[25] According to a chi square test with a Bonferroni correction; df=4, X2=49.47, p=***.

[26] According to a chi square test with a Bonferroni correction; df=4, X2=22.65, p=***.

[27] Cosme (2008) found coordination to be a frequent correspondence of *ing*-clauses also in Dutch translations.

coordination like in (30) below, as it is for a translator from English to choose a coordination when translating a supplementive *ing*-clause, as seen in, e.g., (13).

(30) Stiegs mormor följde med och bodde tillsammans med dem under den första tiden. (LEGS; SW ST)
"Stieg's maternal grandmother went too and lived together with them during the initial period"
His grandmother went too, initially also *living with them*.

There is a significant difference for PPs in both German and Swedish, PPs being more frequent in originals.[28] There is also a significant difference for subordinate clauses in Swedish. This is a result of all subtypes of subordinate clauses (adverbials, sentential and post-modifying relative clauses) being rarer in the originals than in the translations of *ing*-clauses. The reason is probably that most of these subordinate clauses in Swedish source texts can be conveniently translated using the same structures in English, e.g., subordinate clauses introduced by *when* or *since* corresponding to the frequent Swedish *när/eftersom*. The less explicit supplementive *ing*-clauses less readily emerge as alternatives in these cases. An example of a Swedish temporal adverbial clause translated into a supplementive *ing*-clause is seen in (31). A temporal clause similar to the English gloss would have been an acceptable alternative translation. Supplementive *ing*-clauses nevertheless provide advantages for translators into English. By using a supplementive *ing*-clause the translator avoids having to repeat the same referent twice (*she*). This strategy is used at the expense of decreased explicitness in the link between the subordinate clause and main clause.

(31) *När hon på kvällen kom fram till S:t Göran* fick hon det chockartade beskedet. (LEGS; SW ST)
"when she that evening arrived at the S:t Göran [hospital] she got the devastating news"
Arriving at the hospital that evening, she was given the devastating news.

The greater use of prepositional phrases in German and Swedish source texts than in German and Swedish translations from English is difficult to explain in terms of translation strategies and general language-specific preferences, even though there is a consistency in that both German and Swedish source texts contain more prepositional phrases than the German and Swedish target texts. The numbers are low and the individual source texts seem to have a strong influence on the outcome.

Figure 5 shows that the high frequency of main clauses in German target texts produces a significant difference in comparison with German source texts. Using Dirdal's (2014) terminology, it is thus more likely that translators into German 'build' main clauses from supplementive *ing*-clauses than English translators 'reduce' German main clauses to supplementive *ing*-clauses. *Ing*-clauses also seem to be the most frequent source in English-to-Norwegian clause building (2014: 127), but these *ing*-clauses are neither clearly defined nor analyzed in detail by Dirdal. The greater tendency for sentence building from supplementive *ing*-clauses in German is perhaps unexpected in view of Solfjeld's (2008) suggestion that information density in German non-fiction leads to a high degree of sentence splitting in translations into Norwegian. The reason for our divergent finding is probably our restriction to a single English construction lacking a German counterpart. It seems reasonable

---

[28] According to a chi square test with a Bonferroni correction; German df=4, X2=21.44, p=***; Swedish df=4, X2=11.81, p=***.

that it is easier to divide information-dense non-fiction sentences into two than merging two into one. It nevertheless seems that there is a language-specific preference for more clause building in the German than in the English texts in the LEGS corpus (apart from the greater use of semicolons mentioned above): the five German translated texts contain more full stops (+9%) than their English originals, while English translations from German instead contain fewer full stops (−8%) than their German originals (other punctuation marks being very rare).

Below, in example (32), is one of the fairly rare instances of two German main clauses being reduced to a main clause and a supplementive *ing*-clause. The translation avoids mentioning the subject (*Merkel/she*) twice (as also seen in (31)), and condenses (in (32) by eliding the subject *sie* and the adverbial *lieber*, while nevertheless adding the verb *choosing*) and backgrounds the information from the second main clause into the subordinate clause (cf. Kortmann 1991: 113) while maintaining the order of the direct object preceding the dative object / prepositional object.

(32) Staatsbesuche absolviert Merkel überhaupt nur in begrenzter Zahl. *Diese höchsten protokollarischen Ehren überlässt sie lieber dem Bundespräsidenten.* (LEGS; GE ST)
"these highest honours she rather leaves to the President"
Merkel makes only a small number of state visits, *choosing to leave this highest honour to the President of the Republic.*

The final significant difference between German source texts and target texts relates to the greater proportion of other equivalents in German originals. This is an effect of most of these minor categories (infinitive, NP, participle, rephrased, VP, adjP and advP) combining to increase the frequency in originals with only one (omission/addition) marginally going against the trend.

4.3.3 *Explicitation*
As seen above, some of the structures used to translate the often semantically indeterminate supplementive *ing*-clauses involve more explicit structures in German and Swedish. This is in line with translations typically being more explicit than their originals, as suggested by Baker (1996: 180). In the following discussion of explicitation we will include conjunctions, adverbials, German *um zu* ('in order to') infinitives and Swedish infinitives consisting of prepositions together with the infinitive marker *att* as explicitation devices (for explicitation in translations of *ing*-clauses into Norwegian, see Behrens 1998: 259ff.).

Table 4 presents the explicitations occurring more than ten times in each target language as well as the number of other structures found.

Table 4. Explicitations in German and Swedish target texts from English originals.

Translations into German		Translations into Swedish	
als ('when')	16	*när* ('when')	23
da ('since')	16	*eftersom* ('because')	18
dabei ('thereby/at the same time')	11	*där* ('where')	16
indem ('while/by')	11	*efter att* ('after that')	14
so dass/sodass ('so that')	11	*så att* ('so that')	13
		därmed ('thereby/thus')	12
other	124	other	58
Total	**189**	**Total**	**154**

The Swedish translations contain fewer instances of explicitation. There are more instances and more variation in German translations which produce a large proportion of 'other'

explicitation strategies.[29] The most frequent explicitations involve temporal and causal conjunctions in both languages, as exemplified in (2), (14) and (25) above. The higher proportion of explicitations in German is largely due to the use of pronominal adverbs, which come in many different forms and express various relations. Kortmann (1991: 110, 122) notes that English lacks conjunctions expressing instrument (such as German *indem*), accompanying circumstance (*wobei*) and those making two events form a unit (*und dabei*, *wobei*). Thus, an English structure largely absent in German is sometimes translated by a German class of words not occurring in English. Apart from the frequent *dabei* and *indem* listed in Table 4,[30] other examples include *dadurch* ('thus/thereby'), *womit* ('whereby') and, as in (33), *nachdem* ('after'). The Swedish translator uses a different type of explicitation: a preposition together with the infinitive marker *att*, expressing a similar meaning to the German. This is one of the 64 instances explicitated in both target languages.[31]

(33) *Having studied their gardens,* these volunteers were asked to repeat the exercise in one countryside habitat, chosen at random from a range of options.
Nachdem sie dieses Experiment in ihrem Garten durchgeführt hatten, wurden die Freiwilligen gebeten, es in einem ländlichen Habitat zu wiederholen, das nach dem Zufallsprinzip ausgewählt wurde.
"after they had performed this experiment in their garden"
Efter att ha studerat trädgården ombads de frivilliga att upprepa övningen vid ett av olika alternativ slumpmässigt valt förekomstställe på landsbygden.
"after INF. have studied the garden"

Our findings suggest that there are partial overlaps between the most frequent German and Swedish lexical manifestations of explicitation (see Table 4). However the German preference for various pronominal adverbs also produces notable differences between the target languages.

5.   Conclusions

This study has shown that there are both similarities and differences between the German and Swedish correspondences of English supplementive *ing*-clauses. Coordination is the most frequent correspondence in translations both into and from German and Swedish. This seems to be due to the fact that coordination, just as the supplementive *ing*-clause, are compact and semantically rather indeterminate. Coordination occurs in three to four out of ten instances in both languages, which means that the correlation between this German and Swedish construction and the English supplementive *ing*-clause is not very strong, however. The second most common correspondence, subordination, backgrounds the subordinate clause to the main clause, similar to what the *ing*-clause does. Main clauses and prepositions are the final major alternatives in both German and Swedish. Main clauses are a significantly more common alternative in German translations than in Swedish ones. This is probably an effect

---

[29] Explicitation is used in 26.7% (189/709) of the German and 21.7% (154/709) in the Swedish translations. There is a marginal statistical significance for a greater German preference for explicitation (chi-square test: p = 0.035; phi coefficient = 0.058).

[30] Behrens and Fabricius-Hansen (2005: 4) write that *dabei* is used "quite often" in German translations of *ing*-clauses expressing 'accompanying circumstance', but do not provide quantitative support for this claim.

[31] This means that 33.9% of the 189 explicitated German instances are also explicitated in the Swedish translations, and that, conversely, 41.6% (64/154) of the explicitated Swedish instances are also explicitated in the German translations.

of an ongoing shift in German from hypotaxis to parataxis, as noted by Becher (2011) and Bisiada (2013). Like coordination, a prepositional phrase is a compressed alternative. They often occur as correspondences of *ing*-forms that appear to be preposition-like, such as *using*. Among the minor alternatives, it can be noted that participles only occur as correspondences in 2% of the instances, which shows that German and Swedish participles are not close equivalents of the English supplementive *ing*-clause.

In view of the semantically indeterminate nature of supplementive *ing*-clauses, it is noteworthy that a number of translations contain overt explicitation markers. The most common types involve subordinators such as *als/när* ('when'), *da/eftersom* ('since/because') and German pronominal adverbs (e.g., *dabei*, *indem*). German translations produce slightly more explicitations than the Swedish ones, possibly because of the wide range of German pronominal adverbs. It is nevertheless evident that more in-depth analyses are needed on the relation between translations and the semantics of supplementive *ing*-clauses.

One considerable advantage with the LEGS corpus is that it makes it possible to compare each source text with two target languages (cf. Egan 2016). This is particularly fruitful when comparing translations of a structure lacking a productive equivalent in more than one target language. Although English supplementive *ing*-clauses are semantically indeterminate, the German and Swedish translators in this study choose the same translation solutions for almost half the English source-text instances. This suggests that there is systematicity in the translation choices which can be explained both by the target languages being closely related structurally and by the source-text structure steering the translators in specific directions. This is illustrated by the fact that only four translation categories (coordination, subordination, main clause and PP) account for 90% of all translations into both German and Swedish.

At this stage the LEGS corpus is fairly restricted as regards the number of texts. Still it is large enough to produce adequate numbers of instances of medium-frequency phenomena such as supplementive *ing*-clauses, which previously have only been studied in relatively small and partly opportunistically collected data-sets.

References

Baker, M. 1993. Corpus linguistics and translation studies: implications and applications. In *Text and Technology: In Honour of John Sinclair*, M. Baker, G. Francis and E. Tognini-Bonelli (eds), 233–250. Amsterdam and Philadelphia: John Benjamins.

Baker, M. 1996. Corpus-based translation studies: the challenges that lie ahead. In *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*, H. Somers (ed.), 175–186. Amsterdam and Philadelphia: John Benjamins.

Becher, V. 2011. Von der Hypotaxe zur Parataxe: Ein Wandel im Ausdruck von Konzessivität in neueren populärwissenschaftlichen Texten. In *Satzverknüpfungen. Zur Interaktion von Form, Bedeutung und Diskursfunktion*, E. Breindl, G. Ferraresi and A. Volodina (eds), 181–209. Berlin: de Gruyter.

Behrens, B. 1998. *Contrastive Discourse: An Interlingual Approach to the Interpretation and Translation of free ING-participial Adjuncts*. PhD Thesis, Oslo.

Behrens, B. and Fabricius-Hansen, C. 2005. The relation accompanying circumstance across languages. Conflict between linguistic expression and discourse subordination?. *SPRIKreports* no. 32. `https://folk.uio.no/cfhansen/cfhAccCicumst10.pdf` [last accessed 9 October 2017]

Behrens, B. and Solfjeld, K. 2014. Discourse role guiding clause type and position in translation. In *Adverbials in Use: From Predicative to Discourse Functions*, L. Sarda, S. Carter-Thomas, B. Fagard and M. Charolles (eds), 267–296. Louvain-la-Neuve: Presses universitaires de Louvain.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bisiada, M. 2013. Changing conventions in German causal clause complexes. A diachronic corpus study of translated and non-translated business articles. *Languages in Contrast* 13(1), 1–27.

Blensenius, K. 2006. *Particip med andra ord. En korpusstudie av svenska motsvarigheter till engelska ing-satser*. Göteborg: Göteborgs universitet, Institutionen för svenska språket.

Blum-Kulka, S. 2004 [1986]. Shifts of cohesion and coherence in translation. In *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition*, J. House and S. Blum-Kulka (eds), 17–35. Tübingen: Narr. Reprinted in *The Translation Studies Reader*, 2nd ed, L. Venuti, (ed.), 290–305. London: Routledge.

Chesterman, A. 2004. Hypotheses about translation universals. In *Claims, Changes and Challenges in Translation Studies*, G. Hansen, K. Malmkjaer and D. Gile (eds), 1–14. Amsterdam and Philadelphia: John Benjamins.

Cosme, C. 2008. A corpus-based perspective on clause linking patterns in English, French and Dutch. In *'Subordination' vs. 'Coordination' in Sentence and Text: A Cross-linguistic Perspective*, C. Fabricius-Hansen and W. Ramm (eds), 89–114. Amsterdam: John Benjamins.

Dirdal, H. 2014. Individual variation between translators in the use of clause building and clause reduction. In *Corpus-based Studies in Contrastive Linguistics*, S. O. Ebeling, A. Grønn, K. R. Hauge and D. Santos (eds). *Oslo Studies in Language* 6(1), 119–142.

Dirdal, H. 2017. Student translators and the challenge of ing-clauses. In *Language, Learners and Levels: Progression and Variation*, P. de Haan, S. van Vuuren and R. de Vries (eds), 203–225. Louvain-la-Neuve: Presses universitaires de Louvain.

Egan, T. 2016. Contrasting translations. *Nordic Journal of English Studies* 15(3), 80–101.

Fischer, K. 2013. *Satzstrukturen im Deutschen und Englischen. Typologie und Textrealisierung*. Berlin: Akademie (Konvergenz und Divergenz 1).

Hasselgård, H. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press.

Johansson, S. 2002–2003. Viewing languages through multilingual corpora, with special reference to the generic person in English, German, and Norwegian. *Languages in Contrast* 4(2), 261–280.

Kortmann, B. 1991. *Free Adjuncts and Absolutes in English*. London: Routledge.

Lindquist, H. 1989. *English Adverbials in Translation: A Corpus Study of Swedish Renderings.* Lund Studies in English 80. Lund: Lund University Press.

Malá, M. 2005. Semantic roles of adverbial participial clauses. In *Theory and Practice in English Studies 3: Proceedings from the Eighth Conference of British, American and Canadian Studies*. Brno: Masarykova univerzita. 91–97.

Nida, E. A. 1964. *Toward a Science of Translating*. Leiden: E. J. Brill.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Solfjeld, K. 2008. Sentence splitting and discourse structure in translations. *Languages in Contrast* 8(1), 21–46.

Stump, G.T. 1985. *The Semantic Variability of Absolute Constructions.* Dordrecht: Reidel.

Visser, F. Th. 1972. *An Historical Syntax of the English Language*. Part III/2. Leiden: Brill.

Authors' address

Jenny Ström Herold / Magnus Levin
Department of Languages
Linnaeus University
Universitetsplatsen 1
351 95 Växjö
Sweden

jenny.strom.herold@lnu.se / magnus.levin@lnu.se

The Swedish modal auxiliary *ska/skall* seen through its English translations

Karin Aijmer

University of Gothenburg (Sweden)

Abstract: The Swedish modal auxiliary *ska/ll* is interesting typologically because it is highly multifunctional with different functions in the systems of tense and modality. The aim of the paper is to investigate the translations of *ska/ll* into English as a methodology to throw light on its multifunctionality and the linguistic and extralinguistic factors determining its usage as a marker of obligation. The translations with *must* and the imperative point to the performative uses of *ska/ll* such as commands and orders. However, the translations also show that *ska/ll* is involved in offers, recommendations, suggestions, advice where the strength of imposition has been weakened. Other translations reveal more unexpected uses of *ska/ll*, for example to express negative evaluation if the context of the utterance could be interpreted as emotional or argumentative. The translations also draw attention to linguistic and non-linguistic factors aiding to distinguish functions of *ska/ll (*or choosing a particular translational correspondence).

Keywords: *ska*, *skall*, obligation, multifunctionality, parallel corpus, English, Swedish

1. Introduction

Polysemy can be regarded as "an outstanding feature of modal verbs in the Germanic languages" (Narrog 2016: 98). The Swedish modal auxiliary *ska/ll* (English *shall,* German *sollen*, Dutch *zullen*) is, for example, interesting typologically because it is highly multifunctional with different functions in the systems of tense and modality.[1] The Swedish reference grammar (Teleman *et al.* 1999: 312) describes *ska/ll* as having a temporal, deontic and epistemic meaning. The meanings are illustrated with their translations into English in examples (1)–(3).

(1) Det <u>ska</u> bli mitt livs verk. (GT1)
This <u>will</u> be my life's work.

(2) Du <u>ska</u> hem och äta! (ARP1)
You <u>must</u> go back and eat.

---

[1] The notation *ska/ll* is used to signal that the form *ska* includes the more formal and conservative *skall*.

(3)   Jo, men han lär <u>ska</u> ha en pojkbyting i alla fall. (AL1)
        Well, but <u>they do say</u> he has a boy, all the same.

In example (1) the reading of *ska* (like English *will*) refers to the future. However, in (2) *ska* is translated with *must* and refers to an obligation imposed on the hearer most likely by the speaker. *Ska/ll* can also refer to what people say as evidence for what is claimed, as in (3) (a type of evidential modality), in which case it is more closely related to epistemic (modal) than deontic meaning.

Modal auxiliaries such as Swedish *ska/ll* and its correspondences in English have generally been analysed from a semantic point of view. In the literature on modality they are, for example, associated with such properties as necessity, moral desirability, subjectivity. According to Teleman *et al.* (1999: 316), *ska/ll* suggests that somebody demands a certain action by someone who is not expected to carry out the action on his or her own initiative (see example 2). The demand can be anchored in a social or functional norm. On the other hand, the categories needed to analyse the functions of modal auxiliaries may be different from those focusing on their semantic properties. In discourse *ska/ll* has for example functions to perform specific speech acts as well as other, less obvious uses.

The present study sets out to investigate the translations of *ska/ll* into English as a methodology to throw light on its multifunctionality and the linguistic and extralinguistic factors determining its usage. The paper takes the viewpoint that the different meanings of *ska/ll* and their frequencies can be described in terms of the well-known phenomenon of grammaticalization described for example by Hopper and Traugott (1993: xv): "the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions" (quoted from Leech 2004: 75). The grammaticalization of *ska/ll* involves the use both as a modal auxiliary and as a future marker. However the present study is modest in its scope since the focus is on *ska/ll* as a modal marker. The study includes the present forms *ska/ll*[2] and not the preterite form *skulle* (*should*) since *skulle* can be assumed to have a different interactional profile. The research questions are:

- What are the functions of *ska/ll* revealed by the translations into English and what do the translations tell us about the polysemy and multifunctionality of *ska/ll*?

- What are the linguistic and extra-linguistic factors or parameters determining how *ska/ll* has been translated (and indirectly what uses it has)?

- How can we explain the translation paradigms in terms of grammaticalization and ongoing semantic changes?

The paper is organised as follows. Section 2 gives an overview of research on factors playing a role for the uses of modal markers of obligation. The methodology and data are discussed in Section 3, and Section 4 presents the translations of *ska/ll* into English. Section 5 suggests a categorization of *ska/ll* based on the translation data. Section 6 is the conclusion.

---

[2] However, Hilpert (2006) found that *ska* and *skall* differed substantially in their collocational patterns, which suggests that they are associated with different meanings.

2.   A review of relevant research on modal markers of obligation

2.1   Swedish *ska/ll*

Previous work may be an aid to suggest factors determining how *ska/ll* has been translated. According to Teleman *et al.* (1994: 316), *skall* is used in sentences where someone or something demands a certain action from someone (or some body or institution). The requirement can be anchored in a functional or social norm (such as duty, custom, what is appropriate). It is also pointed out that *skall* often refers more directly to power. However, Teleman *et al.* do not discuss how the source of imposition (or other factors) can distinguish between the different modal meanings.

Hilpert (2006, 2008) has studied the grammaticalization of *ska* and *skall* with future meaning in a synchronic perspective. In a corpus study he examined what verbs were 'significantly attracted' to *ska* or *skall* in future and obligation meanings. The investigation showed that out of the three verbs which were most distinctive for *ska* (*bli* 'become', *ske* 'happen' and *göra* 'do') *bli* and *ske* strongly preferred inanimate subjects thus ruling out other meanings such as obligation (or intention). However, other verbs are strongly linked to *ska* with obligation meaning. Hilpert provides the following example with *betala* 'pay' where *ska* = 'must':

> På något sätt ska de betala för vad de har gjort
> *on some way must they pay for what they have done*
> 'In some way they will have to pay for what they have done' (Hilpert 2006: 158)

Hilpert's study is interesting for the present study because it suggests that the meaning of the infinitival verb is important for how *ska/ll* is translated. Another result is that the item which should be the basis for the semantic description may be *ska/ll* followed by the infinitival verb and not the independent item *ska/ll*.

2.2   English *should, must, have to*

There is a rich literature on the linguistic and extra-linguistic factors playing a role for describing the uses of the modal markers of obligation in English. The most frequent obligation markers are *should, must, have to*. They have a large number of different meanings, which makes it difficult to categorize them. A basic factor is whether the modal markers of obligation are used performatively to indicate speech acts. This function is particularly associated with *must*, which is used for orders and commands, although, as pointed out by Narrog (2016: 111), basically all modal markers can have a performative function under the right conditions.

The study of obligation markers in English has also been influenced by recent research on the semantic changes they have undergone. On the basis of matching corpora from the 1960s and the 1990s it has been reported that *must*, which has associations with direct speaker authority or power, is being replaced by *have to* where the source of imposition is external or objective (Smith 2003: 242; Leech 2004). The avoidance of an overt expression of power can also be a factor which can be studied in the ways *ska/ll* has been translated.

Many linguists have drawn attention to the factors strength and the source of imposition. However, when Myhill and Smith (1995) studied the discourse functions of obligation markers in different languages they found it difficult to apply these criteria to their data. A more useful criterion was based on the observation that obligation markers often expressed (negative) evaluation or affectedness usually on the part of the hearer. They

describe the following example of *have to* as an interesting non-obligation use of *have to* illustrating the function of negative evaluation:

> Why did he have to go out? Why didn't someone hearer him… Why did he have to hear noises in the night? (Myhill and Smith 1995: 248; simplified example).

Should as a modal auxiliary of obligation has been discussed less than *must* and *have to*. Verhulst and Heyvaert (2015) have described the difference between *should* and *ought to* in American English. They draw attention to the use of *should* to express the speaker's personal opinion in contexts of disagreement. For instance, in the example below, the speaker (a mother) is criticizing her daughter for going out without permission:

> You shouldn't go out without telling us, Marilyn (Verhulst and Heyvaert 2015: 573).

3.   Methodology

Translations are suitable for a usage-based or functional approach to language since they show what translators choose as the equivalent on the basis of a general idea of what is meant by the source item, the larger discourse context, knowledge of genre conventions, etc. The methodological assumption behind this type of study is that translations into a different language provide a window into its multifunctionality (see Ivir 1981, 1987 and the discussion by Altenberg and Granger (2002)). This idea has been further developed by Dyvik (1998, 2004) who argues that the ambiguity or vagueness of multifunctional items are reflected in their translations into another language and that it is possible to establish a function or sub-function on the basis of how the item is interpreted by the translator. In this perspective, a 'translational paradigm' showing how translators have interpreted *ska/ll* can be argued to give a more objective picture of its multifunctionality. Dyvik demonstrated how sense distinctions can be made by establishing translation correspondences starting both from originals and translations. However *ska/ll* has no obvious correspondence in the target language which could be used as the starting-point for examining the translations in the reverse direction. The translation correspondences used for establishing sense distinctions therefore represent translations in one direction only.

Crucial for the contrastive analysis is the availability of parallel corpora. The translations of *ska/ll* are taken from the English-Swedish Parallel Corpus (ESPC) (Altenberg and Aijmer 2000, Altenberg *et al.* 2001). The ESPC is a bidirectional corpus (with translations from English into Swedish and vice versa) consisting of roughly 3 million words of both fiction and non-fiction texts. For the present study I used translations from Swedish into English and restricted myself to fiction (1.5 million words).

The procedure followed in this study has been to extract the sentences with *ska/ll* together with their translations into English. Depending on the translational counterparts of *ska/ll* we can categorize the meanings of *ska/ll* as belonging in the domains of future and obligation (and epistemic modality/evidentiality). The translations can, however, also be assumed to give rise to a more detailed description of what *ska/ll* is doing in the discourse in different domains.

4.   The translations

4.1   Overview

The translations provide a fine-grained analysis of the different meanings or functions of *ska/ll.* The translations in Table 1 have been grouped into three major categories reflecting their uses to refer to future time, to mark obligation and evidential meaning. The examples where the translation is idiosyncratic or too far removed from the original have not been classified ('other').

Table 1. The translations of *ska/ll* in different meanings in the ESPC.

	Translation	ska	skall	Total
Future	will	129	39	168
	be going to	55	16	71
	shall	33	8	41
	present tense of the main verb[3]	9	-	9
	be+-ing	8	1	9
	be about to	6	-	6
	infinitive[4]	26	12	38
	be set on	1	-	1
	be due to	1	-	1
	intend to	1	-	1
	begin to	1	-	1
	Sub-total	**270**	**76**	**346**
Deontic/ Obligation	should	59	12	71
	must	20	4	24
	have/has to	20	2	22
	be to	11	16	27
	be supposed to	11	5	16
	imperative[5]	11	3	14
	wouldn't you like, (do) you/they want	7	3	10
	may, might	6	-	6
	need	1	1	2
	can	11	6	17
	I want to/I should like	4	6	10
	be meant to	3	-	3
	(perhaps) would	1	1	2
	have got to	1	-	1
	be allowed to	1	-	1
	indeed	1	-	1
	Sub-total	**168**	**59**	**227**
Evidential	be supposed to	6	-	6
	they do say	1	-	1
	Sub-total	**7**	**-**	**7**
Other		48	16	64

[3] For example: *När ska vi segla? – When do we sail?* When the verb is in the present tense it is strictly not a translation of *ska* but of the whole verb phrase where *ska* occurs.
[4] An example is: *När vill du att jag ska komma?* (lit. "When do you want that I shall come") , translated as *When do you want me to come?* (BL1). The infinitive does not translate *ska* but *ska* followed by the main verb.
[5] The imperative translates the whole verb phrase rather than *ska* alone.

Karin Aijmer

4.2   Discussing the translations

Ska/ll only rarely corresponds to its semantic and formal cognate *shall* in the English translation. Instead it has a wide range of translations rendering aspects of its meaning. The translations into English give a rough picture of the polysemy of *ska/ll* and the parameters motivating the different meanings. In other words, "the one-to-many relationships among formal correspondents in pairs of contrasted languages is merely a reflection of similarity relationships between meanings and linguistic units in single languages" (Ivir 1981: 56). We can distinguish between translations where *ska/ll* has deontic or obligation meaning and examples with future or epistemic reference. In the obligation meaning the most frequent translations were modal auxiliaries (*should, must, can may, might, need, would*) or semi-modals: *have to, have got to, be to,*[6] *be supposed to, be meant to*.[7] The translators have also used the imperative to translate *ska/ll* or a construction with a lexical verb (*want*).

Deontic modality is closely related to future meaning and it is challenging to keep the meanings apart. This is illustrated by *shall* in the translation which has not lost its obligation meaning when it refers to the future.[8] *Ska/ll* has been regarded as having future meaning when it was translated by *shall, will, be going to, be+ing* form (e.g., *the sign is coming down*), *be about to, be due to, be set on*, the use of the present tense of the main verb, *intend to, begin to*. The meanings include prediction (*will*), intention (*I'm going to*), ingressive aspect (*begin to*), scheduled future event (present tense), an event expected to take place (*be due to*). The frequency of *will* is very high followed by *be going to* (see Hilpert 2008: 50 on different future interpretations).

The translations provide interesting information both about the frequency of different future meanings and the factors which are relevant when we talk about events in the future. The future meaning of *ska/ll* dominates (53.73%) as reflected by the high frequency of *will* with future meaning (see Table 2). The majority of other meanings fall into the area of deontic or obligation modality. On the other hand, the evidential meaning is infrequent. It is represented by *they do say* and *it is supposed* in the translations.

Table 2. The frequency of future meaning, obligation (deontic) meaning and evidential meaning.

Future	346 (53.73%)
Obligation	227 (35.23%)
Epistemic/evidential	7 (1.19%)
Other	64 (9.94%)
Total	644 (100%)

5.   Categorizing the functions of *ska/ll* meaning obligation

The obligation meaning has been rendered in different ways in the translations suggesting that *ska/ll* has more than one meaning. The most frequent translations were *should, must, have to* and *be to* (see Table 1).

---

[6] With some hesitation I regarded *be to* as mainly having obligation meanings. According to Collins, *be to* typically has both modal uses and uses which "are more temporal than modal" Collins (2009: 85).
[7] *Must, have to* and *be supposed* to also have modal meanings (inference). These are generally easy to distinguish from the deontic or obligation meaning.
[8] Thus, Collins (2009: 137) refers to different uses of the deontic *shall*.

144

5.1 *Ska/ll* translated as *should*

Obligation *ska/ll* was generally translated as *should* rather than *shall*[9] (in the meaning obligation).[10] *Should* was more frequent as a translation than *must* or *have to* (which have a strong deontic meaning). Several functions can be distinguished.

5.1.1 Should *with the function to convey the speaker's personal opinion or wishes*
The source of the imposition is the speaker as suggested by combinations such as *I guess I should, perhaps we should, you don't think we should* to make a suggestion or give a piece of advice.

Should was found in a subordinate clause under *I guess* to express the speaker's personal opinion, as shown in (4). *I guess* has the function to weaken the obligation meaning. The speaker wants to avoid threatening the hearer's face by mitigating his/her opinion.

(4)   - En sån hamnar aldrig på mitt huvud, sa Wallander. Man <u>kanske ska</u> vara glad över att man inte längre är ordningspolis. (HM1)
        Lit. "One perhaps should be glad…"
        "<u>I guess I should</u> be glad I 'm not a cop in uniform anymore."

Perhaps we should in (5) and (6) presents what is said as a tentative suggestion.

(5)   Men <u>vi ska kanske</u> akta oss lite för att ge den här mannen ett råd. (SC1)
        <u>Perhaps we should</u> be rather careful about giving this man advice.

(6)   — <u>Kanske ska vi</u> hoppa över bakgrunden till dopnamnet och bara tala om vad han heter-? (ARP1)
        "<u>Perhaps we should</u> skip the background of his baptism and just say what his name is..."

In (7) *perhaps* translates *eller* ('or'). The function is to recommend an action or make a polite suggestion:

(7)   <u>Eller ska</u> vi, som ännu inte gjort vårt, ta och lösa problemet åt honom? (SC1)
        <u>Perhaps</u> we who are not yet old age pensioners <u>should</u> solve it for him?

Speakers avoid a direct expression of obligation unless the context is argumentative. In (8) the translation contains a subordinate clause with *should* under the polite formula *you don't think*. The speaker is making a polite suggestion that it would be better for the hearer to wait until the rain stops.

---

[9] The translations with *shall* in my data are mainly found in questions. Such examples are treated as deontic by Collins (2009:137) For example: "Vill du jag <u>ska</u> ta henne nu?" (AL1) with the translation "<u>Shall</u> I take her now?" However, a larger number of translations with deontic *shall* would be found in non-fiction (especially legal texts).
[10] On the other hand, there was no example of *ought to* in the translations. However this may be expected against the background that *ought to* has declined generally in English (Leech 2004: 70) or that *ska* has a different meaning than *ought to*.

(8) Det är värst vad det regnar, <u>ni ska inte vänta</u> tills...
— Det kommer att ge sig, sa Aron. (GT1)
Boy, is it raining! <u>You do n't think you should wait</u> until..."
"It 's going to stop," said Aron.

Should in the translation was also found in interrogative sentences where the question form and the negation are strategies mitigating the imposition on the hearer by giving the hearer the option not to follow the suggestion:

(9) - <u>Ska</u> vi <u>inte</u> ta ut några patrullbilar? undrade Martinson. (HM2)
"<u>Should n't</u> we take a few patrol cars?" wondered Martinson.

5.1.2 Should *in contexts of disagreement*
When *should* is used in the translation of *ska/ll* this raises the question whether the context is argumentative. *Should* has the interpersonal function to take up a challenging position to the hearer or what is said in contexts where no agreement is presupposed (Verhulst and Heyvaert 2015: 573). In (10) the context suggests that the hearer disagrees with the speaker's opinion.

(10) - Jag tänkte att du borde veta om det, sa han. Förstår du vad det är som håller på att hända med polisen?
- Nej, sa Wallander. Och det <u>ska</u> du uppfatta som ett både uppriktigt och uttömmande svar. (HM1)
"I thought you ought to know about it," he said.
"Do you have any idea what 's going to happen to the police force?"
"No," said Wallander. "And you <u>should</u> take that as the complete truth."

The factor involved in the choice of *should* in the translation is the different opinions of people discussing what is going to happen to the police force.
   In (11), *should* is motivated by the argumentative context and by the emphasis on what the speaker has always said.

(11) Barn <u>ska</u> vara barn, det har jag alltid sagt. Låt dom behålla oskuldens vita änglavingar så länge som möjligt. (ARP1)
"Children <u>should</u> be children, I 've always said. Let them keep their white angel wings of innocence as long as possible.

Should in the translation indicates that the norm is what the speaker thinks (rather than some general recommendation).

(12) Om det är en verkligt fin årgång är det ett helgerån att hälla upp sherryn på karaff — det är ju rena mordet på bouqeten. En gammal fin sherry <u>ska</u> serveras direkt ur buteljen. (ARP1)
"If it 's really a genuinely fine vintage, then it 's blasphemy to decant it into a decanter — that 's straight murder of the bouquet. A fine old sherry <u>should</u> be served straight out of the bottle."

As in other examples where *should* has been chosen in the translation the context in (12) involves the existence of other opinions.

5.1.3 Should *with the function of expressing a negative emotion*
Should was found in rhetorical *why*-questions with the illocutionary function of complaint:

> (13) Varför <u>ska</u> vi släpa på tunga trasmattor när det finns bra såna i plast. (SC1)
> Why <u>should</u> we drag heavy rag rugs around when you can get such good plastic ones?

Example (14) is similar although *skall* does not occur in a *why*-question. Instead, the sentence is a statement asserting (or at least implying) that it should not be so difficult.

> (14) Att det <u>skall</u> vara så svårt. (SCO1)
> That it <u>should</u> be so difficult.

In (15) the "foolishness that you should stay in America" conveys the speaker's negative attitude (disapproval, irritation):

> (15) — Den där fånigheten att du <u>skall</u> vara kvar Amerika är ingenting att diskutera, sade Gunnar. (JMY1)
> "This foolishness that you <u>should</u> stay here in America is not to be discussed," said Gunnar.

Summing up, the translations with *should* illustrate several ways in which *ska/ll* can be used and confirm earlier analyses discussing the functions of *should*.

- *Should* is subjective and performative. It is used in non-argumentative (neutral) contexts to express the speaker's personal opinion in a tentative way (*I think I should*) and to make a tentative (*you don't think you should*), or co-operative suggestion (*perhaps we should*).

- *Should* can be used in argumentative contexts where the speaker disagrees with attitudes actualized by the preceding context.

- It is used (particularly in *why*-questions) to express negative effects such as irritation or impatience that things are not different.

5.2 *Ska/ll* translated as *have to*

5.2.1 Have to *in the translation expressing negative feelings*
Ska/ll can also be translated by a modal auxiliary or semi-auxiliary expressing strong necessity such as *must* or *have to.* The general analysis of *have to* is that it has to be chosen when the deontic source is external to the speaker (Collins 2009: 60). However this analysis does not cover all the ways it is used in discourse. As shown by Myhill and Smith (1995: 111), *have to* (unlike *must)* is often used in emotional contexts where it indicates negative affectedness (involving the speaker or the hearer) or something which is bad for everyone (cf.

also Lewis 2015). This is illustrated by example (16), where *have to* conveys the speaker's negative feelings (irritation, impatience, complaint) with a situation where he has to pretend to think the sun is shining (while he is being examined by the doctor):

(16) — Varför <u>ska jag låtsas</u> att det är solsken? frågade jag där jag låg avklädd på den vita båren. (MS1)
"Why <u>do I have to think</u> of the warm sun?" I asked as I lay undressed on the white examination couch.

The translation with *have to* conveys the speaker's negative feelings such as annoyance with an undesirable situation. The question is not about the reason for a certain state of affairs but a complaint ("I shouldn't be thinking of the warm sun"). According to Narrog (2016: 111), "such usage does not appear to be directly performative, but aids in forming a specific type of illocution (in this case a complaint), and indicates negative affectedness, in this case on the part of the speaker".

In (17) *have to* has been chosen in the translation to convey the speaker's negative feelings (annoyance, irritation) with an unwanted situation ('I shouldn't really have to tell you this'):

(17) Inga byxor, hur många gånger <u>ska</u> man <u>behöva</u> säga det? (PP1)
No swimming trunks, how many times <u>do I have to</u> tell you?

The Swedish original contains *ska behöva* (lit. 'shall need to'). However *need to* is not used in the translation since it would not have conveyed the speaker's negative emotions.

In (18) the unpleasantness of "experience anything like this" is obvious from the situation and the use of *have to* in the translation:

(18) Jag hoppas ingen av er nånsin <u>ska behöva</u> uppleva det. (HM1)
I hope none of you ever <u>has to</u> experience anything like this."

In (19) the translator has used *can* rather than *have to* without any apparent difference of meaning to translate the construction *skall behöva* (rather than just *skall*). The question is rhetorical and *can* implies that "it should not be possible for anyone to sink so low":

(19) Hur lågt <u>skall</u> man <u>behöva</u> sjunka! (MR1)
How low <u>can</u> you sink?

Examples (17)–(19) suggest that it is the construction *ska behöva* ('shall need to') which is translated into English *have to* (or *can*) and serves as the basis for the interpretation of the negative meanings complaint, annoyance, irritation, etc.

In (20), the negative evaluation is additionally expressed by the superordinate clause (*it's only that it's silly*):

(20) Det är bara det att det är konstigt att det <u>ska</u> vara mil till en skvätt mjölk. (SC1)
It 's only that it 's silly to <u>have to</u> go miles for a drop of milk.

Have to in the translation indicates that the meaning is evaluative rather than obligation:

(21) Den svenska sommaren är för vacker och för kort för att sånt här <u>ska behöva</u> hända. (HM1)
      The Swedish summertime is too beautiful and too short for something like this to <u>have to</u> happen.

In (22), the negative interpretation is conveyed by the context and by *have to* in the translation. What is negative in this context is 'today's traffic' and how it threatens survival.

(22) vi ungdomar, Sveriges framtid och ljusa hopp, <u>ska</u> ju <u>överleva</u> dagens trafik så vi kan växa upp och bidra till morgondagens. (PP1)
      we young hopefuls, the guardians of Sweden's future, <u>have to survive</u> today's traffic in order to grow up and contribute to tomorrow's.

5.2.2 Have to *in the translation with performative meanings*
Have to, like *should*, can have several different meanings. In example (23) it could be argued that *have to* is used performatively (instead of *must* or an imperative) to make a command or a request:

(23) -Vad heter hunden? frågar jag.
      -Först <u>ska</u> du svara på en fråga! (PCJ1)
      -What is the dog's name? I ask.
      -First you <u>have to</u> answer a question!

5.2.3 *Translations with* have to *indicating habitual meaning*
When the subject of *have to* is a third-person noun phrase, *ska/ll* is anchored to some kind of norm (according to a regulation or with reference to what is habitual). In example (24) the verb is passive: the agent is unexpressed and *have to* has habitual meaning (cf. Smith 2003: 259):

(24) Varje helg <u>skall</u> marsvinen på vinden vägas. (PCJ1)
      The guinea pigs in the attic <u>have to</u> be weighed every weekend.

To sum up, *have to* (rather than *must*) is used in the translation where the speaker is negatively affected by the verbal action or evaluates it negatively as a complaint. Especially in questions, the translations convey that *has to* has the function to express annoyance, complaint and irritation. Finally, performative and habitual meanings were also attested.

5.3   Translations with *be supposed to, be meant to, that means*

In the examples in the corpus *be supposed to* implies a negative evaluation of the hearer or the situation.

(25) – Jag har inte talat med henne ännu.
      <u>Är det meningen att hon ska</u> följa med? (Lit: "is it the intention that she shall go along")

- Nej. (HM1)
I have n't talked to her about it yet."
"Is she supposed to go along?"
"No."

The translation by the passive *be supposed to* without an agent is more closely related to *have to* than to *must* since it is evaluative rather than performative. It refers to the previous scheduling of an event or to a previous agreement between the speaker and the hearer as the deontic source.

In (26) *ska* has been translated as *was meant to*. As with *have to*, the implication could be that the speaker is critical or reproachful (it shouldn't be simple):

(26)   Vem har sagt att det ska vara enkelt. (PE1)
       Who said it was meant to be simple.

In (27) the translator has chosen a construction with *do you want sb. to* since the negative meaning (reproach) is already conveyed by the context:

(27)   "Är det meningen att Borka ska ta över allt rövande också här i Mattisskogen", undrade hon bistert, (AL1)
       "Do you want Borka to take over all the robbing in Matt's Forest as well?" she asked sharply,

Summing up, the translations with *be supposed to* give additional evidence that *ska/ll* can express negative evaluation. The source of the imposition is what the speaker and hearer have agreed to do.

5.4   Translations in the 'request domain'

Modal markers of obligation can be assumed to have to do with action, and there is a close link between deontic modality and speech acts (Palmer 1986, Narrog 2016). In fact, many uses of *ska/ll* belong to the 'request domain' and have a performative or subjective quality. In the appropriate context (e.g. when it is translated by *must* or *should) ska/ll* can be interpreted as making a request, order, command or more positively as expressing advice, suggestion or recommendation.

5.4.1 *Translations with* must
Must is typically used in the translation in scenarios where the speaker has authority over the hearer and performs a command or an order.

(28)   Du ska hem och äta! (ARP1)
        You must go back and eat.

In (28), *you must* places the hearer under the obligation to act (you must go back and eat since I am telling you). *Must go* translates *ska* without a main verb but before an adverbial indicating direction (Teleman *et al*. 1999: 312).

Translations with *must* can also reveal non-performative uses of *ska/ll*. In (29), where the deontic source is not specifically the speaker, *must* is not performative but the deontic source refers to social norms (religion, tradition, etc.):

(29) Man <u>ska</u> väva livet till slut, sade hon. (TL1)
     "We <u>must</u> weave life to the end," she said.

5.4.2 *Translations with an imperative*
Among other translations of *ska/ll* in its performative use is the imperative. Imperatives can be regarded as illocutionary force indicators with a direct and performative function. Translations with an imperative come close to translations with *must* in assuming that the speaker has authority or power over the hearer. All the imperative examples translating *skall* express what Coates (1983: 32) refers to as "strong obligation" and paraphrases as "I order you to" (cf. Section 5.4.3 where *ska/ll* is translated as *I want you to* to convey a more polite meaning):

(30) Det <u>ska</u> du då inte bekymra dig om, sa hon. (KE2)
     "<u>Do n't you</u> worry your heart about it," she replied.

(31) — Gå — naturligtvis <u>ska</u> du gå, säger han upprörd. (MS1)
     "Run along then," he says outraged.

5.4.3 *Translations having a close association with politeness strategies*
A. Translations referring to the speaker's wishes:

I want you to can be used in social situations where the speaker has authority over the hearer and it is therefore sufficient to express what the speaker wants the hearer to do:

(32) <u>Du ska</u> säga efter mig på ungerska, säger han med sammanbitet allvar. (MS1)
     <u>I want you</u> to repeat something after me in Hungarian," he says with grim seriousness.

The translation with *I should like us to*, as in (33), indicates that the speaker is making a suggestion rather than an order:

(33) <u>Jag vill att</u> vi alltid <u>ska</u> kunna prata vid varandra. (TL1)
     <u>I should like</u> us always to be able to talk to each other.

B. Translations directed to fulfilling the hearer's wishes:
In example (34) *ska* has been translated by a polite *I want to*. *I want to* aids in performing speech acts such as offers where the speaker promises to perform an action which is favourable to the hearer.

(34) — Men först <u>ska</u> jag spela någonting för dig, (MS1)
     "But before we leave I <u>want to</u> play you something."

C. Translations as questions about the hearer's wishes:
The translation with a question about the hearer's wishes in (35) indicates that the speaker is making an offer. More indirectness is not needed since the speaker is granting a favour (Leech 2014: 153).

(35)   <u>Ska du</u> ha dej lite kaffe? (GT1)
        <u>You want</u> a little coffee?

The translation *would(n't) you like* in (36) is a polite alternative to make an offer since it makes it easier for the hearer not to comply.

(36)   <u>Ska</u> faster <u>inte ha</u> en liten sherry i alla fall? (ARP1)
        "<u>Would n't you like</u> a little sherry after all, Auntie?"

The translations in (35) and (36) have in common that they draw attention to the interaction between *ska/ll* and polite strategies for performing speech acts such as asking about the hearer's wishes (an indirect way to make a polite offer).

In (37) the translation by *do you want* is associated with the function to make a suggestion:

(37)   <u>Ska vi</u> dra det genast… (SW1)
        <u>Do you want</u> to hear about it?
        Lit. "'ska' we go through it at once"

6.   Conclusion

As a result of grammaticalization *ska/ll* has many different meanings. The translations showed that nearly half of the examples of *ska/ll* had a future meaning (53.73%) while the deontic meaning represented roughly 35% of the examples. The evidential meaning was found only in a few examples. However, roughly 10% examples were difficult to classify.

The focus in this paper has been on what the translations can show about the functions of *ska/ll* and the factors governing the choice of translation (and distinguishing between different senses). The translations with *must* and the imperative point to the performative uses of *ska/ll* associated with imposed obligation and authority such as commands and orders. However, the translations also showed that *ska/ll* is involved in speech acts such as offers, recommendation, suggestion, advice where the strength of imposition has been weakened. The translation with *I want you to* or *should like you to* is another indication that the avoidance of strong imposition threatening the hearer's face is involved in how *ska/ll* is translated.

The translations also pointed to more unexpected uses of *ska/ll*. The translations with *should*, *have to* or *be supposed to* nearly always expressed negative evaluation if the context of the utterance could be interpreted as emotional or argumentative. *Should* and *have to* were found in usages where they were closely associated with illocution types such as reproach or complaint.

The translations also draw attention to linguistic and non-linguistic factors aiding to distinguish functions of *ska/ll* (or choosing a particular correspondence). Such factors are, for example, (a) the interplay with the meaning of the infinitival verb, (b) performativity (whether the speaker uses the obligation marker to perform a speech act or to report on

habitual obligations), (c) the type of discourse context (whether there is a case of disagreement or high emotionality), and (d) the speakers' tendency to avoid explicitly expressing their authority or power. The translations could for example be explained by their interaction with politeness strategies.

References

Altenberg, B. and Aijmer, K. 2000. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In *Corpus Linguistics and Linguistic Theory. Papers from the 20th International Conference on English Language Research on Computerized Corpora* (*ICAME 20*) *Freiburg im Breisgau 1999,* C. Mair and M. Hundt (eds), 15–33. Amsterdam & New York: Rodopi.

Altenberg, B. and Granger, S. 2002. Recent trends in cross-linguistic studies. In *Lexis in contrast. Corpus-based approaches*, B. Altenberg and S. Granger (eds), 3–48. Amsterdam/ Philadelphia: John Benjamins.

Altenberg, B., Aijmer, K. and Svensson, M. 2001. *The English-Swedish Parallel Corpus (ESPC): Manual.* Department of English, Lund University. Available at: http://www.sol.lu.se/engelska/corpus/corpus/espc.html.

Coates, J. 1983. *The Semantics of the Modal Auxiliaries*. London and Canberra: Croom Helm.

Collins, P. 2009. *Modals and Quasi-modals in English.* Amsterdam/New York: Rodopi.

Dyvik, H. 1998. A translational basis for semantics. In *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, S. Johansson and S. Oksefjell (eds), 51–86. Amsterdam/New York: Rodopi.

Dyvik, H. 2004. Translations as semantic mirrors: from parallel corpus to wordnet. In *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora*, K. Aijmer and B. Altenberg (eds), 311–326. Amsterdam: Rodopi.

Hilpert, M. 2006. A synchronic perspective on the grammaticalization of Swedish future constructions. *Nordic Journal of Linguistics* 29(2): 151–172.

Hilpert, M. 2008. *Germanic Future Constructions. A Usage-based Approach to Language Change*. Amsterdam/Philadelphia: John Benjamins.

Hopper, P. and Traugott, E.C. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Ivir, V. 1981. Formal correspondences vs translation equivalence revisited. *Poetics Today* 2(4): 51–59.

Ivir, V. 1987. Functionalism in contrastive analysis and translation studies. In *Functionalism in linguistics*, R. Dirven and V. Fried (eds), 471–481. Amsterdam/Philadelphia: John Benjamins.

Leech, G. 2004. Recent grammatical change in English: data, description, theory. In *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora*, K. Aijmer and B. Altenberg (eds), 61–81. Amsterdam/New York: Rodopi.

Leech, G. 2014. *The Pragmatics of Politeness*. Oxford: Oxford University Press.

Lewis, D. 2015. A comparable corpus approach to the expressions of obligation across English and French. *Nordic Journal of English Studies* 14(1): 152–173.

Myhill, J. and Smith, L. 1995. The discourse and interactive functions of obligation expressions. In *Modality in Grammar and Discourse,* J. Bybee and S. Fleischmann (eds), 239–292. Amsterdam: John Benjamins.

Narrog, H. 2016. The expression of non-epistemic modal categories. In *The Oxford Handbook of Modality and Mood*, J. Nuyts, J. and J. van der Auwera (eds), 89–116. Oxford: Oxford University Press.

Palmer, F. R. 1986. *Mood and Modality.* Cambridge: Cambridge University Press.

Smith, N. 2003. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In *Modality in Contemporary English*, R. Facchinetti, M. Krug and F. Palmer (eds), 241–266. Berlin/New York: Mouton de Gruyter.

Teleman, U., Hellberg, S., and Andersson, E. 1999. *Svenska Akademiens Grammatik.* Stockholm: Norstedts.

Verhulst, A. and Heyvaert, L. 2015. Root modal uses of *should, ought to* and *be supposed to* in present-day British English: From patterns and profiles to proficiency guidelines. *English Studies* 96 (5): 562–595.

Author's address

Karin Aijmer
Department of Languages and Literatures
University of Gothenburg
PO box 200,
405 30 Göteborg
Sweden

karin.aijmer@eng.gu.se

Swedish *orka* viewed through its English correspondences – ability, insufficient strength/energy or insufficient volition?

Mats Johansson, Lene Nordrum

Lund University (Sweden)

Abstract: This paper explores the Swedish auxiliary *orka* and its English correspondences as reflected in English-Swedish parallel corpora. *Orka* is interesting from a contrastive perspective since it lacks a straightforward equivalent in English. We show that most of the English correspondences, both in the direction Swedish original to English translation and in the direction English original to Swedish translation, indicate a semantic analysis of Swedish *orka* involving a combination of two meaning components 1) *ability* and 2) *sufficient physical or mental strength/energy*. We suggest an analysis inspired by Nadathur (2016) where *ability* is the core semantic property, but sufficiency of physical or mental energy is included in the lexical meaning of the verb as a potential obstacle to ability. In addition, our material includes correspondences reflecting a second meaning involving *sufficiency of volition* (von Fintel, 2006), which we assume is derived from the meaning above. We also note that *orka* predominantly occurs in negative polarity contexts, and speculate that the relatively recent use of the imperative form *orka* in informal talk, meaning something along the lines of 'I could not be bothered', might have its source in such negative connotations. In addition to offering a specified semantic description of the Swedish verb *orka*, our study contributes to cross-linguistic studies of expressions of *sufficiency* meanings and sufficiency as a meaning component in verbs (Fortuin, 2013).[1]

Keywords: sufficiency verbs, *orka*, participant-internal possibility, parallel corpora, English/Swedish

1.   Introduction

The Swedish verb *orka* is interesting from a contrastive perspective since it lacks a straightforward correspondence in English. *Norstedts Swedish-English Dictionary* (Petti, 2000: 665), for example, provides the following idiomatic correspondences of *orka*:

(1)   Jag (du etc.**) orkar (orkade)** + inf.
       He (you etc.) **can (could)** + inf.

(2)   Han **orkar** arbeta -
       He **is able** to work…He **is capable** of working.

(3)   Nu **orkar** jag inte [hålla på] längre.
       I **cannot** go on any longer; I am **too tired** to go on.

(4)   Om jag **orkar** [med det] ska jag…
       (har **krafter nog** [have sufficient strength]): if I **have strength** enough (the strength) for it…
       (har ork [have 'energy']): if I **can manage** (**am up to**, **have energy enough** for) it…

As indicated in Norstedt's list, *orka* can function both as a main verb, especially in sense 4, and as an auxiliary, but it is considerably more frequent as an auxiliary and has been since medieval times (Lagervall, 2014: 341). Our focus in this article is on *orka* as an auxiliary verb. In terms of the meaning of *orka*, we see that Norstedt's list includes *ability* verbs (*can, be able to*) as well as expressions of the type *having strength (enough)*, which raises the question whether *orka* is polysemous. In the major Swedish reference grammar, *Svenska Akademiens Grammatik* (Teleman *et al.* 1999(4): 288), however, the two meanings, *ability* and *having strength*, are described as intertwined: *orka* is defined as a *potential auxiliary* which indicates whether the subject referent has sufficient strength or stamina to bring about the state of affairs in question, and further described as "a more specified meaning than *kunna* (can)" [our translation].

        Observations such as these raise the question of what type of semantic properties compose the meaning of *orka,* and how these properties are related. In other words: How can we account for Teleman *et al.*'s observation of a more specified meaning of *kunna* ('could') in semantic terms? Drawing on the general methodology proposed by Johansson (2007), we have earlier shown that the semantic account of another Swedish potential auxiliary, *hinna*, roughly 'having enough time to be able to do something', can be sharpened through the lens of parallel corpora (Johansson and Nordrum, 2016). In this study, we use the same general approach to explore the semantics of *orka*. The idea is that if we consider word-to-word translations of a lexical item as well as lexical material typically occurring in the context of the item, the semantic properties figuring in the interpretation of the item are brought to the surface, and it becomes possible to say something about how these properties are related. We consider both how *orka* is translated from Swedish original texts into English and cases where *orka* shows up in translations of English original texts.

1.1   *Orka* on Modality's Semantic Map

The starting point for our semantic analysis is Teleman *et al.*'s (1999(4): 288) observation that *orka* is a specialization of the modal auxiliary *kunna* ('could'). A model suitable for illustrating such specialization and for pointing out contrastive differences with verbs in other languages is van der Auwera and Plungian's (1998) account of modality: *Modality's semantic map*. On Modality's map, typologically general modal meanings are illustrated as separate boxes, but the map also allows for specialized modal sub-meanings as sub-domains inside the more general meaning categories. A lexical item in one language can thus cover a large area of several general meanings (several separate boxes) whereas its most frequent correspondence in another language might cover a more limited space on the map, or perhaps only a sub-domain of one category. English *may* and its Dutch counterpart *mogen* are a case

in point, where the latter is restricted to encoding deontic modality, whereas *may* has a wider range of meanings (see van der Auwera and Plungian, 1998: 89).

As illustrated in Fig. 1, the general meaning categories on Modality's semantic map are the semantic domains of *possibility* and *necessity*. The map is thus fairly restricted in terms of what is included as modality, focusing on what van der Auwera and Plungian (1998: 86) refer to as "a reasonable representation of core modality". Other accounts of modality can be more inclusive and include additional meanings, such as for example *volition* (see discussion in van der Auwera and Plungian, 1998: 84). The possibility/necessity distinction on Modality's map (Fig. 1) cuts across three types of modality: *participant-internal modality*, *participant-external modality* and *epistemic modality*. In addition, *deontic modality* is included as a sub-domain of participant-external modality (van der Auwera and Plungian, 1998: 82).

Participant-internal possibility	Participant-internal necessity
Participant-external possibility	Participant-external necessity
Deontic possibility	Deontic necessity
Epistemic possibility	Epistemic necessity

Figure 1. *Modality's semantic map* **(based on van der Auwera and Plungian, 1998: 82).**

Of the meaning categories in Fig. 1, *participant-internal possibility* and *necessity* capture possible and necessary states of affairs related to a core participant's capacities, mostly the Subject, whereas *participant-external modality* involves possibilities or necessities mediated by some participant-external circumstance. In the case of *deontic modality*, which is a specialization of participant-external modality, the external circumstance is some authority or social norm. *Epistemic modality*, lastly, expresses the speaker's judgment of the likely truth-value of a proposition.[2] Examples (5a)–(5d) illustrate the different types of possibility and examples (6a)–(6d) the different types of necessity (van der Auwera *et al.*, 2009: 274).

(5)   a. I **can** swim.
            (Participant-internal possibility)
        b. To get to the station, you **can** take bus 66.
            (Participant-external possibility)
        c. You **can** stay home – you have my permission.
            (Deontic possibility)
        d. He **may** come, or he may not – I don't know.
            (Epistemic possibility)

(6)   a. I **have** to have a cup of coffee, otherwise I can't function.
            (Participant-internal necessity)
        b. In order to get to the station, you **have** to take bus 66.
            (Participant-external necessity)
        c. You **must** stay home now, and this is an order.
            (Deontic necessity)

---

[2] An important distinction between epistemic modality and participant-internal and participant-external modality is that epistemic possibility takes scope over the entire proposition whereas the latter two types deal with "aspects internal to the state of affairs that the proposition reflects" (van der Auwera and Plungian, 1998: 82).

d. He **must** be home now; he left for the office a long time ago.
(Epistemic necessity)

Returning to *orka*, we note that Teleman *et al.*'s description places *orka* in the domain of *participant-internal possibility*. Specifically, in Teleman *et al.*'s (1999(4): 288) account, *orka* is part of a category referred to as *potentiella hjälpverb* ('potential auxiliaries'), which gives a latent possibility for something to happen or for someone's ability to do something. In the case of *orka*, the meaning proposed is that the Subject referent has enough strength or stamina to perform a certain action (Teleman *et al.*, 1999(4): 315). In our view, the reference to 'enough strength or stamina' relates *orka* to an interesting category of verbs referred to as *verbs of sufficiency* (Flint, 1980) or *sufficiency verbs* (Chesterman, 2007; Johansson and Nordrum, 2016).

1.2   Sufficiency verbs

The Scandinavian languages and the languages spoken around the Baltic Sea feature a number of sufficiency verbs, but to our knowledge sufficiency verbs are relatively infrequent outside of this region. The largest set of sufficiency verbs are found in Finnish: In her dissertation *Semantic structure in the Finnish lexicon: verbs of possibility and sufficiency*, Flint (1980) lists 45 such Finnish verbs. In Flint's study, the sufficiency verbs are given the features +[possibility] and +[sufficiency] and the sufficiency feature can then be further specified to something like [time] or [strength]. Table 1 illustrates four Finnish verbs with correspondences in Swedish.

As the glosses in Table 1 indicate, the semantic category of sufficiency can be encoded in other ways than as a semantic component of verbs, and such expressions of sufficiency are more commonly discussed in the literature. Many languages include a special sufficiency marker (like English *enough* and Swedish *nog*), a construction with a purposive clause, as well as other means of encoding sufficiency. In addition, sufficiency may be contextually derived, e.g. by simple juxtaposition (see Fortuin, 2013 for a survey of ways of expressing sufficiency across 59 typologically diverse languages). It is rare, however, for sufficiency to be incorporated in the meaning of verbs. As noted above, Finnish has a rich inventory of sufficiency verbs, English has *dare* and the more general *suffice*, whereas the Scandinavian languages fall somewhere in between with a handful of sufficiency verbs which specify the type of sufficiency (e.g. sufficiency with respect to a notion of strength/energy for *orka*).

Table 1. Four sufficiency verbs in Finnish and Swedish with English glosses

Sufficiency category	Finnish Verb	Swedish verb	English gloss
Time	*ehtii*	*hinna*	To have enough time
Energy	*jakssaa*	*orka*	To have enough strength/energy
Space	*passaa*	*passa*	'Fit' [to be big/small enough]
Boldness	*kehtaa*	*våga*	'Dare' [to be bold enough]

An interesting question is thus whether 'the sufficiency component' is the only semantic property of a sufficiency verb, and if it is not, whether it is the most salient property. In a previous study of the Swedish sufficiency verb *hinna* ('have enough time to be able to [do sth]') (Johansson and Nordrum, 2016), we argue that the English correspondences of *hinna* reflect a meaning where *time sufficiency* is the most salient, at-issue, meaning of the verb, but that the verb also reflects an *ability* meaning, which is backgrounded. In the terminology of van der Auwera and Plungian's Modality's map (1998), we view *time sufficiency* as

participant-external possibility and *ability* as participant-internal possibility, and *hinna* thus covers two contiguous areas on the map.

For *orka*, in contrast, the sufficiency component cannot really be disentangled from participant-internal possibility since both 'having enough energy' and 'being able to' reflect participant-internal modality. Rather, as suggested by Teleman *et al.*, (1999(4): 315), sufficiency seems to be a sort of 'specialization' of participant-internal modality. In this study, we use translation data in parallel corpora to shed light on how this specialization can be described, and explore whether there are other meaning components of *orka* that are not captured by the description in the Swedish reference grammar or by dictionary entries.

2.   Method and material

Following the general methodology 'seeing language through multilingual corpora' (Johansson, 2007), this paper explores the meaning of *orka* as mirrored in parallel translation corpora. The material was drawn from *The English-Swedish Parallel Corpora* (ESPC), comprising approximately 650,000 words (Altenberg and Aijmer, 2000),[3] and from the Swedish-English components of *The European Parliament Proceedings Parallel Corpus* (Europarl) and the *Amsterdam Slavic Parallel Aligned Corpus* (ASPAC). The two components comprise 33,406,922 words and 1,516,943 words, respectively. All corpora are available from the search interface KORP,[4] managed by the Swedish Language Bank ('Språkbanken'), University of Gothenburg (Borin *et al.*, 2012). The ESPC includes written texts, non-fiction and fiction, mainly published between 1980 and 2000, while the material from Europarl contains non-fiction texts in the form of scripted speeches from the late 20th century to the present day. The ASPAC material, lastly, contains fiction texts from the beginning of the 20th century to the present day. Both the direction Swedish original to English translation (SO→ET) and English original to Swedish translation (ST←EO) were investigated. Since the Europarl and ASPAC corpora do not allow searches that distinguish between the original texts and the translated text respectively, the data from these corpora were sorted manually into original and translations based on the metadata.

Data collection and analysis involved a sequence of steps. First, instances of *orka* were retrieved and manually cleaned so that only instances of *orka* as an auxiliary verb remained. Next, English correspondences were extracted from the parallel corpora's sentence alignment interface, transferred to a database, and supplemented with information about the type of English correspondence and the syntactic and semantic context in both the original and the source text. The examples were then sorted according to translation correspondence and further analysed for patterns.

Some limitations should be noted. First, it is clear that the material does not provide a balanced empirical foundation for a comprehensive study of the meaning of *orka* across registers and dialects; for one, informal talk is not included in the material and some of the examples are retrieved from written texts from the beginning of the 20th century. It follows that *orka* may include meanings and meaning components that are not reflected in our material, and our study is therefore mainly exploratory. Further, the number of examples is limited, which means that frequency can only inform arguments, not validate them. Despite these shortcomings, however, our data has the unique advantage that meaning can be explored through the lens of authentic interpretations of *orka*. That is, neither the source texts

---

[3]   A description of the corpus is also available at http://www.sprak.gu.se/english/research/research-activities/corpus-linguistics/corpora-at-the-dll/espc/.
[4] The KORP interface is available at https://spraakbanken.gu.se/korp.

nor the translations were produced with a theoretical agenda (see Aijmer, 2004), which ensures authenticity of data.

3. Exploring the English correspondences and contexts of *orka*

Table 2 gives the English correspondences of the lemma *orka* in our material. In this section, we use these correspondences to shed light on the semantic profile of *orka*, but it should be noted that the correspondences are not necessarily addressed in the order in which they are listed in Table 2. The section is organized so that we first deal with translations that seem to focus on ability meaning, such as *can* in Table 2 (Section 3.1), and then on such examples that seem to focus on sufficiency of energy or strength, such as *have the energy strength to* in Table 2 (Section 3.2). Next, we move to a more general discussion of sufficiency (Section 3.3), before addressing the observation that the overwhelming majority of our *orka* examples occur in non-assertive contexts (Section 3.4). We end the section with some observations of *orka* in translations with *be bothered/fed up with* (Table 2)*,* which we argue focus on *volition* rather than *ability* (Section 3.5).

Table 2. The English correspondences of *orka* in the parallel corpora*

English correspondence	SO→ET		ST←EO	
	N 104	%	N 29	%
Can	29	27.9	10	33.5
Be (un)able to	12	11.5	4	14
(Can) Manage	17	16.3	2	7
Can bear/cope/stand/face/bring	9	8.7	1	3.5
HAVE (enough/the) *strength/energy to*+inf	8	7.7	1	3.5
Be bothered/be fed up/feel like	7	6.7	1	3.5
too Adj to+inf	5	4.8	3	10.5
Can find the energy/have energy, get strength	3	2.9	-	-
Zero	5	4.8	3	10.5
Other	9	8.7	4	14
SUM	104	100	29	100

*The term 'correspondence' refers to both sources and translations of *orka* (Johansson, 2007: 23). Table 2 gives both congruent correspondences and divergent correspondences of *orka*. Congruent correspondences are those where the source syntax is kept intact (Johansson, 2007: 23-26) and *orka* corresponds to an auxiliary or semi-auxiliary, while divergent correspondences are those with a different syntactic structure, as in the correspondence *too Adj to* + inf. It should be noted, however, that none of the correspondences listed in Table 1 involves major syntactic changes.

3.1 Ability

Roughly half of the translations and the sources of *orka* in Table 2 correspond to a modal auxiliary reflecting 'ability' meaning, which indicates that ability is a prominent meaning component in *orka*. Under ability, we include correspondences with *can, be (un)able to,* and *(can) manage.* However, a closer look at the correspondences with *ability* suggests that the meaning component *sufficiency of mental/physical strength* is usually reflected as well, or is supplied by the context.

In the vast majority of the ability correspondences, an aspectual/durative adverbial indicating that the ability is subject to exhaustion is present in the context, thus adding a sufficiency element, a pattern that is also noted in Norstedt's dictionary (see entry 3 in (1), p.

156). Example (7) illustrates the context with an aspectual/durative adverbial in the direction SO→ET, and example (8) in the direction ST←EO.[5]

(7)   <u>Hur länge</u> skall jag *orka* vara dessa barns samlande punkt? (GT1)
      <u>How long</u> will he *be able to* be the center of these children's lives? (GT1T)

(8)   He shut his eyes and wondered if he *could* hold on <u>any longer</u>. (JRT1)
      Han slöt ögonen och undrade om han skulle *orka* hålla sig fast <u>längre.</u> (JRT1T)

We note that the aspectual/durative adverbial is always present in both the English original and in the Swedish translation in our material, but can be added in the English translations, as in (9).

(9)   Han *orkade* inte prata utan gick och <u>lade sig</u>. (SL1)
      He *could* hardly talk <u>any more</u>, so he went to bed. (SL1T)

We would argue that the addition of an aspectual adverbial in (9) is a way for the translator to capture that *orka* encodes *sufficiency* as well as *ability* meaning. Even if it is not possible to claim that the adverbial *any more* mirrors a sufficiency component, it does reflect the possibility that the Subject's ability to talk is not permanent, but subject to exhaustion, which can then be related to sufficiency. In (9), the sufficiency relation comes in via the co-ordinated clause, *so he went to bed*, which gives the likely reason why the Subject cannot talk any longer: *he* is probably too tired, implying that *he* lacked sufficient strength to talk due to tiredness. Without *any more,* in contrast, the English translation in (9) risks being interpreted as permanent inability since *could* on its own does not encode sufficiency meaning.

The presence of sufficiency meaning is even clearer in another context noted among the ability correspondences. The context in question involves an adjective modified by *so…that* indicating that a state of exhaustion has been reached which interferes with the Subject's ability. The context only shows up in the direction SO→ET, and is illustrated in (10) and (11).

(10) Är vi helt enkelt trötta — är dagens unga kvinna <u>så trött att</u> hon inte längre *orkar* säga det hon känner att hon måste säga — eller har vi ingenting mer att säga varandra på det sedvanliga debattspråket? (MS1)
      Is the woman of today <u>so tired that</u> she is *unable* to say what she ought to be saying in accord with her feelings? Or have we no more to say to each other in the current jargon? (MS1T)

(11) Han var <u>så dåsig, att</u> han knappt *orkade* göra sig mödan att sänka blickarna. (SL1)
      He was <u>so drowsy that</u> he *could* barely rouse himself enough to lower his glance. (SL1T)

It can also be noted that the Swedish original in (10) includes an adverbial, *längre*, which can be compared to *any more* in the English translation above (9).

---

[5] All the examples in the text mark *orka* and its source translation in italics, whereas other structures relevant in the discussion are underlined. In each case, the initials of the author of the text are provided in parentheses, where T indicates that the text is a translation. The source texts can be found in the search interfaces for the corpora listed under references.

A variant of the *so…that* context is the non-congruent correspondence *too* Adj *to*+inf, where it could be maintained that the expression involves a covert ability modal encoded as an infinitive (see Meier, 2003; Hacquard, 2005). This correspondence figures both as a source and as a translation of *orka*, and includes adjectives such as *tired, exhausted, feeble, ill, sick and weary*. Examples (12) and (13) are illustrations.

(12) Nej, uppriktigt sagt inte. Men jag *orkar* inte leta efter den andra texten nu. (MG1)
No, to be honest, it wasn't. But I 'm *too tired to* look for the other passage now. (MG1T)

(13) There they dropped off one by one into uncomfortable sleep full of horrible dreams, as evening wore to black night ; and there we must leave them for the present, *too sick and weary to* set guards or take turns watching. (JRT1)
Så föll de en efter en i en orolig sömn full av mardrömmar, medan kvällen övergick i svart natt, och där måste vi lämna dem för ett tag, alltför trötta och dåliga *att orka* sätta ut vakter eller turas om att speja. (JRT1T)

We suggest that both the ability correspondence (e.g., *be able to*) combined with an adjective modified with *so…that* and the *too* Adj *to*+inf correspondences provide a *sufficiency condition* related to the potential exhaustion of the Subject's ability. This sufficiency condition is specified by an adjective associated with sufficient/necessary amount of energy/strength (such as *tired* or *ill*) and thus related to a point on a scale that marks when the Subject has the necessary amount of some physical or mental entity to carry out a certain act.
In one example in the material, (14), the idea of a point on a scale is made specific by the sufficiency marker *enough*.

(14) Jag var tillräckligt stark för att *orka* knuffa ned en fyllgubbe i vattnet. (SL1)
I was strong enough to shove an old drunk into the water. (SL1T)

In (14), then, we would argue that ability is realized as covert modality in the *to*-infinitive and the sufficiency condition is realized in general sufficiency terms: with the adverb *enough*. (see e.g. Fortuin, 2013). The Swedish original also has *tillräckligt* 'enough', which, in a sense, 'doubles' the sufficiency component of *orka*.

3.2 Sufficiency

Table 2 also includes a HAVE (enough/the) *strength/energy to* + inf correspondence, illustrated in (15) and (16).

(15) Han *orkade* inte bruka årorna, men i stället satte han sig att vagga och gunga i eka. (SL1)
He *had not strength enough* to use the oars, but instead, he seated himself to swing and rock in the scow. (SL1T)

(16) Och framför allt *orkade* han inte ta itu med den biten samtidigt som han skulle ge sig på Wennerström. (SL1)
Above all, he did not *have the energy* to deal with that problem at the same time as he was tackling Wennerström. (SL1T)

We initially observe that there is a potential pattern in the material that *strength* is used for *physical strength* and *energy* for *mental strength*, but this pattern is not clear. For one, it is not always straightforward whether the reference to strength/energy is to physical or mental strength, as in (17), where both possibilities seem plausible.

(17) Det har snart gått två veckor sedan jag skrev sist. Jag har inte *orkat*. (LH1)
      It has been almost a fortnight since I've written anything. I have n't *had the strength.* (LH1T)

Another observation for the correspondences with *strength/energy* is that the ability component of *orka* seems lost in translation. However, we would argue that it is not. In line with our observation about the *to*-infinitives in the *too* Adj *to*+inf correspondences (section 3.1), we note that the correspondence HAVE (enough/the) *strength/energy to* + inf arguably contains covert ability modality encoded in the infinitive. On this view, the correspondence profiles both a sufficiency condition, *have enough strength/energy*, as well as ability in the infinitive, roughly*: have enough strength/energy to be able to do something.*

Interestingly, our material includes only one example, (18), where *orka* is introduced as the correspondence of an expression with *strength* or *energy* in an English original text.

(18) "Don't mind the house, child. I know it 's a mess but I *ai n't got the strength* I once had to keep it tidy". (GN1)
      "Bekymra dig inte för oredan, lilla vän. Jag vet att det är rörigt, men jag *orkar* inte hålla det så fint som förr". (GN1T)

The obvious reason why HAVE (enough/the) *strength/energy to* + inf is not a common source of *orka* is that the parallel corpora include very few examples of this construction, nine in total. We note, though, that six of these nine instances are translated with a corresponding phrase in Swedish, as exemplified in (19) and (20).

(19) Would she *have the energy* to seek one out? (FW1)
      Skulle hon *ha energi nog* att leta upp en? (FW1T)

(20) It seemed that bingo afternoons left her so exhausted both physically and emotionally that she never *had enough energy* left to cook an evening meal. (RD1)
      Det verkade som om hennes bingo eftermiddagar gjorde henne så utsjasad både fysiskt och psykiskt att hon aldrig *hade tillräckligt med energi kvar* för att laga middag åt dem. (RD1T)

This strategy may reflect a tendency for translators to pick a translation that represents the closest formal correspondence to the source item, in line with some principle of equivalence, a principle that has been debated from numerous angles in translation theory over the years (Nida and Taber, 1982; Baker, 1992), but this explanation remains speculative.

What can be concluded, however, is that similarly to the ability translations commented on in section 3.1, the pattern with an aspectual/durative adverbial is also prominent for the HAVE (enough/the) *strength/energy to* + inf translations, as in (21).

(21) "Inget barn behöver svälta ihjäl, <u>så länge</u> Karlsson *orkar* släpa fram korv och potatis". (AL2)

"No child need starve, <u>as long as</u> Karlson *has the strength to* deliver the sausage and mash". (AL2T)

At this point, then, we note that a precise semantic definition of *orka* seems to include two semantic components: *ability* and *sufficiency of mental/physical strength*. The relation between the two components is difficult to tease out, but it is clear that the ability component is always overtly reflected in some form in the correspondence itself, whereas the sufficiency component can come in via the context. For the translations where *orka* corresponds to an ability modal such as *could* or *be able to*, sufficiency comes in via the expressions *so* Adj *that* in the context, or a sufficiency condition is suggested by an adverbial/aspectual adverbial. In the translations with *energy/strength*, ability is expressed by means of covert modality in the form of a *to*-infinitive and sufficiency is reflected in the HAVE (enough/the) *strength/energy* translation. For this correspondence too, an aspectual/durative adverbial is generally present in the context.

3.3   Sufficiency as an obstacle to ability

A possible way to view the relationship between the two meaning components, ability and sufficiency, is that the sufficiency component highlights that there is a potential *obstacle* to the ability component, specified to a point where the Subject's physical or mental strength is insufficient or running out. This type of analysis has been forwarded by Nadathur (2016) to explain the semantics of Finnish *jaksaa*, a sufficiency verb corresponding to Swedish *orka*. Nadathur's analysis is part of a discussion of implicative verbs, a discussion that is beyond the scope of our article, but some of her observations are relevant to our interpretation of *orka*. Inspired by a proposal by Baglini & Francez (2016), Nadathur proposes that *jaksaa* involves *sufficient strength* as a causally necessary, but contextually insufficient, factor for the realization of *jaksaa's* complement verb. On this analysis, *orka* would reflect a Subject's ability to do something, but also highlight that this ability is potentially obstructed by the amount of physical or mental strength available to the specific individual in the specific situation. The proposal that the factor is contextually insufficient explains that even if a Subject has sufficient strength to be able to do something, it does not follow that he or she actually does it – there may be other factors involved in this decision. Interestingly, Nadathur (2016: 1010) points out that *jaksaa* is similar to *manage* in assuming a potential obstacle, but that *manage* does not specify what the obstacle is.

The idea of sufficiency of strength as an obstacle fits with our data. For example, the translations *(can) manage* (Table 2) can be viewed as a generalization of the obstacle *sufficient strength*. In such instances, specification to *sufficiency of strength* is brought in by the context. In (22), the reference to the Subject being "ill and miserable" seem to achieve such specification.

(22)   "Det ska bli ganska intressant att se, om jag *orkar* mer än halvvägs, <u>sjuk och eländig</u> som jag är". (AL2)
"It will be quite interesting to see if I *manage to get* further than halfway, while I am all <u>ill and miserable</u> like this". (AL2)

Also, somewhat tentatively, we interpret another category of verbs found in the translations of *orka* as mirroring a general obstacle. The verbs in question are *bear*, *cope* and *face*. As shown in Table 2, the material includes nine translations where these verbs combine with the *ability* modal *can*, all in non-assertive contexts. Example (23) is an illustration.

(23) Han påminde Lisbeth om att klockan var <u>mitt i natten</u> och påstod att han <u>inte</u> *orkade* tänka på saken. (SL1)
It was <u>the middle of the night</u>. He *could <u>not</u> face* thinking about the whole thing now. (SL1T)

In (23), we would argue that *ability meaning* is reflected in *can*, and that *face* in the English translation profiles a potential obstacle, but the context is necessary to specify the obstacle to sufficient strength. In (23), such specification is suggested by the reference to time, the middle of the night, which suggests that the Subject, *he*, might need sleep and thus does not have sufficient mental or physical strength to be able to think properly.

We also note that in most of the examples with *orka*, the obstacle is not overcome, i.e. *orka* appears in a non-assertive context. On an intuitive note, the non-assertive contexts can be expected from the type of obstacle profiled in *orka*: If you do not have sufficient strength to do something, you may want to make this obstacle salient. If, on the other hand, you have sufficient strength to do something, it might be more likely that you simply say that you did it. On a more formal note, however, the negative contexts raise the question of whether *orka* can be viewed as a so-called *Negative Polarity Item* (NPI), i.e. a lexical item that can only appear in contexts with negative grammatical polarity.

3.4 The negative association of *orka*

In the vast majority of the occurrences of *orka*, both in the SO and the ST texts, *orka* appears in non-assertive contexts where it has been shown that negative polarity items (NPIs) would be triggered (see Lawler, 2005 for an extensive list of NPIs and NPI triggers in English). This is also noted by Teleman *et al.* (1999(4): 311). Below, we refer to these contexts as *NPI triggers*. By far the most common NPI trigger in the material is sentence negation, as in (24).

(24) Ronja förstod att det var synd om Skalle-Per som var så gammal, men hon förstod inte varför knektar och trindskallar skulle komma och bråka vid Vargklämman. Förresten var hon sömnig och *orkade* <u>inte</u> bry sej om det <u>heller</u>. (AL1)
Ronia knew it was sad for Noddle-Pete to be so old, but she did not understand why soldiers and fools would come and make trouble at the Wolf's Neck. In any case, she was sleepy and *could <u>not</u> be bothered* thinking about it. (AL1T)

Sometimes, as in (24), *orka* is accompanied by an NPI (*heller* 'either', in this case), but more often it is not. Negation may also be in a higher clause, as in (25), where *orka* occurs in the complement clause of *förstod* 'understood', and the main clause is negated by *inte*. We may gloss the first sentence with *orka* in (25) as 'she did not understand that they ORKA-past'. This particular sentence is left untranslated in ARP1T, where only the next (also negated) sentence receives a translation with *couldn't be bothered*.

(25) Och Henry instämde med kraft, och det blev en lovsång till Moralen och Ansvaret och Människokärleken som fastern hade kunnat klara sig utan. Hon förstod <u>inte</u> att de *orkade*. Själv orkade hon inte sitta här i den eländiga fåtöljen länge till, och när hon såg att klockan började närma sig två bad hon dem att skynda sig på. (ARP1)
It would have been indecent not to set to work," he said. Henry forcefully concurred, and then it all grew into a hymn of praise to Morality and Responsibility and Love of

Mankind which Auntie could have done without. She *couldn't be bothered* to sit there in that wretched armchair much longer, and when she saw it was almost two o'clock in the morning, she asked them to get a move on. (ARP1T)

In addition to sentence negation with *inte* ('not'), the material includes some other contexts that have been noted to trigger NPIs (e.g. by Lawler, 2005), such as negative adverbs like *aldrig* ('never') and negative pronouns functioning as Subjects or Objects, as in (26) and (27).

(26) Man tror sig veta att vi finns där men det är <u>aldrig någon</u> som *orkar* lägga märke till oss. (KOB1)
"They think they know we 're there, but <u>nobody ever</u> *manages* to notice us". (KOB1T)

(27) […] och <u>ingen</u> hade under hela hösten *orkat* börja igen. (GT1)
And <u>no one</u> had *been able to* start again throughout the fall. (GT1)

Another common NPI trigger in the material is questions, as in (28).

(28) Där får han också syn på sitt bärande; just när han lagt henne ifrån sig och det far genom honom: "Hur länge skall jag *orka* samla barna?" (GT1)
There he glimpses his burden, just after he has put her down. It tears through him: "How long *can* I *manage* to keep the kids together?" (GT1T)

Finally, our data contain a small assortment of other NPI triggers, including a conditional clause, as in (29), a comparative clause, as in (30), and a relative clause modifying *den ende* ('the only [family member]'), as in (31).

(29) Han hade verkligen massor av kakel i denna källare <u>bara</u> han *orkade* ta sig ner i den och vågade trotsa kloakdofterna och öppna på dörren. (LG1)
He actually had hoards of tiles in the cellar <u>if he</u> *could* only *manage* to get himself down there and face the stench of the lavatory and open the door. (LG1T)

(30) Han höll kvar greppet <u>så länge</u> han *orkade.* (JG1)
He held the grip <u>as long as</u> he *could.* (JG1T)

(31) Karin hade varit <u>den enda</u> i släkten som kunnat och *orkat* ta sig an Nora. (MG1)
Karin had been <u>the only</u> family member willing and *able to* take Nora on. (MG1T)

Conspicuous as this tendency may be, we nevertheless argue that it does not warrant an analysis of *orka* as an NPI. The obvious reason is that *orka* also occurs in unambiguously assertive contexts, as in (32).

(32) Men... det var knappast den slags hästar som frestade hans spel. Den lät sig spännas för barnens trilla redan första söndagen. Och *orkade* dra dem alla. (SCO1)
Then he showed me the pony, which he had also won. But... that kind of horse was hardly what tested his gambling. It allowed itself to be harnessed to the children's trap on the very first Sunday, and *managed to* pull them all. (SCO1T)

Although there is only one such example in the SO texts and none in the ST texts, the assertive context in (32) alone seems to invalidate an analysis of *orka* as a strict NPI, at least on the standard assumption that NPIs are barred from assertive contexts. Further, we note that the use of *orka* in (32) is perfectly idiomatic, and has no additional interpretation.

Nevertheless, the overwhelming majority of *orka* in non-assertive contexts, all but one example, is real and needs to be accounted for. One potential approach is to regard *orka* as an NPI with a complex distribution across assertive and non-assertive contexts. English *much* is a notorious example of such complex distribution. For example, there is a sharp difference in acceptability between (33) and (34) (from Huddleston and Pullum, 2002: 826–827).

(33) a. *I enjoy sailing *much*.
       b. I don't enjoy sailing *much*.

(34) This means *much* to the American tradition.

Huddleston & Pullum argue that *much* is an NPI in (33) only, as indicated by its unacceptability in an assertive context, and that fairly fine-grained distinctions need to be made between different syntactic properties of a potential NPI. For *orka*, however, we see no comparable distinction between different uses of the verb that could account for the acceptability of example (32) above.

A second approach is to regard *orka* not as an NPI, but as part of a negative collocation associated with NPIs, an approach taken by van der Wouden (1994). While his approach is ultimately a formal semantic one, which is meant to account for the variability of use of NPIs within a particular language or across different languages, we adopt the basic idea. Specifically, the definition of collocations as co-occurrence tendencies between lexical items or, in an extended sense, between lexical items and syntactic/semantic contexts, allows the classification of *orka* as a polarity-sensitive verb with collocational tendencies to occur in non-assertive contexts.

A third interesting approach to explain the negative contexts of *orka* is suggested by Flint (1980: 134ff), who notes that Finnish verbs of possibility and sufficiency are very common in negative contexts. Her explanation, based on observations of the use of these verbs in interaction, is that they offer excuses, for example in response to requests or invitations. In other words, declining an invitation or refusing a request may be mitigated by making reference to a lack of sufficient time, energy, courage etc. For *orka*, then, which incorporates sufficiency with respect to mental or physical strength/energy, it is possible that its preponderance in non-assertive contexts stems from its frequent use as an excuse for non-compliance with invitations or requests. The exchange in (35) illustrates this use in our data.

(35) — Var är Eva-Liisa, jag vill att hon också...Gå och hämta henne.
       — Jag *orkar* <u>inte</u>. (GT1)
       "Where 's Eva-Liisa? I want her, too... Go get her".
       "I *can't*". (GT1T)

Notice, incidentally, that, in this exchange, that the ability modal *can* serves much the same purpose, by referring to inability as an excuse for non-compliance with a request. As has been noted repeatedly in the literature on indirect speech acts, reference to ability is a common strategy both for making polite requests and for declining to comply with them (e.g. by Searle, 1975). What sufficiency verbs like *orka* contribute, then, is a more specific (in)ability

than 'bare' ability modals like *can*. In our study of Swedish *hinna* 'have enough time' (Johansson and Nordrum, 2016), we do not explicitly address its use in negative contexts, but this verb, too, can certainly be used in a similar manner to express non-compliance, by referring to having insufficient time.

3.5   *Be bothered* and volitional modality

So far, we have argued that all correspondences of *orka* profile ability and involve sufficiency of strength/energy as an obstacle to ability. Further, we note that *orka* has collocational tendencies to figure in non-assertive contexts where *insufficient* strength/energy is highlighted as an obstacle to the Subject's ability. However, an interesting (albeit small) group of translations does not entirely fit this description. This group involves phrases like *be bothered to* and *feel like* (Table 2), which may be understood as expressing a lack of sufficient *volition* rather than ability. Examples are given in (36)-(38).

(36)   Själv *orkade* hon inte sitta här i den eländiga fåtöljen länge till, och när hon såg att klockan började närma sig två bad hon dem att skynda sig på. (ARP1)
She *couldn't be bothered to* sit there in that wretched armchair much longer, and when she saw it was almost two o'clock in the morning, she asked them to get a move on. (ARP1T)

(37)   Nej, hon vill inte. Hon *orkar* inte. (MR1)
No, she did not want to. She *couldn't be bothered*. (MR1T)

(38)   Först på tisdagen *orkade* hon ta sig upp ur sängen.
She did not *feel like* getting up until Tuesday.

We take translations of this type to reflect a meaning that is derived from the sufficiency readings of *orka*. Some support for our assumption is that *The Swedish Academy Dictionary* (SAOB) explicitly claims that this meaning is secondary, and, moreover, that it is mainly found in informal registers today (although their earliest citation is from the 16th century).

3.6   Orka!

We end this section on a more speculative note, related to the volitional use of *orka* reflected in translations such as *couldn't be bothered*. In the speech of young Swedes, there is a relatively new development in the use of *orka*, common enough to be listed in internet dictionaries. Examples (39) and (40) are from *www.slangopedia.se*.

(39)   Ska vi gå en promenad?
shall we go [for] a walk
- Men *orka*! Vi tankar en film istället.
"but ORKA we download a film instead".

(40)   - *Orka* gå till skolan när betygen redan är satta!
"ORKA go to school when the grades are already set".

We interpret *orka* as volitional in these cases, i.e. what is conveyed is not a lack of sufficient energy to go for a walk or to school, but an unwillingness to do so. What is interesting about examples like these, however, is both that *orka* occurs in its base form (presumably as an imperative rather than an infinitive), and that the intended sense is negative, although there is

no explicit negation. This second point relates to the observation above about the overwhelming tendency for *orka* to occur in non-assertive contexts. In informal terms, it appears that in (29) and (30), *orka* itself is capable of conveying this negative sense.

There are (at least) two ways of regarding the negative sense conveyed by *orka* in these examples. First, we may assume that we are simply dealing with irony or sarcasm, either of which can be understood in terms of some notion of 'saying the opposite of what you mean'. A similar explanation has been proposed within a variety of frameworks (e.g. Grice's co-operative principle). A corresponding example from English would be the use of *Care!* In the sense 'I don't care', illustrated by (41) (from *The Urban Dictionary*).

(41) Person one: oh wow! look at me i'm so great.
      Person two: care.

A second approach to the negative use of *orka* is as a phenomenon referred to as *hyponegation* by Horn (2009), which can be found in expressions like *I could care less*, which is often castigated as a corrupted version of *I couldn't care less*. Some support for this approach is that just like *I could care less*, *orka* licenses NPIs, like *längre* ('any longer'), as illustrated in (31) and (32). [6]

(42) I could care less about <u>ever</u> going back to school. (= Horn's 28d)

(43) Orka vänta på Kalle <u>längre</u>! (our example)

The near synonym *palla* in Swedish is also used in this way, i.e. with predominantly volitional meaning and negative import but no overt negation. To what extent hyponegation and this kind of meaning are systematically related is an interesting topic for further study.

4.   Concluding discussion

In this study, we aimed to shed light on the semantic profile of the Swedish verb *orka* as reflected in parallel corpus data. Our data support an analysis where most examples of *orka* reflect two meaning components: participant-internal ability *and* sufficiency of mental or physical strength. In line with Nadathur's (2016) proposal for the semantic profile of Finnish *jaksaa,* a verb similar to *orka*, we suggest that the sufficiency component of *orka* reflects an assumed obstacle, i.e. the presence of sufficient strength/energy, to the event reflected in *orka*'s complement. We also assume that *orka* is a polarity-sensitive verb with collocational tendencies to occur in non-assertive contexts. This analysis is a specification of Teleman *et al.*'s (1999(4): 288) observation that *orka* refers to "whether the subject referent has sufficient strength or stamina to bring about the state of affairs in question, or "a more specified meaning than *kunna* 'can'" [our translation]. In addition, our material includes some examples which reflect a second meaning of *orka* developed from the meaning presented above. In these cases, the sufficiency meaning concerns sufficient *volition* rather than physical or mental strength.

---

[6] An anonymous reviewer points out that a third possible approach to the imperative use of *orka* is the concept of *semantic prosody* (Sinclair 1991; Louw 1993). That is, it could be argued that *orka* has attained negative semantic prosody by way of its association with non-completion due to a lack of strength or willingness, and that this negative semantic prosody can be one contributing factor to imperative *orka*. This is an interesting approach for future studies.

Our analysis of *orka* might contribute to a description of sufficiency as a semantic component relevant to modality. In light of our data, we tentatively suggest that rather than viewing the sufficiency component in *orka* as a specification of ability – in effect a subcategory of participant-internal possibility – the sufficiency meaning comprises a separate meaning layer that stretches across the modal meanings on modality's semantic map and also extends beyond the map.

Viewing sufficiency as a separate layer of a modality map makes sense since we see combinations of core modality and sufficiency all across the map. As we have shown here, *orka* combines ability (participant-internal possibility) and sufficiency. Further, we have previously shown that Swedish *hinna* 'have the time to' does the same, with sufficiency being participant-external (Johansson and Nordrum, 2016). Additionally, at least one other Swedish verb, *slippa* 'not have to' has a sufficiency component superimposed on participant-external deontic modality. Example (44) illustrates *slippa*.

(44) Det fanns inget han hellre önskade än att hon dog. Då skulle han vara ensam kvar. Han skulle *slippa* att ringa henne, han skulle snart ha glömt hur hon ens såg ut. (HM1)
There was nothing he wanted more than for her to die. Then he'd be left alone. He would*n't have to* call her <u>anymore</u>, and soon he'd forget what she even looked like. (HMT1)

Our interpretation of *slippa* could be glossed as 'not undergoing sufficient external force to do sth'. Notice that the translation introduces an indication, *anymore,* that the external pressure to call her is exhausted at some point.

Another observation relevant to viewing sufficiency as a separate layer on Modality's map is the difference between so-called *anankastic conditionals* and similar constructions with *only* (see e.g. Huitink, 2005; von Fintel and Iatridou, 2007). This difference can be understood in terms of deontic modality with superimposed sufficiency. For example, the example of participant-external necessity offered in (6c) (section 1.1), and repeated here in (45), gets a sufficiency reading when supplemented by *only*.

(45) In order to get to the station, you *have* to take bus 66.
In order to get to the station, you *only have* to take bus 66.
(i.e. it is sufficient that you take bus 66)

Our assumption that *orka* is a polarity-sensitive verb with collocational tendencies to occur in non-assertive contexts deserves some comments. We speculate that this collocational tendency is due to the status of *orka* as primarily an ability verb. Specifically, the added sufficiency component seems redundant in assertive contexts, where it is less relevant why someone was able to do something. Thus, we expect 'pure' ability modals where ability is asserted. Moreover, claiming that one has insufficient energy to do something by using negated *orka* serves the interactional purpose of politely declining invitations, refusing requests, etc., as pointed out by Flint (1980). This last observation raises a couple of interesting questions for further research. For example, it is possible that the use of *orka* in excuses has been instrumental in the development of a secondary volitional meaning, which, in addition, may carry the negative force on its own, as in the speech of young Swedish speakers.

References

Aijmer, K. 2004. The semantic path from modality to aspect: *Be able to* in a cross-linguistic perspective. In *Corpus Approaches to Grammaticalization in English*, H. Lindqvist and C. Mair (eds), 57–79. Amsterdam/Philadelphia: John Benjamins.

Altenberg, B. and Aijmer, K. 2000. The English-Swedish Parallel Corpus: a resource for contrastive research and translation studies. In *Corpus Linguistics and Linguistic Theory*, C. Mair and M. Hundt (eds), 15–33. Amsterdam and Atlanta, GA: Rodopi.

Baglini, R. and Francez, I. 2016. The implications of managing. *Journal of Semantics* 33(3): 409–448.

Baker, M. 1992. *In Other Words: a Course Book on Translation*. London: Routledge.

Borin, L., Forsberg, M., and Roxendal J. 2012. Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC, 2012*. Istanbul: ELRA. 474–478.

Chesterman, A. 2007. What is a unique item? In *Doubts and Directions in Translation Studies*, Y. Gambier, M Shlesinger, and R. Stolze (eds), 3–13. Amsterdam: Benjamins,

Flint, A. 1980. *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Fortuin, E. 2013. The construction of excess and sufficiency from a crosslinguistic perspective. *Linguistic Typology* 17:31–88.

Grice, H. P. 1975. Logic and conversation. In *Syntax and Semantics, vol. 3*, P. Cole, J.L. Morgan (eds), 41–58. New York: New York Academic Press.

Hacquard, V. 2005. Aspects of *too* and *enough* constructions. In *Proceedings of Semantics and Linguistic Theory*, E. Georgala and J. Howell (eds), 80–97. Ithacha, NY: Cornell University.

Horn, L. R. 2009. Hypernegation, hyponegation, and parole violations. In *Annual Meeting of the Berkeley Linguistics Society*, 403–423.

Huddleston, R. and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Huitink J. 2005. Analyzing anankastic conditionals and sufficiency modals. *Proceedings of ConSOLE XIII*. 135–156

Johansson, M. and Nordrum, L. 2016. Swedish *hinna* viewed through its English correspondences – have time or be able to? *Nordic Journal of English Studies* 15: 171–199.

Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam/Philadelphia: John Benjamins.

Lagervall, M. 2014. *Modala Hjälpverb i Språkhistorisk Belysning*. Göteborgsstudier i Nordisk Språkvetenskap 23. Göteborgs universitet: Institutionen för Svenska Språket.

Lawler, J. 2005. *Negation and NPIs*. http://www-personal.umich.edu/~jlawler/NPIs.pdf [last accessed 15 November 2017].

Louw, B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*. M. Baker, G. Francis and E. Tognini-Bonelli (eds), 157–175. Amsterdam: John Benjamins.

Meier, C. 2003. The meaning of *too*, *enough* and *so … that*. *Natural Language Semantics* 11: 69–107.

Nadathur, P. 2016. Causal necessity and sufficiency in implicativity. In *Proceedings of SALT 26*, M.Moroney, C-R Little, J. Collard and D. Burgdorf (eds), 1002–1021.

Nida, E.A. and Taber, C.R. 1982. *The Theory and Practice of Translation*. Leiden: E.J. Brill.

Petti, V. (ed). 2000. *Norstedts Stora Svensk-Engelska Ordbok. 3rd ed*. Stockholm: Norstedts ordbok.

Searle, J. R. 1975. Indirect speech acts. In *Syntax and Semantics, vol. 3: Speech Acts*, P. Cole and J. L. Morgan (eds), 59–82. New York: Academic Press

Sinclair, J. 1991. *Corpus, Concordance, and Collocation*. Oxford: Oxford University Press.

Teleman, U., Hellberg, S., and Andersson, E. 1999. *Svenska Akademiens Grammatik, vol. 1–4*. Stockholm: Svenska Akademien.

van der Auwera, J. and Plungian, V. A. 1998. Modality's semantic map. *Linguistic Typology* 1998(2): 79–124.

van der Auwera, J., Kehayov, P., and Vittrant, A. 2009. Acquisitive modals. In *Cross-linguistic Semantics of Tense, Aspect and Modality*, L. H. Hogeweg, d. Hoop and A. L. Malchukov (eds), 271–302. Amsterdam: John Benjamins.

van der Wouden, T. 1994. *Negative Contexts*. Groningen Dissertations in Linguistics. Groningen: University of Groningen.

von Fintel, K. 2006. Modality and language. In *Encyclopedia of Philosophy – Second Edition vol. 10*, D. M. Borchert (ed.), 20–27. Detroit: MacMillan Reference USA.

von Fintel, K. and S. Iatridou. 2007. Anatomy of a modal construction. *Linguistic Inquiry* 38: 445–83.

Corpora and online dictionaries:

"care, v." *Urban Dictionary* September 2003, `https://www.urbandictionary.com/define.php?term=Care` [last accessed 15 November 2017].

"orka" *Slangopedia* August 2008, `http://www.slangopedia.se/ordlista/?ord=orka` [last accessed 15 November 2017].

"orka" *Svenska Akademiens Ordbok* (SAOB) 1950 `https://svenska.se/saob/?id=O1055-0260.QoBq&pz=7` [last accessed 15 November 2017].

The Amsterdam Slavic Parallel Aligned Corpus (ASPAC) Available at `https://spraakbanken.gu.se/korp` [last accessed 15 nov 2017].

The English-Swedish Parallel Corpus (ESPC) Available at `http://www.tekstlab.uio.no/cgi-bin/omc/PerlTCE.cgi` [last accessed 15 November 2017].

The European Parliament Proceedings Parallel Corpus (Europarl) Available at `https://spraakbanken.gu.se/korp` [last accessed 15 November 2017].

Authors' address

Mats Johansson / Lene Nordrum
Centre for languages and literature
Lund University
Box 201
221 00 Lund
Sweden

mats.johansson@englund.lu.se / lene.nordrum@englund.lu.se

The FAIL TO construction: A contrastive perspective

Thomas Egan

Inland Norway University of Applied Sciences (Norway)

Abstract: This article considers the various senses of the English FAIL TO construction in the light of its Norwegian translation correspondences. This construction has been alleged to be in the process of grammaticalising as a marker of negation in Present-day English. The paper tests the hypothesis that if FAIL TO is in the process of grammaticalising as a marker of negation pure and simple, we should find it used to translate, or find it translated by, pure negation markers (the equivalent of English *not*) in other languages. The particular language of translation investigated in this paper is Norwegian in which the default negation marker is *ikke*. It is shown that translations in both directions lend support to the hypothesis.

Keywords: failure, periphrastic negation, inherent negation, grammaticalisation, English/ Norwegian

1. Introduction

The topic of this article is the construction consisting of the verb *fail* followed by a *to*-infinitive, the FAIL TO construction. This construction may be used with several, to some extent related, senses. Three main senses are listed in the *Oxford English Dictionary* (OED), "to be unsuccessful in an attempt or exercise' (sense 12a), 'to leave undone, omit to perform" (sense 10a), and "not to have the effect of" (sense 10c). The first of these senses will be referred to in this article as the Effort sense, the second as either the Duty or Expectation sense, and the third as the Negation sense. The decision to distinguish between two separate sub-senses of 'omit to perform' is prompted by the conviction that there is a germane distinction between omitting to carry out an action the subject has a duty to perform and omitting to carry out an action that the speaker/writer was expecting the subject to perform, without the subject being independently bound to do so. Note that the corpus-based *The New Oxford Dictionary of English* (Pearsall & Hanks 1988) distinguishes the two senses, defining them as 'neglect to do something' and 'behave in a way contrary to hopes or expectations by not doing something', respectively. The four senses of FAIL TO are illustrated by (1)–(4), the first of which encodes implied unsuccessful effort, the second implied neglected duty, the third implied disappointed speaker expectation, and the fourth negation pure and simple. All four are taken from the English-Norwegian Parallel Corpus (ENPC: see section 2).

(1) It was a deep disappointment to him when the Scots *failed to gain* their objectives… (MH1)[1]

(2) If a Member State persists in *failing to put into practice* the recommendations of the Council… (MAAS1)

(3) Incredibly, even now the German High Command *failed to perceive* either the invasion fleet or the significance of the massive Allied activity. (MH1)

(4) Another family of ceratoids *fails to develop* large nostrils …. (SJG1)

It has been suggested by various scholars, including Mackenzie (2008) and Egan (2010, 2016) that the FAIL TO construction as instantiated by instances like (4) is undergoing grammaticalisation. If this is the case one might hypothesise that it would be frequently translated by sentences containing pure negation markers, for instance the default Norwegian negation marker *ikke* ('not'). However, one would not expect FAIL TO to be employed to the same extent as a translation of Norwegian *ikke*, since the default means of coding negation in English is by means of expressions containing *not*. One would therefore hypothesise that there would be more instances of the FAIL TO construction in the English original texts than the English translations. It is these two hypotheses that form the backbone of this article, which explores the question of how instances of FAIL TO in original English texts in the ENPC are translated into Norwegian, and what sort of expressions in original Norwegian texts are translated into English by FAIL TO. The motivation for the study is the wish to test whether or not the contention that FAIL TO is functioning in Present-day English in at least some contexts as a negation marker is supported by the evidence of the translation corpus.

The article is structured as follows: section 2 introduces the corpus, the search queries and criteria for categorising the instances into the semantic categories employed in the analysis. Section 3 contains a brief discussion of the concepts of negation and grammaticalisation, as these may apply to the FAIL TO construction. Section 4 presents the results of the corpus search and the analysis of the translations from and into English. Finally, section 5 contains a summary and conclusions.

2.   The corpus, queries and criteria for analysis

The data for the present study comprise all instances of the FAIL TO construction in both original English texts and English translations in the English-Norwegian Parallel Corpus (ENPC: see Johansson 2007). The ENPC contains extracts from 50 English and 50 Norwegian original texts, 30 fictional and 20 non-fictional, with their translations into the other language.[2] All instances of the lemma FAIL were extracted. The reason for not restricting the search to instances of FAIL TO was the possibility of there being examples with adverbials intervening between the matrix verb and the infinitive marker. In fact there is only one such example, in the English original texts, cited here as (5).

(5) That this was *the* landing, they *failed* for many days *to understand*. (MH1)
    At det var den virkelige landsettingen, *gikk ikke opp for* dem før flere dager senere. (MH1T)
    "…they did not realise..."

---

[1] (MH1) indicates the provenance of the token, 'MH' being the initials of the author. (MH1T) would serve to indicate a translation of the same text, 'T' standing for 'translation'.
[2] For details of the corpus texts see <http://www.hf.uio.no/ilos/english/services/omc/enpc/>.

Having downloaded the relevant instances, I classified them into the four categories introduced in section 1, Effort, Duty, Expectation and Negation. The classification was based on the answers to the following three questions:

(i) Does the predication involve the expenditure of effort on the part of the subject?

(ii) If no effort is involved, does the complement clause encode a situation which the subject had an objective duty to realise?

(iii) If there is no connotation of either effort or duty, does the complement clause encode a situation which the speaker indicates he or she would expect to have been realised?

Since the classification was made by just one researcher, the results were set aside and the analysis of the instances repeated independently after a three-month interval, with any instances placed in a different category, of which there were a handful, subjected to closer analysis.

With respect to the first question, the identification of the presence or absence of an element of effort on the part of the syntactic subject does not as a rule pose any problems for the analyst. (6) may stand as a representative example.

(6) He believed that the reason he *had failed to win* the girl was because all along he had been too timid, too sensitive, too afraid of rejection, pain, ridicule and loss of face. (RF1)
Han mente at grunnen til at han *ikke hadde klart å vinne* piken var at han hele veien hadde vært for sky, for nærtagende, for redd for avvisning og smerte, for å bli holdt for narr eller tape ansikt. (RF1T)
"…had not managed to…"

There can be no doubt that the subject in (6) made an objective effort to gain the affections of the girl. As well as instances like (6), where it is the subject that is the source of the effort, there are also instances in which, while there is obviously an expenditure of effort, there is a metonymic relationship between the syntactic subject and the person or persons who are making the effort. (7) is a typical example of such a relationship.

(7) The drive for Cherbourg *had* thus *failed to achieve* its principal strategic purpose, and when the Americans renewed their attack southwards, they made slow progress through the bocage. (MH1)
Kampen om Cherbourg *hadde* dermed *ikke nådd* sitt viktigste strategiske mål, og da amerikanerne fornyet angrepene sørover, gikk det langsomt over bocage-landskapet. (MH1T)
"…had not reached…"

Instances like (7) were classified as instantiating Effort FAIL TO. (6) and (7) are objective in the sense of Traugott (2010), since the lack of success encoded in both examples is not related to the attitude of the speaker.[3] Duty FAIL TO, which the second question is aimed at identifying, resembles Effort FAIL TO in this respect. It is exemplified here by (8).

---

[3] I write 'in the sense of Traugott', since Langacker (2008, for example) has promulgated an influential, but quite different, interpretation of what constitutes subjectivity.

(8)   As long as a Member State *fails to comply with* a decision taken in accordance with paragraph 9, the Council may decide to apply the following measures: (MAAS1)
Så lenge en medlemsstat *ikke etterkommer* et vedtak etter nr. 9, kan Rådet bestemme å ta i bruk eller skjerpe ett eller flere av følgende tiltak: (MAAS1T)
"…does not comply with..."

It is clear in (8) that the member states in question are duty bound to comply with the relevant decisions. It is not simply the case that the speaker would expect them to do so, although of course it may well be considered the norm for parties to fulfil their legal obligations. The existence of such an external, objective requirement is generally easy to recognise.

The third question is aimed at identifying examples of the Expectation sense of FAIL TO. This sense, in which it is merely the predictions of the speaker that are disappointed, is characterised by the absence of any element of either effort or duty. Consider in this respect example (9).

(9)   Certainly he had a strong streak of solipsism in him, but surely not so strong that he *would have failed to notice* a riot going on around him. (RF1)
Riktignok var han mye av en solipsist, men avgjort ikke så mye at han *ikke ville lagt merke til* at det foregikk opptøyer rundt ham. (RF1T)
"…would not have noticed…"

In (9) the speaker expresses surprise (*surely not so strong that*) at the idea that the subject would have failed to notice the riot in question. There is no implication that the subject had made a conscious effort to do so, nor that he had any duty to do so. There is, nonetheless, no doubt that the speaker would have expected him to do so. Expectation FAIL TO is subjective, in the sense of Traugott (2010), in that it is only the expectations of the speaker that are disappointed.

The final sense of the FAIL TO construction, which will be discussed in more detail in section 3, is the Negation sense, exemplified here by (10). It is characterized by the absence, not only of any element of effort on the part of the subject, or the existence of a duty which the subject neglects to fulfil, but also by the absence of any connotation of expectation on the part of the speaker. In other words, if the answer to all three questions posed by the analyst is 'no', the token is classified as an example of Negation FAIL TO.

(10) To human eyes many of the body markings appear to be beautiful and highly conspicuous, but this is because we so often see them in artificial environments where their camouflage quality *fails to show itself*. (DM1)
I menneskets øyne virker mange av forsvinningsdraktene både vakre og svært iøynefallende, men det skyldes at vi så ofte ser dem i kunstige omgivelser der deres kamuflasjeeffekt *ikke kommer til sin rett*. (DM1T)
"…does not prove of use…"

There is no implication in (10) that the speaker would expect the camouflage quality to show itself. Section 4.1 contains details of how many of the four types of FAIL TO are found in the original English texts and in the translations from Norwegian.

3. FAIL TO, negation and grammaticalisation

Millions of words have been written about the concept of negation, which raises many thorny questions for philosophers as well as linguists (see, for instance, the introduction to Horn 2001). Indeed, it is by no means certain that an answer can be found to many of these questions since, according to Dixon (2012), "Negation is an intrinsic notion in the world, and in language. It is not something which can be defined or even explained." (Dixon 2012: 89) In the present section, however, I will adopt a common-sense definition of negation as encoding the non-realisation of some situation or other. This simple definition would, I imagine, meet with general approval, at least in cases of predications that do not involve irony or sarcasm.

The present article is not the first to touch on the topic of FAIL TO as encoding negation in the texts in the ENPC. Johansson (1997: 205) discusses the following example.

(11) Men jeg greidde visst *ikke* [lit. did not manage] å få det forklart. (KF1)
But apparently I *failed* to make her understand. (KF1T)

Johansson includes this example under the heading of 'Implicit Negation' and states that in such examples the English verb "can be described as semantically negative". Various scholars employ terms equivalent to 'Implicit Negation' to describe the sort of negation coded by FAIL TO. Karttunen (1971) speaks in terms of negative implicativity, Dixon (2012: 124) employs the term 'inherent negativity', while Huddleston and Pullum (2002: 835) refer to *fail* as one of a class of 'covertly negative lexical items with clausal or clause-like complements'. Other such verbs they mention are *avoid*, *decline*, *forget*, *neglect* and *refrain*.

The description of the sort of negation coded by FAIL TO as implicit, inherent or covert is prompted by the fact that it does not contain a negative particle like *not*. It rests, in other words, on formal grounds. One might, on the other hand, argue that functionally the construction is maximally explicit, since the situation in the complement clause is never ever realised. It differs in this respect from *forget*, which entails the non-realisation of a situation when followed by a *to*-infinitive complement, but its realisation when followed by an *-ing* complement (see Egan 2008: 286). Rather than say that FAIL TO implies negation, one could argue that it entails negation and may, in addition, imply effort, duty or expectation. Indeed, Bolinger writes that *fail* codes 'categorical negation', giving as an example "*There failed to appear* (*there did not appear*) *the very one we needed most*" (Bolinger 1977: 122). This ability of FAIL TO to occur with *there* as subject is taken by Mackenzie as evidence that *fail* is a subject-raising verb. Mackenzie also provides evidence of its negative character. He points out that *fail* is not independently modifiable by an adverbial when it occurs in Negation FAIL TO and that it tends to occur with negative polarity items (Mackenzie 2008: 61).

According to Dixon "Recent work had provided a number of tests for whether a sentence in English should be considered negative – addition of a positive tag, or an addition commencing with *and neither*, or one commencing with *not even*" (Dixon 2012: 93). Egan (2010) used the *and neither* test to investigate whether the matrix verb FAIL TO is backgrounded at the expense of the complement predicate. Boye and Harder (2009, 2012) maintain that this sort of discursive backgrounding is typical of the process whereby a formerly lexical element acquires grammatical status. Egan (2010) contains evidence of discursive backgrounding in the form of a selection of corpus tokens in which phrases like 'x failed to do something' are followed by 'and neither did y'.

Given that FAIL TO functions to all intents and purposes as a pure negation marker in at least some contexts, the question naturally arises as to how this usage has come about. According to Mackenzie, "If the lexical verb *fail* lends itself to a usage which is entirely equivalent to the grammatical strategy of negation, this suggests that, in this usage, it has been subject to a degree of grammaticalization" (Mackenzie 2008: 54). The development of a negation maker from a verb such as *fail* is by no means unusual in the world's languages, as attested by Dixon (2012: 94). Traugott and Dasher (2002) describe grammaticalisation as being "properly conceived as the change whereby lexical material in highly constrained pragmatic and morphosyntactic contexts is assigned functional category status, and where the lexical meaning of an item is assigned constructional meaning" (Traugott and Dasher 2002: 81). According to Diewald, "the process of grammaticalization is a process whereby linguistic items gain grammatical function while reducing their lexical-descriptive function" (Diewald 2010: 18). We have seen that there are contexts in which FAIL TO displays functional as opposed to lexical content. This shows that FAIL TO is grammatical; there remains the question of when it has been grammaticalised.

There are, to the best of my knowledge, no instances of Negation FAIL TO in Middle English. I write 'to the best of my knowledge' since the fact that FAIL TO was itself almost invariably negated before Late Modern English (the earliest non-negated example in the OED is from 1810), with two morphemes encoding non-realisation of the complement situation, one of which serves to cancel the other, renders inappropriate the sort of tests which are normally applied to determine the presence of negation. According to Mackenzie the "notion of disappointed expectation plays a crucial role in the understanding of how *fail* has come to be used" (Mackenzie 2008: 55). It does, indeed, seem more likely that it is the bleaching of the element of speaker expectation that has resulted in FAIL TO functioning as a negation marker. In addition to instantiating semantic attrition of the element of speaker expectation, Negation FAIL TO instantiates attrition of the element of subjectivity itself. Kranich (2010: 118) maintains that while in the early stages of grammaticalization "the newly emerging constructions are often made use of by speakers to express subjective shades of meanings, such meanings tend to get lost in later stages of grammaticalization". While it would be overstating the case to say that FAIL TO is in the later stages of grammaticalization, it certainly exhibits, in examples of Negation FAIL TO, the sort of loss of subjective shades of meaning to which Kranich is referring.

4. Contrastive analysis

Since the goal of this paper, as stated in the introduction, is to investigate the translation correspondences of FAIL TO in order to cast light on whether the construction has come to be used to encode negation pure and simple, the presentation of the Norwegian expressions in this section will concentrate on their function in holding up a mirror to their English correspondences, rather than the semantics of these expressions in their own right. The section is divided into three parts. Section 4.1 presents the overall results for the verb FAIL in all three constructions in which it is commonly employed, in both original English texts and translations into English. Section 4.2 looks more closely at Norwegian translations of the English instances of FAIL TO and section 4.3 at the use of FAIL TO in English translations of Norwegian originals.

4.1 The overall results

The number of instances of FAIL TO in both original English texts and translations into English are shown in Figure 1, which also includes totals numbers of the other two constructions containing FAIL, the intransitive construction and the construction with a nominal object (see examples 12–13).
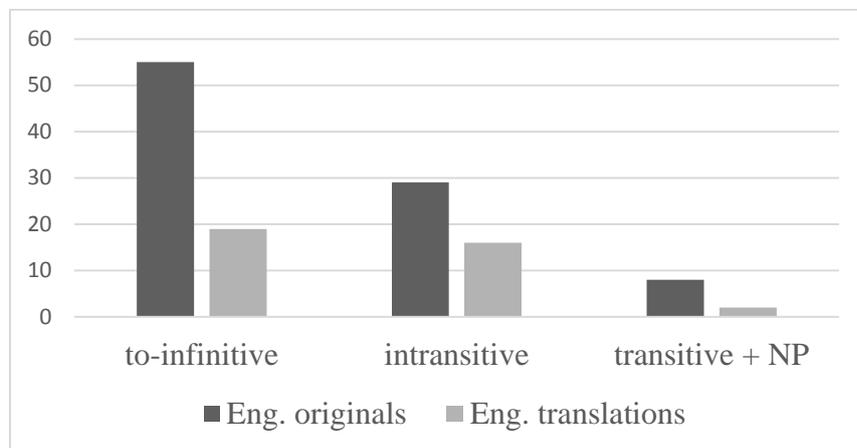

Figure 1: Raw numbers of instances of FAIL in three constructions.

As shown by Figure 1, in all three of the main syntactic roles of FAIL, intransitive, transitive with a nominal (frequently reflexive) object and as a matrix verb with a *to*-infinitive complement, there are considerably more instances in the English originals, confirming one of two hypotheses outlined in section 1. Why do we find fewer instances of all three in the English translations than originals? With respect to the FAIL TO construction, we will see in sections 4.2 and 4.3 that the expenditure of unsuccessful effort tends to be coded in Norwegian by constructions containing a verb denoting effort combined with the negative particle *ikke*, corresponding, for example, to 'did not manage' in English. The maximally congruent, and by far the most common, form of English translation of such phrases contains *not* in place of Norwegian *ikke*. The same point may be made about the other two constructions, exemplified here by (12) and (13).

(12) Maybe she was one of those women who succeeds in business and *fails* in relationship with men. (SG1)
      Kanskje hun var en av de kvinnene som lykkes i karrieren og *ikke får til* sine forhold til menn. (SG1T)
      "…does not succeed in…"

(13) Kate *failed* her eleven plus. (MD1)
      Kate *klarte ikke* opptagelsesprøven til høyeste kursplan på ungdomsskolen… (MD1T)
      "…did not manage…"

If we restrict our attention to instances of the FAIL TO construction, and inquire as to the distribution of the four main semantic types in the English originals and translations, we find that the Effort type is the most common in the translations, while the Negation type is most common in the originals. Since, as is shown in Figure 1, there are more than twice as many instances of FAIL TO in the English original texts, the incidence of the four semantic types is given in Figure 2 in percentages rather than raw figures, for ease of comparison.
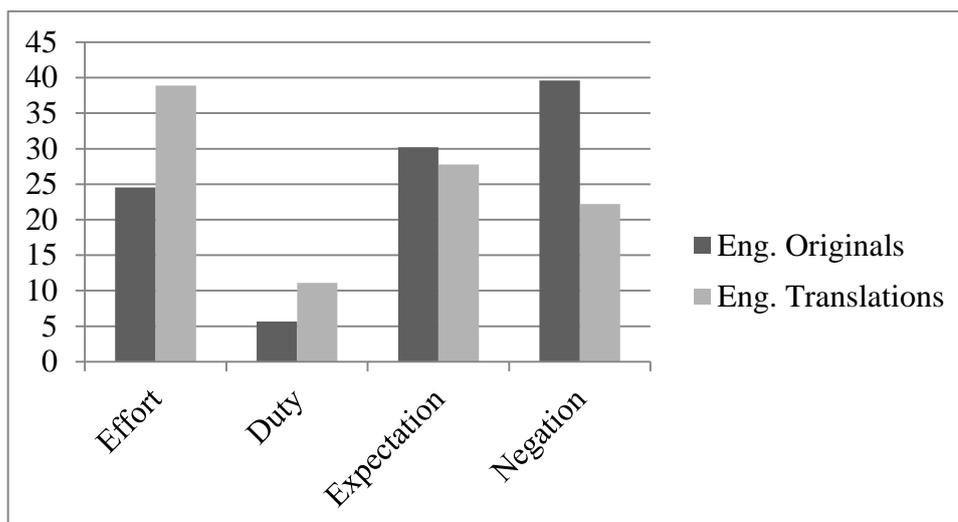
Figure 2. Percentages of four semantic types in English originals and translated FAIL TO.

In sections 4.2 and 4.3 we will look more closely at the translation correspondences of the various senses of FAIL TO.

4.2   Translations of English originals

There are 55 instances of FAIL TO in the English original texts in the ENPC. Instances of the construction are found in just under half of the texts in the corpus (24 out of 50). Their translation correspondences are listed in Table 1, in which near-synonyms are grouped together.

Table 1. Main Norwegian correspondences of English original FAIL TO.

Form of translation	Meaning	Instances
ikke (+ hverken)	not (+ neither)	21
ikke klare, ikke greie, ikke få, mislykkes i, ikke lykkes i	not manage	16
(ikke) unngå	(not) avoid	6
ikke makte, være ute av stand til	be unable	3
unnlate	omit	2
glemme	forget	2
nekte	refuse	1
gi opp	give up	1
aldri slutte	never stop	1
være nødt til	have to	1
Ø		1

The fact that as many as 21 of 55 instances of FAIL TO are translated by either *ikke* or *hverken* ('neither') may be taken as a first indication of support for the hypothesis in section 1 that FAIL TO may be translated by pure negation markers. Since FAIL TO, in all of its senses, always precludes the realisation of the situation in its complement clause, it is no surprise that most of its translation correspondences display negative implicatures. The two exceptions, *aldri slutte* ('never stop') and *være nødt til* ('have to'), which both imply the realisation of the complement situation, are used to translate negated FAIL TO. Example (14) may serve as an illustration.

(14) On the one hand, she *cannot fail to feel* respect for her father's achievements… (MD1)
På den ene siden *er hun nødt til å ha* respekt for det faren har oppnådd. (MD1T)
"…she needs to…"

Similarly, the single token with a zero correspondence, (15), translates negated FAIL TO.

(15) It was a sight that made me think of warm kitchens and well-seasoned stews, and it never *failed to make me ravenous*. (PM1)
Det var et syn som fikk meg til å tenke på varmen på kjøkkenet og duftende lapskaus *og gjorde meg skrubbsulten*! (PM1T)
"...and made me ravenous."

Readers familiar with Norwegian may wonder why *ikke makte* is classified as indicating lack of ability, rather than unsuccessful effort. The reason is that in the single example of this construction in the corpus, cited here as (16), *maktet ikke* translates the Expectation rather than the Effort sense of FAIL TO.

(16) But today even this wonder, continually repeated, *failed to comfort* his spirit. (PDJ3)
Men i dag *maktet* selv *ikke* dette underet, som gjentok seg gang på gang, *å gi ham ro i sinnet*. (PDJ3T)
"…was unable to..."

It is obvious that the subject in (16), *this wonder*, is not agentive; i.e. that it did not try to comfort the spirit of the person in question. There is rather an implication of disappointed expectation on his part.

If we make a crude distinction between the instances in which FAIL TO codes mere negation, or affirmation in the case of (15), and instances where there is an additional implication of Effort, Duty or Expectation, we find that translations either add or subtract such implications in about a third of all cases, indicating perhaps that the alleged semantic distinctions are not felt as keenly by translators as by the present linguist, at least. I will exemplify with some instances containing the implication of effort, starting with instances (17) and (18) where this implication is present in the original and preserved in the translation.

(17) He believed that the reason he had *failed to* win the girl … (RF1)
… at han *ikke* hadde *klart* å vinne piken… (RF1T)
"…had not managed to…"

(18) … which the air forces *failed* to destroy by bombardment. (DL2)
… som flyvåpenet *ikke greide* å ødelegge med bombing… (DL2T)
"…did not manage to…"

The Norwegian matrix verbs in (17) and (18) both imply the expenditure of effort on the part of the subject to realise the complement situation. We may contrast these with the simple negation in the translation in (19).

(19) He *failed to get* this and was instead matriculated as an Arts 11 student… (RF1)
Dette fikk han ikke... (RF1T)
"This he did not get…"

The *this* in (19) refers to a scholarship and, since these are generally viewed in a positive light by prospective students, the translator may have felt it unnecessary to indicate the expenditure of effort to achieve this award on the part of the subject. [4]

We see an example of the opposite tendency, i.e. the addition of the implication of effort, in (20).

(20) We need the concrete and real, as he did; and we *fail to see* this…. (OS1)
 Og vi *mislykkes i å forstå…* (OS1T)
"And we do not succeed in understanding…'

In (20) there is no implication that the subject (we) makes any effort to incorporate the 'concrete and real' into our conception of the cognitive sciences, the topic dealt with in the text from which the example is taken. It merely states that we do not do so. It is the translator who adds the implication of the expenditure of effort.

Unlike (20) the majority of translations of Negation FAIL TO do not add implications of effort. (21) is representative of these translations.

(21) These intense areas automatically draw the eyes of the predator and as he stares at them he *fails to notice* the larger shape which carries them. (DM1)
Disse iøynefallende områdene tiltrekker seg automatisk rovdyrets blikk, og når et stirrer på dem, *legger* det *ikke merke til* den større formen de befinner seg på. (DM1T)
"...does not notice…"

It is instances like (21) that provide the clearest indication that the translator in question has taken FAIL TO to code negation, and nothing more.

4.3 Translations of Norwegian originals

Table 2 contains details of the instances in the original Norwegian texts that are translated by English FAIL TO. The 18 instances are found in 12 texts.[5]

---

[4] We may note in passing that 'ikke få' codes negative implicature when employed as a matrix verb with a past participle complement. This is the case when *failed to deposit* (JB1) is translated as *fikk ikke deponert*.
[5] I have omitted one token from the analysis, since the original Norwegian text contains a quotation in English naturally rendered verbatim in the English translation.

Table 2. Main Norwegian correspondences of *FAIL TO* in English translations.

Form of original	Meaning	Instances
ikke, ingenting	not	8
unnlate, la være	omit to	3
ikke greie	not manage to	2
uten (NP)	without (NP)	2
utebli, ryke ut	other (intrans),	2
Ø (translated by negated *fail to*)	Ø	1

    (22)   Men jeg *greidde* visst *ikke å få* det forklart. (KF1)
           "…did not manage to…"
           But apparently I *failed to make* her understand. (KF1T)

Eight of the instances mark negation either by *ikke*, as in (23), or the negative pronoun *ingenting* ('nothing').[6]

    (23)   Hvis *ikke* den utlagte faren *erkjenner* farskapet. (LSPL1)
           "…does not acknowledge…"
           If the putative father *fails to acknowledge* paternity… (LSPL1T)

In (23) both the original Norwegian text and its English translation instantiate negation, pure and simple, there being no reason to expect the putative father to acknowledge paternity. Indeed quite the opposite is the case, as indicated by the heading of the text section from which the sentence is taken: "Establishment of paternity by a court decision".

The majority of translations resemble (22)–(23) in that they neither add nor subtract implications present in the original formulations. Indeed, there are no cases in which an implication of effort, duty or expectation is subtracted. This follows largely from the fact that most of the Norwegian originals denote negation pure and simple, there being thus no implications to subtract. There are, however, four instances in which implications are added (or at least formalised), three involving Effort, and one Duty. Two of the cases where an implication of effort is added take the form of a construction in Norwegian consisting of the preposition *uten* (without) followed by a noun phrase, as in (24).

    (24)   Det hadde imidlertid liten virkning, for ved valget i 1989 endte det samlede Venstre igjen *uten mandater til Stortinget*. (UD1)
           "…without seats in parliament..."
           This had little effect, for after the 1989 election, the new Liberal Party *failed to win* a single seat in the national assembly. (UD1T)

It is, of course, natural to assume that a political party taking part in a parliamentary election will endeavour to win seats in parliament, but our understanding of this effort is based on our knowledge of contemporary politics and government in the case of the Norwegian original in (24). It is more directly implied in the English translation.

---

[6] To these may be added the non-negated Norwegian original …*alle hadde naturligvis lagt merke til…* (EG1), which merely states "naturally everyone had noticed", but which is translated …*no one could have **failed** to notice...*

The token in which a dereliction of duty is implied in the translation is cited here as (25).

(25) De anså det hele for å være en fillesak og *sikret seg ikke* navn og adresse på noen av vitnene. (KA1)
"…did not get hold of..."
They saw the whole thing as a trivial case and *failed to take* the names and addresses of any of the witnesses. (KA1T)

Whereas the Norwegian original in (25) merely asserts that *they* (the police) did not note the names and addresses of witnesses, the English translation implies neglect on their part in not so doing.

5.   Summary and conclusions

The purpose of the study reported on in this article was to determine whether translation correspondences of the FAIL TO construction could lend support to the contention that this construction is now functioning, in at least some contexts, as a marker of negation pure and simple. The translation correspondences examined were Norwegian, and the data, consisting of translations both from and into English, were taken from the English-Norwegian Parallel Corpus.

Two hypotheses were advanced in section 1. The first predicted that there were likely to be more instances of FAIL TO in the English original texts than in the translations into English. The reasoning behind this hypothesis was the assumption that, if FAIL TO is actually functioning as a negation marker, translators into English of Norwegian negative predications would be more likely to employ the default negation marker *not* than a periphrastic negative. The second hypothesis was that if FAIL TO is indeed functioning as a periphrastic negative, we should expect to find instances translated by the default Norwegian negation marker *ikke*, and possibly also examples where FAIL TO is used to translate *ikke*.

Before presenting the results of the corpus study, I looked, in section 3, at some English language-internal evidence that FAIL TO can function as a negation marker. I also touched on the question of whether the construction has grammaticalised in English and argued that Negation FAIL TO has evolved from Expectation FAIL TO with the gradual bleaching of the element of speaker expectation in the course of the last couple of centuries.

Section 4 contained an overview of the results of the corpus study. The data show clearly that FAIL TO is three times as common in original English than translated texts, thus lending support to the first hypothesis. The reason for this discrepancy is presumably that *not* is the default translation correspondence of negated Norwegian expressions. The second hypothesis is supported by the results presented in sections 4.2 and 4.3, which show that FAIL TO is often translated by, and used to translate, the Norwegian negation marker *ikke*. The fact that it can be used to translate *ikke* is particularly strong evidence of its function as a periphrastic negative since the translator into English has chosen it in place of the default negation marker *not*.

In conclusion, the results presented in this article add inter-linguistic evidence of the negative character of FAIL TO to the sort of intra-linguistic evidence mentioned in section 3. If one accepts that the more evidence one has, and the more different sorts of evidence, the better, it is clear that in some contexts FAIL TO fails to mean more than *not*.

References

Secondary sources:

Bolinger, D. 1977. *Meaning and Form*. London: Longman.

Boye, K. and Harder, P. 2009. Evidentiality: Linguistic categories and grammaticalization. *Functions of Language* 16(1): 9–43.

Boye, K. and Harder, P. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1): 1–44.

Diewald G. 2010. On some problem areas in grammaticalization studies. In *Grammaticalization: Current Views and Issues*, K. Stathi, E. Gehweiler and E. König (eds), 17–50. Amsterdam: John Benjamins.

Dixon, R. M. W. 2012. *Basic Linguistic Theory: Volume 3: Further Grammatical Topics*. Oxford: Oxford University Press.

Egan, T. 2008. *Non-finite Complementation: A Usage-based Study of Infinitive and -ing Clauses in English*. Amsterdam: Rodopi.

Egan, T. 2010. The 'Fail to' construction in Late Modern and Present-day English. In *English Historical Linguistics 2008. Selected Papers from the Fifteenth International Conference on English Historical Linguistics. Volume 1: The History of English Verbal and Nominal Constructions*, U. Lenker, J. Huber and R. Mailhammer (eds), 123–141. Amsterdam: John Benjamins.

Egan, T. 2016. The subjective and intersubjective uses of FAIL TO and 'not fail to'. In *Aspects of Grammaticalization: (Inter)subjectification and Pathways of Change*, H. Cuyckens, L. Ghesquière and D. van Olmen (eds)*, 168–196. Berlin: De Gruyter.

Horn, L. 2001. *A Natural History of Negation. Reissued with new introduction.* Stanford, CA: Center for the Study of Language Information.

Huddleston, R. D. and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Johansson, S. 1997. In search of the missing *not*: Some notes on negation in English and Norwegian. In *From Ælfric to the New York Times: Studies in English Corpus Linguistics*, U. Fries, V. Müller, and P. Schneider (eds)*, 197–214. Amsterdam: Rodopi.

Johansson, S. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Karttunen, L. 1971. Implicative verbs. *Language* 41. 340–358.

Kranich, S. 2010. Grammaticalization, subjectification and objectification. In *Grammaticalization: Current Views and Issues*, K. Stathi, E. Gehweiler and E. König (eds), 101–121. Amsterdam: John Benjamins.

Langacker, R. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Mackenzie, J. L. 2008. Failing without trying. *Jezikoslovlje* 9(1-2): 53–85.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language.* London: Longman

Traugott, E. C. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In *Subjectification, Intersubjectification and Grammaticalization*, K. Davidse, L. Vandelotte and H. Cuyckens (eds), 29–7. Berlin: De Gruyter.

Traugott, E. C. and Dasher, R. B.. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Primary sources:
CLMET: Corpus of Late Modern English Texts. De Smet, H. 2005. *ICAME Journal* 29, 69–82.
NODE: *The New Oxford Dictionary of English*, J. Pearsall and P. Hanks (eds). 1998. Oxford: Oxford University Press.
OED: *The Oxford English Dictionary,* 2nd edition on compact disc. 1992. Oxford: Oxford University Press
ENPC: The English-Norwegian Parallel Corpus. Oslo: Oslo University: see <http://www.hf.uio.no/ilos/english/services/omc/enpc/>.
WebCorp: Birmingham: Birmingham City University: see <http://www.webcorp.org.uk/live/>

Author's address:

Thomas Egan
Inland Norway University of Applied Sciences
Holsetgata 31
2318 Hamar
Norway

thomas.egan@inn.no

The expression of interpersonal and textual functions in Czech and in English

The Czech postfix *-pak* and its translation counterparts[1]

Denisa Šebestová, Markéta Malá

Charles University, Prague (Czech Republic)

Abstract: The study explores the possibility to use translation counterparts as "markers" (Malá, 2013), or "methodological anchors" (Gast, 2015), of discourse functions, i.e. formal correlates of interpersonal and textual functions, which make it possible to detect these functions in the text, and to compare their expression cross-linguistically. We focus on Czech expressions containing the postfix *-pak* (such as, *copak* – 'what + *pak*', *kdepak* – 'where + *pak*'). The postfix *-pak* is shown to be a polyfunctional indicator of discourse function (cf. Grepl and Karlík, 1998). The expressions ending in *-pak* were found to have content / speaker-related functions (such as deliberative meaning, emotional evaluation, (im)possibility) as well as communication / addressee-oriented functions (appeal, establishing/maintaining contact) (cf. Aijmer, 2013; Šebestová and Malá, 2016).

Keywords: postfix *-pak*, interpersonal functions, textual functions, parallel corpus, Czech/English

1. Introduction

The present study examines Czech expressions containing the postfix *-pak* as seen through their English translation counterparts. In most grammars of present-day Czech, the description of such expressions is not systematic, usually being limited to the characteristics of some notable individual particles and pronouns which contain the postfix, and few examples of their typical uses (see Section 2). This study aims to provide a more comprehensive description of the postfix in terms of the discourse functions it can signal. To identify the various functions and uses of the postfix, we employ the methodology of contrastive analysis, using material from a Czech-English parallel translation corpus (Section 4). In this approach, English plays a dual role. Assuming that "linguistic structure is language-specific while the cognitive and functional-communicative substance which

---

constrains it is potentially universal" (Boye, 2012: 7), we may treat the Czech sentences comprising words with the postfix *-pak* and their English translations counterparts as functionally equivalent. This makes it possible for us first, to identify the functions of the Czech expressions by exploring the functions of the English counterparts, and second, to compare the means used to perform these functions in the respective languages.

2.   Theoretical background: the postfix *-pak*

The Czech postfix *-pak* is attached at the end of the word after all inflectional suffixes,[2] cf. e.g., the declension of the pronominal adjective *jakýpak* ("what-*pak*"), *jakéhopak*, *jakémupak*, *jakémpak*, *jakýmpak*; hence it is termed a "postfix" in grammars of present-day Czech (Karlík *et al.*, 2000: 109). Etymologically, the postfix *-pak* evolved from the particle *pak,* as witnessed by forms such as *co pak, kde pak* (Šmilauer, 1969: 28). These forms are considered obsolete today, and are not attested in present-day Czech corpora.[3] Relating the postfix *-pak* to this particle, some Czech grammars refer to it as an enclitic particle (Karlík *et al.*, 1995: 679).

Czech expressions containing the postfix *-pak* are found in several word classes: pronouns, e.g. (1a), pronominal adverbs (1b), particles[4] (1c) and interjections (1d), resulting in homonymous items, such as *copak*.

(1)   a. **Copak** jste tam koupila? (Karlík *et al.*, 1995: 694) – pronoun
       "What-*pak* you bought there?"[5]

   b. **Kampak** jsi to dal? (Havránek *et al.*, 1960: 826) – pronominal adverb
      "Where-*pak* you put it?

   c. **Copak** jsi jiná než ostatní ženy? (Štícha, 2013: 773) – particle
      "*Copak* you are different from other women?"

   d. **Copak**, oni to dnes nehrají? (Havránek *et al.*, 1960: 222) – interjection
      "*Copak,* they aren't performing it today?"[6]

The expressions containing *-pak* are expressive (Komárek *et al.*, 1986: 393) and stylistically marked (*ibid*.: 100). They are frequent primarily in spoken language (Čermák, 2012: 181; Balhar *et al.*, 2011: 570). Particles with the postfix *-pak* are classified as emotional (Komárek *et al.*, 1986: 236), or interrogative contact particles (*ibid*.: 231). Some of them may carry modal meanings, notably the particle *jestlipak,* which marks a question as deliberative, i.e. posed to oneself (Dušková *et al.*, 2012: 313), or simultaneously to oneself and another addressee (Zouharová, 2008); see example (2).

---

[2] Being a predominantly synthetic language, Czech is rich in inflection. The declension system of nouns, pronouns and adjectives comprises seven case forms.
[3] SYN 2015, ORAL, InterCorp Czech. Nor are these forms listed in Filipec, Kroupová *et al.* (2005).
[4] In Czech, particles (*částice*) are traditionally viewed as a separate word class and defined as follows: they are not integrated into the syntactic structure of a clause and they express the speaker's relationship towards the content or form of the communication, to the addressee, etc. (Komárek *et al.*, 1968: 228). Czech particles tend to be emotionally expressive. Many of them are homophonous with members of other word classes.
[5] Unless indicated otherwise, the translations are by Šebestová and Malá.
[6] For the purposes of our analysis, the part-of-speech classification can be disregarded as the major functions tended to occur across part-of-speech boundaries.

(2)   **Jestlipak** sis to už přečetl*?* (Komárek *et al.*, 1986: 231)
       "*Jestlipak* you have read it yet?"

Expressions with the postfix *-pak* may express diverse types of the speaker's stance: different sources list different characteristics, none of them focussing systematically on the semantics of the postfix. Examples of semantic features of various *-pak* expressions mentioned in the literature include: surprise, apprehension (Komárek *et al.*, 1986: 236), admiration, understatement, curiosity, or indignation (Havránek *et al.*, 1960: 222).

      In general, the pronouns, adverbs, particles and interjections ending in *-pak* (3a) may be viewed as elements of what Poldauf (1964) described as "the third syntactical plan", i.e. elements relating the content of an utterance to the individual and "his specific ability to perceive, judge and assess", and expressing the individual's "concern" with the content of the communication or with its form (*ibid.*: 242). Poldauf's findings suggest that English and Czech differ substantially in the means used to express speakers' stance. Czech employs a wide repertory of such means, often morphological or lexical ones, such as free datives, e.g. (3b), various particles or interjections, e.g. (3c) and (3d), respectively). As the part-of-speech classification of some expressions tends to be problematic (e.g. some may fall into different parts of speech depending on their position within the given sentence, they may or may not be syntactically integrated etc.), Poldauf subsumes them under the broader terms "signals" or "formulas".

(3)   a. **Kdepak** asi je? (*ibid.*: 253)
         "Where-*pak* possibly he is?"
         I wonder where he is. / Where is he, I wonder? (*ibid.*)

       b. Čas **mu** utíkal pomalu. (*ibid.*: 249)
         "Time him$_{DAT}$ passed slowly." (*ibid.*: 255)
         He found time pass too slowly. (*ibid.*: 249)

       c. **Že** je dnes dusno? (*že* = particle, syntactically integrated, *ibid.*: 247)
         "*Že* it is sultry today?"

       d. Dnes je dusno, **že**? (*že* = interjection, syntactically non-integrated, *ibid.*)
         "Today it is sultry, isn't it?"

In English, the elements of the "third syntactical plan" appear to be considerably restricted in comparison with Czech, and they tend to comprise mainly grammatical means (e.g. specific syntactic structures). This is partly due to typological differences between the two languages, cf. for example, free datives in inflectional Czech and their analytical counterparts in English (*ibid.*: 248). The difference may be accounted for also by a generally higher degree of expressivity in Czech (*ibid.*: 254).

      Within the "third syntactical plan", we may differentiate between two types of evaluation: "emotional" in (4) as opposed to "intellectual", depending on the basis for the individual's stance (5). According to Poldauf, English shows an overall preference for 'intellectual' evaluation (*ibid.*: 253), frequently to the point of styling expressions of emotional evaluation as intellectual. In example (6), for instance, the speaker's primary aim is to express his feelings, yet the formal representation corresponds to intellectual evaluation (*I wish*).

(4)   **Copak** potřebuje skútr?
       "*Copak* he needs a scooter?" (*ibid.*: 247)

(5)   **Jestlipak** to víte?
      I wonder if you know it? (*ibid*.: 253)

(6)   **I wish** you were here. (*ibid*.)

Both types of evaluation are close to modal meaning (*ibid*.: 244, 247). The modal meaning related to evaluation tends to be epistemic. In example (7a), the speaker expresses his commitment to the truth of a previous statement. In (7b), with the particles *aby tak*, the epistemic modal meaning is deliberative, the speaker is pondering whether it could be raining, while expressing his stance on the possibility of rain.

(7)   a. Je to **myslím** přesně tak. (intellectual, *ibid*.: 244)
         "It is I think exactly so." (*ibid*.: 253)

      b. **Aby tak** venku pršelo. (emotional, *ibid*.: 247)
         "*Aby tak* it was raining outside."

3.   The aims of the study

To return to the topic of the present study, the postfix *–pak* can signal various discourse functions in Czech. Our aim is to offer a comprehensive description of the repertory of these functions. Generally, the same discourse functions can be assumed to be expressed in Czech originals as in their English translation counterparts, since languages share their "needs of expression and communication" (Mathesius, 1936: 95). However, the means of expression as well as the extent to which the functions are explicitly marked are likely to differ in different languages (*ibid*., cf. also Haspelmath, 2010; Martinková, 2014). For instance, epistemic stance and appeal indicated by the particle *jestlipak* in example (8a) is signalled by *I wonder if* in English.[7] In example (8b), no direct overt counterpart of the particle *copak* can be identified.

(8)   a. **Jestlipak** máte ještě tu tlustou knihu?
         "*Jestlipak* you have still that thick book?"
         I wonder if you still have that thick book?

      b. Ale **copak** se to nedalo vymyslit nějak jinak?
         "But *copak* was it impossible to arrange it in another way?"
         But was there no other way to arrange things?

The study pursues two closely intertwined goals: first, a comprehensive description of the discourse functions of sentences containing words with the postfix *-pak*; second, an overview of the means used in English to perform the same discourse functions. The latter aim relates to the wider topic of typological differences between the two languages, bearing in mind the relatively restricted possibilities of emotional evaluation in English (Poldauf, 1964).

Several types of English counterparts are expected to occur as correspondences of the Czech *-pak* expressions, drawing on the literature and the findings of our pilot study

---

[7] Unless indicated otherwise, all examples in Sections 3-6 are from the parallel translation corpus *InterCorp*: the Czech sentences are the originals; their English counterparts (translations) are given below them. Where necessary, literal English translations of the Czech sentences (by Šebestová and Malá) are inserted between the originals and translations, and marked by inverted commas.

(Šebestová and Malá, 2016). For example, Czech questions introduced by the particle *copak*, as in (9), are often translated by polar rhetorical questions (Dušková *et al.*, 2012: 316).

(9)   **Copak** chceš být vyloučen ze školy?
      "*Copak* you want to be expelled from school?"
      Do you want to be expelled from school? (*ibid*.)

Another type of expected English counterparts is negative polar questions. In English, these questions suggest a change in the speaker's previous assumptions or views; usually, they express the speaker's unpleasant surprise (Dušková *et al.*, 2012: 314).

(10) **Copak** nemůže přijít?
      "*Copak* he can't come?"
      Can't he come? (Peprník, 1984: 30)

Our pilot study showed that the occurrence of explicit markers, typical of the Czech third syntactical plan, is also possible in English: the introductory signal *I wonder if* corresponded to the Czech particle *jestlipak*; see (8a) above, idiomatic expressions of emphatic negation corresponded to the interjection *kdepak*, as in (11).

(11) **Kdepak**, teď už bych nic neufoukal.
      "*Kdepak*, now I could not blow anything."
      Not a hope. Couldn't blow now.

4.   Material and method

The material was drawn from the Czech-English fiction subcorpus of the parallel translation corpus *InterCorp*, version 9. The examined material was limited to Czech originals and their English translations,[8] resulting in a subcorpus of 26 Czech original modern novels (complete texts, published mostly between 1950 and 2010) and their English counterparts. The size of the subcorpus (Czech texts) is 2 708 811 tokens.[9]

In total, 576 Czech concordance lines containing expressions with the postfix *-pak* and their English translation correspondences were analysed. Most of the expressions with the postfix *-pak*, as described in grammars and dictionaries of present-day Czech, were represented in the sample, albeit with different frequencies of occurrence (Table 1). The Czech sentences were classified according to the word class of the *-pak* expression. Since the PoS tagging of the corpus proved unreliable and sometimes at variance with our criteria, the classification was performed manually.

---

[8] It would be interesting to complement the results obtained relying on Czech source texts and English translations by an analysis drawing on the opposite direction, i.e. exploring English originals which were translated into Czech using *-pak* expressions. However, this is beyond the scope of the present paper.
[9] Admittedly, the material used is to some extent problematic as the analysis relies on individual translator's choices. Nevertheless, we believe that the size of the corpus and range of texts (26), authors (15) and translators (21) makes it possible to make some generalisations. For a detailed description of *InterCorp* see http://www.korpus.cz.

Table 1. Representation of words comprising -*pak* (lemmata) in our data.

-*pak* expression	no. of instances	%
copak, cožpak, sopak	312	54.2
kdepak, depak, kdepák, depák	98	17.0
jakýpak, jakejpak	37	6.4
jestlipak	32	5.6
kdopak	23	4.0
jakpak, japak, japa	23	4.0
pročpak	17	3.0
kampak	11	1.9
kdypak	9	1.6
zdalipak	4	0.7
natožpak	2	0.5
kolikpak, kolipak	3	0.3
odkdypak	2	0.3
čípak	1	0.3
kterýpak, kerejpak	2	0.2
Total	**576**	**100**

Where possible, an overt counterpart of the Czech -*pak* expression was identified in the English counterparts of the Czech sentences. The English sentences were classified in terms of sentence type and polarity.

The next step − an analysis of the discourse functions of the English counterparts of the Czech sentences with -*pak* expressions − revealed that several broad functional areas can be identified: epistemic modality, the function of appeal, expressing a change in the speaker's assumption, emotional expressivity, and expressing contrast or emphasis. The functions were often found to combine, making it impossible to tease out a dominant one (cf. Aijmer, 2013).

As suggested by Johansson (2007: 1), some characteristics which may not be quite salient in the originals may be revealed by translation correspondences. In our analysis, the English translation counterparts were found to contain additional markers of other functions (e.g. sentence-initial coordinating conjunctions). We therefore decided to look more closely at the Czech original sentences in order to find out whether the corresponding Czech markers could be seen to co-occur with the -*pak* expressions. Based on this approach, we identified additional markers of establishing and maintaining contact, expressing politeness or tentativeness, and building textual coherence. Sometimes the markers were present merely in the English translations; in other cases, corresponding markers were also attested in the Czech originals.

In summary, our contrastive analysis was performed in two steps: we started out from the Czech originals, and examined their English translation counterparts; secondly, the English counterparts drew our attention towards elements in the originals which were worth further examination.

As a result of the two-step analysis, each sentence in our data was assigned at least one of the following primary functions: indicating epistemic modality, voicing appeal, or expressing a change in the speaker's assumption. These functions may either occur on their own or they may combine with each other and with other, secondary functions, such as establishing or maintaining contact, expressing politeness or tentativeness, and building textual coherence (including emphasis or contrast). The secondary functions never occurred alone in our material.

The classes identified empirically in our data correspond to a large extent to Erman's (2001) categorization of pragmatic markers, attesting to the broad range of pragmatic functions indicated by the Czech -*pak*. However, our classification is more fine-grained and goes across Erman's three broad categories. Erman's "textual monitors" fulfil textual

functions, relating to our secondary functions of expressing contrast/emphasis and coherence. Erman's "social monitors" correspond to our functions of appeal (primary) and establishing/maintaining contact (secondary), "[ensuring] that the channel is open between the interlocutors" (*ibid*.: 1339). Finally, Erman's "metalinguistic monitors" are "basically modal" – they relate to our primary functions of expressing epistemic modality and changes in the speaker's assumption.

The following section first introduces the primary functions of the sentences with the expressions containing *-pak*, as identified through their English counterparts, and the means used in English to convey the same function. Then, the secondary functions are described.

5.   Analysis

As mentioned, three major discourse functions signalled by the postfix *-pak* were identified with the help of the English translation correspondences: indicating epistemic modality, voicing appeal, and expressing a change in the speaker's assumption. We call these functions primary, as all the sentences in our material express at least one of these functions. Primary functions may either occur on their own or in combination with secondary or other primary functions, as in example (12), which signals the speaker's changed assumption: the speaker was convinced that a third person knew something, but the addressee's reaction leads the speaker to question his original assumption. At the same time, the sentence voices the speaker's appeal to the addressee to provide a clarification, as reflected in the introductory signal *you mean*.

> (12)   **Copak** von to neví?
>   "*Copak* he doesn't know?"
>   **You mean**, like, he doesn't know?

5.1   Primary functions

5.1.1 *Expressing epistemic modality*
Epistemic modality may be defined as "[the] speaker's attitude to the factuality of past or present time situations", and "qualifications concerning the speaker's knowledge" (Huddleston and Pullum, 2012: 178). Biber *et al.* (1999: 485) term it "extrinsic modality", and define it as "[referring] to the logical status of events or states, usually relating to assessments of likelihood: possibility, necessity, or prediction." Dušková *et al.* (2012: 477-478) list the following epistemic modal meanings: "probability, certainty, doubt based on observation, validity based on other people's beliefs, and limiting the truth value of an utterance".

The definitions of epistemic modality are relatively broad. The English sentences in our material were considered to express epistemic modal meaning if they contained an expression which we viewed as an epistemic modality marker. A list of markers was compiled after examining the corpus material and searching for any expressions which overtly convey epistemic modal meaning, bearing in mind that possible means of expressing modality, other than modal verbs, include lexical verbs, adjectives, adverbs, and nouns (Huddleston and Pullum, 2012: 52, 173-5) as well as certain lexico-grammatical patterns (cf. Daneš *et al.*, 1987: 355). Thus, the markers examined include epistemic modal verbs or modal adverbs, question tags, emphatic *no* negation (epistemic meaning of certainty), and mental verbs.

Question tags may be considered signals of epistemic modality as they "elicit confirmation or agreement" (Biber *et al.*, 1999: 208), i.e. they voice appeal, and, at the same

time, they imply either the epistemic modal meaning of uncertainty (requesting verification) or certainty (requesting agreement), as illustrated in (13).

(13) **Copak** se to nedá napravit?
"*Copak* can't it be fixed?"
Well, that can be put to rights, **can't it**?

Mental verbs also tend to co-occur with modals, expressing "emotions, attitudes, cognitive states" Biber *et al.* (1999: 491); Poldauf (1964: 244) also notes that intellectual evaluation is closely related to modality, cf. (14).

(14) **Copak** já sem debilní?
"*Copak* I am a moron?"
You think I'm a moron?

Epistemic modal meaning seems to be an intrinsic feature of the postfix *-pak*, as it was identified in all examined parts of speech.

As shown in Table 2, the English counterparts pointing towards the primary function of expressing epistemic modality were predominantly interrogative sentences (over 60% of all concordances in which epistemic modality was identified).

Table 2. English counterparts expressing epistemic modal meaning.

English counterpart	no. of instances	%	example
interrogative positive non-rhetorical sentence	63	25.0	**Copak** si šlo Doufala splíst? "*Copak* it was possible to mistake Doufal?" How could you take Doufal for somebody else?
interrogative positive rhetorical sentence	53	21.0	**Copak** vím, kdo k panu asistentovi chodí? "*Copak* I know who comes to see the lecturer?" How do I know who comes to see the lecturer?
verbless clause	39	15.5	Ale **kdepak**, to je poctivý člověk. "But *kdepak*, he is an honest man." **But of course not**, he is an honest man.
interrogative negative rhetorical sentence	29	11.5	**Cožpak** jsem vám to neřekl hned, když jsem vás uviděl? "*Cožpak* I didn't tell you…?" Didn't I tell you the moment I set eyes on you?
declarative negative sentence	28	11.1	**Copak** se máte tak zle? "*Copak* you are that badly off?" You're not that badly off.
declarative positive sentence	26	10.3	**Zdalipak** já budu s to ještě někdy dobýt dívku. "*Zdalipak* I will be ever able to conquer a girl." I wonder if I'll ever conquer a girl...
interrogative negative non-rhetorical sentence	9	3.6	Ale **copak** se to nedalo vymyslit nějak jinak? "But *copak* it could not be arranged another way?" But was there no other way to arrange things?
exclamative sentence	3	1.2	**Jakápak** záchytka! "What-*pak* sobering-up centre!" No centre for him!
Imperative sentence	2	0.8	**Depák**, to byste se hoší krutě přepočítali! "*Depák* you would be brutally miscalculating, boys!" Don't kid cherselfs now, boyos, cause that'd be brutally miscalculatin!
Total	**252**	**100**	

Some of the most prominent English counterparts contained epistemic signals, namely modal adverbs, e.g. *really* and *actually*, as in (15), verbs of thinking, e.g. *think* in (14) above), and modal verbs, cf. *could* in (16), which has no direct counterpart in the Czech original).

(15) Jana Rybářová se rychle probírala k vědomí, čeká nás důležitý den, ale **copak** nějaký den není důležitý?
"…but *copak* some day isn't important?"
Jana Rybářová quickly roused herself from sleep, an important day awaits, but are there **actually** unimportant ones?

(16) A **jestlipak** by i to, co Blběnka s Lídou asi dělávaly, než se Blběnka vyvdala za oceán, probudilo v páně Zawynatchovi jeho masochistický princip slasti.
"And *jestlipak* would even what Dotty and Lida's used to do… have awakened the masochistic pleasure principle in Mr. Zawynatch?"
And **could** Dotty and Lida's probable profession have awakened the masochistic pleasure principle in Mr. Zawynatch?

Some of the interrogatives were rhetorical questions. Although formally interrogative, they had the illocutionary force of emphatic statements (Dušková *et al.*, 2012: 316).[10] In (17), the Czech sentence does contain a modal verb (*mám*), but the translation deviates from its usual correspondence (*should*): *mám* is translated as *can*. The meaning of the Czech *jakpak* is made explicit in the English translation; in this rhetorical question, the modality of *mám* is epistemic, the speaker is making an emphatic statement: "I cannot possibly be cross with the boy."

(17) **Jakpak** se **mám** na hocha zlobit, když byl v právu?
"*Jakpak* should I be cross with the boy, when he was in the right?"
How can we be cross with the boy, when he was in the right?

Approximately half of the English interrogative sentences (32% of the total number of interrogatives) were rhetorical questions (cf. footnote 9, Dušková *et al.*, 2012: 316). Polar rhetorical questions function as statements of the opposite polarity, e.g. (18).[11] Variable rhetorical questions, ((15) and (19) are paraphrasable by statements containing a reversed-polarity universal quantifier (Dušková *et al.*, 2012: 316)).

(18) **Cožpak** jsem vám to neřekl hned, když jsem vás uviděl?
"*Cožpak* I didn't tell you the moment I saw you?"
Didn't I tell you the moment I set eyes on you?
(implied meaning: I did tell you the moment I set eyes on you.)

(19) **Copak** vím, kdo k panu asistentovi chodí?
"*Copak* I know who comes to see the lecturer?"

---

[10] The classification of negative interrogative sentences may be problematic (cf. Dušková *et al.*, 2012: 314-317). In our study, we only consider interrogative sentences rhetorical if they contain an epistemic modality marker, or if the context suggests clearly enough that their illocutionary force is an emphatic reversed-polarity statement. Other instances are viewed as merely expressing a change in the speaker's previous assumption (cf. *ibid.*: 314).

[11] We adopt Huddleston and Pullum's terminology, i.e. polar questions (allowing as its answers a pair of polar opposites, Huddleston and Pullum 2012: 868) as opposed to variable questions, containing "a propositional content consisting of an open proposition, i.e. a proposition containing a variable [...] The answers express closed propositions derived by substituting a particular value for the variable" (*ibid.*: 872).

How do I know who comes to see the lecturer?
(implied meaning: There is no way for me to know that.)

Verbless clauses were also represented among the English counterparts expressing epistemic modality. All of them were counterparts of the Czech interjection or particle *kdepak*, functioning as an emphatic negative response (a variant of English emphatic *no*). Usually, these verbless clauses contained an emphatic element, e.g. emphatic negation, an element fronted through negation (20), idiomatic phrases, such as *no way, some hope*, and *by no means*, as in (21).

(20) …o věcech Boga jsem se ani nezmínil, **kdepak**, já byl rád, že tu můžu ležet…
"…the stuff about Bog I didn't even mention, *kdepak*, I was glad to lie here…"
I didn't even mention that stuff about Bog, **not me**, I was glad to be there…

(21) **Kdepak**, tady umírají především mladí lidé.
"*Kdepak*, here mostly young people die."
By no means. The highest death rates here are among young people.

Another group of English counterparts were negative declarative sentences. They express negative epistemic modal meaning, i.e. the speaker's certainty about an explicit negative statement, proving that the Czech original interrogative sentences are indeed rhetorical, as shown in (22).

(22) **Copak** se máte tak zle?
"*Copak* you are that badly off?"
You're not that badly off.

Epistemic modality may also be expressed by positive declarative sentences (they express the speaker's certainty about the truth of the statement). Many of them contained the verb *wonder* (mostly the initial signal *I wonder*), expressing deliberative meaning (23).

(23) **Jakpak** asi skončí tamta partie, pohlédla ke schodišti.
"How-*pak* will that game end, she looked towards the staircase."
Bridge, is it? the blonde thought, and **wondered** how the other game would turn out.

Epistemic modal meaning may combine with the function of appeal in two types of counterparts, namely in deliberative questions and emphatic statements. Deliberative questions posed to oneself (and potentially also to another addressee) express the epistemic modal meaning of uncertainty, as in (24). Here there are explicit appeal signals, namely the term of address *you bums,* the second person verb form and in English also the possessive pronoun *your*. On the other hand, emphatic statements appealing to the addressee to acknowledge the speaker's assertion express a high degree of certainty: example (25) has the illocutionary force of an emphatic statement (the speaker is convinced that 'we do not have to be like that') as well as containing an appeal signal, the inclusive plural *we*.

(24) **Jestlipak** jste, *vy syčáci*, nezapomněli otčenáš?
"*Jestlipak* you haven't, you bums, forgotten 'Our Father'?"
Could it be, **you bums**, that you have forgotten **your** 'Our Father'?

(25) **Copak** musíme být jak pekař s pekařkou na peci?
      "*Copak* we have to be like the baker and his wife on the stove?"
      Do we have to be like the baker and his wife on the stove?

In idiomatic constructions containing a -*pak* expression and an 'echo' element, the epistemic modal meaning combines with the cohesive function of -*pak* as *jakpak by ne*, or *jakýpak* + 'echo' (26). We are using the ad-hoc term 'echo' element here to refer to a recurrence of any expression from the preceding cotext, such as the word *offense* in example (26). This recurrence is not necessarily verbatim, i.e. the 'echo' element may be, e.g., an anaphoric pronoun. A certain degree of variation within the 'echo' element is possible, cf. the adjective *embarrassing* echoed by the corresponding noun *embarrassment* in (27).

(26) To všechno jsou ovšem jednotlivosti; ale stačí je osvětlit vaším dnešním, přítomným **deliktem**, aby se náhle spojily v celek výmluvně svědčící o vašem charakteru a vašem postoji. – **Ale jakýpak delikt**, křičel jsem. Vyložím přede všemi věci tak, jak se odehrály: jsou-li lidé lidmi, musí se tomu přece smát.
      "…but just look at them in the light of your present offense… – But what-*pak* offense, I shouted. …"
      All these, of course, are isolated facts; but just look at them in the light of your present **offense**, and they suddenly unite into a totality of significant testimony about your character and attitude. – **But what sort of offense!** I'll explain publicly what happened. If people are human they'll have to laugh.'…'

(27) "Představ si, jak by to bylo **trapné**, kdybychom nepřišli," řekl jsem. – „**Copak trapné**, ale přišli bychom o Dvořákův violoncellový koncert!"
      "'Imagine how embarrassing it would be if we didn't come,' I said. - '*Copak* embarrassing, but we would miss Dvořák's cello concerto!'"
      "Just imagine how **embarrassing** it would be if you hadn't remembered and we didn't turn up," I said. – "**Never mind the embarrassment**, think of the Dvořák's cello concerto we'd be missing!"

5.1.2 *Voicing appeal*
Appeal is understood as prompting the addressee to react to and to become actively involved in the interaction – to "do" something verbally. In fact, the interrogative sentence type automatically entails a certain basic amount of appeal expressed by the utterance; any question presents an appeal to the addressee to provide the missing information (Dušková *et al.*, 2012: 311). Therefore, the interrogative sentence type on its own was not regarded as a sufficient criterion for a sentence to classify as expressing appeal. Other indicators of appeal had to be present.

      The presence or absence of an addressee is a crucial factor in questions. If the question is posed to an addressee, its function is that of appeal. On the other hand, if there is no second person present, the speaker poses the question to himself, i.e. the question is deliberative and dubitative, and it expresses epistemic modality. Therefore, we looked for second-person signals in the English counterparts. Considering the communicative contexts, we have arrived at the following set of appeal markers: vocatives, second person finite verb forms,[12] second person pronoun as affected object, as in (28), second person possessive pronoun (29),

---

[12] Unless the 2nd person is a general human agent, in which case the question is rhetorical: *Copak dnes najdeš někoho, kdo by měl trochu odvahy?* 'How often nowadays do you find someone with some courage?'

inclusive plural, e.g. (25) above, question tags (30), verbs referring to the ongoing communication, as *I ask you* or *tell me* in (31).

(28) **Copak <u>ti</u>** udělali ti hoši?
"What-*pak* have these boys done to you?"
What have these boys done **<u>to you</u>**?

(29) **Copak <u>tvuj</u>** táta nebyl vlk?
"*Copak* your dad wasn't a wolf?"
Wasn't **<u>your</u>** dad a wolf, **then**?

(30) Počítač je nám, jako ve většině případů, na nic, ale **copak** neznáme tradiční metody, jak pracovat s fotografií?
"…but *copak* we don't know the traditional processing methods?"
As in most such cases, computers are a fat lot of good. But there are still the good old-fashioned processing methods, **aren't there?**

(31) "**Kdopak ti** udělal monokl?" obrátil se k paní Venuši.
"Who-*pak* gave you the monocle?..."
'**Tell me**, who gave you that monocle?' he turned to Mrs Venus.

In the English counterparts, the function of appeal was fulfilled primarily by interrogative sentences; the majority of them were non-rhetorical, as in (32). Variable questions occurred more frequently than polar ones, and positive more often than negative ones.

(32) **Pročpak** jste napadl na [*sic*] toho plešatého pána?
"Why-*pak* did you attack the bald gentleman?"
Why is it that you attacked the bald gentleman?

Some examples contained mental verbs in the second person, e.g. the introductory signal (*do*) *you mean,* as in example (12), repeated here as (33). Others included emphatic elements: *whever, why on earth*, the intensifier *really*, *it*-clefts or inferential constructions (Delahunty, 1995) of the type *could it be that* (34).

(33) **Copak** von to neví?
"*Copak* he doesn't know?"
You mean, like, he doesn't know?

(34) **Jestlipak** znáte časopis Svět zvířat?
"*Jestlipak* you know the magazine The Animal World?"
Could it be that you know the magazine The Animal World?

The appeal may be voiced explicitly in imperative sentences, or in verbs with meta-communicative reference, e.g. *I ask you* in (35).

(35) **Copak** je to možné?
"*Copak* is it possible?"
I ask you, is it possible?

Appeal may also be voiced by a rhetorical question (with the illocutionary force of a reversed polarity statement) − all the examples were polar questions, as in (36). There were also instances of additional emphatic elements: an *it*-cleft (36) and the intensifier *really.*

(36) **Copak** jsi mi tuto větu ve svém automobilu z ledu sám nevytetoval na stehno?
"*Copak* you didn't tattoo that sentence…?"
Wasn't it you who tattooed that sentence on my thigh in your automobile of ice?

In declarative sentences, appeal often co-occurs with deliberative epistemic modal meaning. Signals of deliberative meaning include *perhaps*, *I wonder*; see example (37)*,* or other mental verbs. Sentences introduced by *I wonder* sometimes end with a question mark (9 out of the total 19 in our material). Nevertheless, we classify all of them as declarative sentences, based on their formal characteristics (word order). The optional question mark most likely reflects their illocutionary force, e.g. (37).

(37) Aha, a **jestlipak** víš, že největší herec všech dob byl Charles Laughton?
"Oh right, and *jestlipak* you know…?"
Oh, right, but **I wonder** if you know that the greatest actor of all time was Charles Laughton?

The function of appeal may co-occur with the expression of change in the speaker's previous assumption, which usually concerns the addressee, i.e. there tends to be a term of address or a vocative, as in (38).

(38) **Copak** vy zase nepatříte ke společenský smetánce, **pane profesore**?
"*Copak* you don't belong to the cream of society again, professor?"
Don't you belong to the cream of society again, **professor**?

Table 3 summarizes the different English counterparts of -*pak* voicing an appeal and their distribution in our data.

Table 3. English counterparts voicing appeal.

English counterpart	no. of instances	%	example
interrogative non-rhetorical positive sentence	89	53.9	**Copak** byl domov ještě domovem? "*Copak* home was still home?" Do you think that home was still home?
interrogative non-rhetorical negative sentence	38	23.0	**Copak** nevíš, jak tě mám rád? "*Copak* you don't know how much I love you?" Don't you know I love you?
declarative positive sentence	13	7.9	**Jestlipak** vůbec víš, že tvůj děda původně pocházel z vesnice, která se menuje Vlčeves. "*Jestlipak* … you know that your grandfather…?" I wonder if you know that your grandfather originally came from a village called Vlčeves
interrogative rhetorical positive sentence (all polar)	8	4.8	**Copak** se mi chtělo? "*Copak* I wanted to?" Do you think I wanted to leave?

interrogative rhetorical negative sentence (all polar)	6	3.6	Ale **copak** vy sám většinou nemluvíte, jenom abyste mluvil? "But *copak* you yourself mostly don't talk just for the sake of talking?" Don't you yourself talk mostly just for the sake of talking?
imperative – positive sentence	6	3.6	…tělo neměl zhrublý a ztěžklý svejma bitvama a už vůbec ne chlastem, **kdepak**, sportoval. "…his body wasn't made coarse and heavy by his battles… *kdepak* he was a sportsman." …his body wasn't all coarse an hard from battle, an forget about booze, this boy was an athlete.
declarative – negative sentence	3	1.8	…**copak** máš pas? "…*copak* you have a passport?" you don't even have a passport!
imperative – negative sentence	2	1.2	**Cák** já. "What-*pak* me." But don't take no account of me.
Total	**165**	**100**	

5.1.3 *Change in assumption*

The third main discourse function signalled by *-pak* is that of expressing a change in the speaker's previous assumption. It is strongly linked to emotional expressivity (especially in cases when the speaker expresses an unpleasant surprise). Virtually all the English counterparts here were interrogative sentences (see Table 4), mostly polar questions. A case in point is example (39) where the second speaker had presupposed that the cousin was rich. This assumption appears to be false, forcing the speaker to reassess the situation.

(39) On má bratránek pro dámy veliké kouzlo. Dámy ho mají za boháče. – **Copak** není bohatý?
"…Ladies take him for a rich man. – *Copak* he isn't rich?"
My cousin has a great appeal for the ladies. They think he's rich. – **And isn't he?**

Table 4. English counterparts expressing a change in the speaker's assumption.

English counterpart	no. of instances	%	example
interrogative positive non-rhetorical sentence	44	51.2	**Copak** von to neví? "*Copak* he doesn't know?" You mean, like, he doesn't know?
interrogative negative non-rhetorical sentence	38	44.2	On má bratránek pro dámy veliké kouzlo. Dámy ho mají za boháče. - **Copak** není bohatý? "…Ladies take him for a rich man. – *Copak* he isn't rich?" My cousin has a great appeal for the ladies. They think he's rich. - And isn't he?
declarative sentence	4	4.7	**Copak**, snad se nebojíte? "*Copak*, hopefully you aren't scared?" You're not scared, are you?
Total	**86**	**100**	

As regards co-occurring elements, verbs of thinking were found, such as in *you think* in example (40) and *you mean* in (33) above. There were also some emphatic means identified in interrogatives, such as *wherever, where on earth,* the intensifiers *really, at all* as in (41), or the inferential construction (*is it that* as in (40)).

(40) Ty vopice jedna, **copak** myslíš, že se budu jen s tebou bavit?
"…*copak* you think that I'd be prattling with you only?"
You singular monkey, **is it that <u>you think</u>** that I'd be prattling with you?

(41) **Copak** nemáš ani trochu slitování?
"*Copak* you don't have a bit of pity?
Have you no pity **at all**?

5.2 Secondary functions

Secondary functions of the postfix *-pak* can be identified on the basis of some additional signals. Sometimes, these signals are present only in the English translation, having no direct counterpart in the Czech original. This suggests that the postfix *-pak* has a certain potential to fulfil the given function. This potential would be difficult, if not impossible, to identify if we only analysed the original texts (cf. Johansson, 2007). This can be exemplified by the occurrence of cohesive ties in the English translation, pointing towards the ability of *-pak* to contribute to textual cohesion. For example, in (42), the English conjunction *and* has no explicit counterpart in the original.

(42) **Kdepak** ses tu vzala, slečinko?
"Where-*pak* have you come from, little lady?"
And where **may** you have come from, little lady?

However, in other cases these additional signals are present in the English translation as well as in the Czech original. Here, we adopt a perspective based on Partington's (2015) theory of evaluative harmony. The potential of the postfix *-pak* to fulfil the given function is strengthened by its co-occurrence with another signal of the same function, e.g. as in (42), where the potential of *-pak* to serve as a means of expressing politeness is supported by the presence of the honorific *slečinko / little lady*.[13] When combined, the co-occurring signals reinforce each other's potential. Simultaneously, they create evaluative harmony, contributing to textual cohesion.

From a methodological point of view, it may be worth noting that the two types of signals described above (those limited to English translations and those occurring in both the translation and the original) can only be linked through contrastive analysis. The secondary functions could hardly be reliably identified in any other way than through a translation corpus study. Moreover, a relatively large quantity of material is needed to arrive at plausible generalisations, as the secondary functions are not ubiquitous, unlike the primary ones.

5.2.1 *Contact function*
The function of establishing or maintaining contact was ascribed to *-pak* based on English translations which contained clearly identifiable, explicit contact signals. These were defined as terms of address (*Sir*, *Miss*), honorifics (*your eminence*), vocatives, contact interjections (*come on, hey, look here*), and greetings (*good afternoon*). The contact function is linked to

---

[13] Even though the vocative may appear ironic here, the co-text rather suggests a "good-natured teasing" interpretation: *A long stick of a man, stooping a little, he had a bald head and a good-natured teasing grin, rather like the mysterious old man who appears in fairy tales. "And where may you have come from, young lady?"*

the function of appeal as well as to emotional expressivity, of which some terms of address were evaluative, functioning as stance markers, e.g. (43).

(43) **Pročpak**, vy jeden siamskej slone, nemyslíte?
     "Why-*pak* don't you, you Siamese elephant, think?"
     How come, you Siamese elephant you, that you don't think?

5.2.2 *Politeness/tentativeness*

The role of *-pak* as a politeness or tentativeness marker was suggested by its co-occurrence with honorifics (therefore coinciding with the contact function), such as *slečinko / little lady* in example (42). Further, some morphological politeness signals were identified in English (the epistemic modal *may* in (42); past tense in (44)). Unlike the other functions, which were distributed evenly across different word classes, the function of a politeness signal was most common in *-pak* interjections. This is exemplified in (45), which contains an explicit comment suggesting the polite tone of the utterance not present in Czech (*he said modestly*).

(44) […] řekla jsem Ludvíkovi, **jestlipak** víte, že jedu za tři dny na Slovácko dělat reportáž o Jízdě králů.
     "[…] I said to Ludvík, *jestlipak* you-know that I am-going in three days to Slovácko to do a feature on the Ride of the Kings?"
     […] I said to Ludvik, **did you know** I **was** going to Moravia for three days to do a feature on the Ride of the Kings?

(45) Vy jste umělec, důstojný pane, řekl jsem. - Bránil se: Ale **kdepak**, pane profesore. To já si s tím jenom tak hraju, když mám trošku času.
     "…But *kdepak*, professor. I just tinker around…"
     You're an artist, Father, I said. - **Oh now**, Mr. Smiricky, he said **modestly**. I just tinker around for fun when I have a little time.

5.2.3 *Cohesive function*

The role of the postfix *-pak* in structuring the text and establishing relationships within the discourse was indicated by linking devices (most frequently the coordinator *and*) and by several instances of English discourse markers in the translations which had no direct counterparts in the Czech originals, viz. *then, so, well,* and *now* (functioning as conversation openers, as *so* in (46)).

(46) **Kdypak** vy jste měli fáro?
     "When-*pak* did you have a car?"
     So when did you have your own wheels?

The cohesive uses of *-pak* included those establishing a relationship of emphasis or contrast, represented mostly by the particle *copak* and the interjection *kdepak* in (47). This particle tends to be followed by an 'echo' element, thus supporting the cohesion by lexical repetition.

(47) **Strejček** chrápal, až se vohejbaly divizny, [...] a snažili se ho křísit, jenže **kdepak strejček**, ten chrápal a chrčel a slintal a vodfukoval...
     "The uncle was snoring…, and they tried to revive him, but *kdepak* the uncle, he snored…"

> The **uncle** was snoring so hard it was making the mullein plants bend over, [...] and they tried to revive him, but the **uncle** just kept snoring, his throat rattling and he was drooling and exhaling loudly...

Finally, through their emotional expressivity, *-pak* expressions contribute to evaluative harmony, or the tendency of elements sharing the same evaluative polarity, i.e. positive or negative, to co-occur and create consistent cohesive evaluative "chains" throughout texts, as in (48), which serves as another means of textual cohesion (Partington, 2015: 283-4).

(48)   Yveta Trojanová, dcera toho **sviňáka** [offensive], [...] proč ona může mít, na co jen ukáže prstem - na tom prstě safír v platině za pět **papírů** [slang, expressive] - vozit si **prdel** [vulgar] v auťáku a každoročně letadlem k moři, **copak** ona má díru do **zadku** [informal] jinde než já?
"…to the sea, *copak* she has a hole in her butt different from mine?"
Yveta Trojanová, the daughter of that **pig** [offensive], [...] why does she get whatever she points her finger at, a finger with a platinum ring with a sapphire that cost five **grand** [slang, expressive] – she drives her **ass** [vulgar] around in a car and flies every year to the sea, is the hole in her **butt** [informal] any different from mine?

6.   Conclusions

The aim of the present study was to contribute to the description of the Czech postfix *-pak* through examining its English translation correspondences with respect to discourse functions. The results of our contrastive analysis suggest that the postfix *-pak* is a polyfunctional indicator of discourse functions. The major functions of the postfix are to mark epistemic modality (such as certainty in rhetorical questions or deliberative meaning in dubitative questions), voice an appeal to the addressee, and mark a change in the speaker's previous assumption. These functions tend to occur together. Beside these primary functions, the postfix has also manifested the ability to mark other pragmatic meanings. These functions are termed secondary as they occur only in combination with the primary ones. The secondary functions were establishing/maintaining contact, signalling politeness and marking textual cohesion. Finally, the study confirmed that the postfix *-pak* tends towards emotional expressivity, and occurs frequently in spoken interaction (all our examples come from direct speech in fiction dialogues).

The present study has also provided an overview of the means which may be used in English to convey the same functions as those signalled by the Czech postfix *-pak*. The English counterparts of the Czech sentences comprising words with the *-pak* postfix constitute a scale ranging from specific sentence types (e.g. negative rhetorical questions, exclamative sentences) via lexico-grammatical structures of varying degrees of fixedness (*I wonder if*) to individual lexical markers of the discourse functions (e.g. intensifiers, lexical negators). In both Czech and English, discourse function indicators have shown a preference for clause-initial position. In English this applies, for instance, to conjunctions reinforcing the contact-maintaining function (*and*), interrogative sentence-opening expressions (e.g. *I wonder if…*, *Is it that…*), or negative idiomatic constructions (e.g. *not a hope*). The analysis also suggests that where the Czech particles indicate negative epistemic modal meaning (certainty about negative polarity) or discourse functions of objection, reproach, disagreement etc., English tends to express the negative meaning explicitly (e.g. by negative declarative clauses or introductory negative expressions, such as *not at all*). The study further supported the assumption that the expression of stance and interpersonal functions in English relies

more on grammatical means, though lexical ones are involved as well (introductory signals such as *I wonder*, discourse markers and conjunctions).

Finally, from a methodological perspective, the present study has shown that a combination of methodological approaches is required to provide a comprehensive overview of the functions which the postfix *-pak* may signal in discourse. The unidirectional contrastive approach proved efficient as a starting point, but when complemented by a focussed analysis of individual recurrent signals, it led to a more exhaustive characteristic of the postfix, including its potential (secondary) functions.

References

Aijmer, K. 2013. *Understanding Pragmatic Markers: A Variational Pragmatic Approach*. Edinburgh: Edinburgh University Press.

Balhar, J. *et al.* 2005. *Český jazykový atlas 5.* Prague: Academia.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman

Boye, K. 2012. *Epistemic Meaning. A Crosslinguistic and Functional-cognitive Study*. Berlin/Boston: De Gruyter Mouton.

Čermák, F. 2012. *Morfématika a slovotvorba češtiny*. Prague: Nakladatelství Lidové Noviny.

Daneš, F., Hlavsa, Z. and Grepl, M. 1987. *Mluvnice češtiny. 3, Skladba.* Prague: Academia.

Delahunty, G.P. 1995. The Inferential Construction. *Pragmatics* 5(3): 341–364.

Dušková, L., Strnadová, Z., Knittlová, D., Peprník, J. and Tárnyiková, J. 2012. *Mluvnice současné angličtiny na pozadí češtiny*. Prague: Academia.

Erman, B. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 33: 1337–1359.

Filipec, J. *et al.* 2005. *Slovník spisovné češtiny pro školu a veřejnost.* Prague: Academia.

Gast, V. 2015. On the use of translation corpora in contrastive linguistics. A case study of impersonalization in English and German. *Languages in Contrast* 15(1): 4–33.

Grepl, M. and Karlík, P. 1998. *Skladba češtiny.* Olomouc: Votobia.

Havránek, B. *et al.* 1960. *Slovník spisovného jazyka českého I; A – M*. Prague: ČSAV.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3): 663–687.

Huddleston, R. and Pullum, G.K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Johansson, S. 2007. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studie*s. Amsterdam/Philadephia: John Benjamins.

Karlík, P., Nekula, M. and Rusínová, Z. 1995. *Příruční mluvnice češtiny*. Prague: Nakladatelství Lidové Noviny.

Komárek, M., Kořenský, J., Petr, J. and Veselková, J. 1986. *Mluvnice češtiny. 2, Tvarosloví*. Prague: Academia.

Malá, M. 2013. Translation counterparts as markers of meaning. *Languages in Contrast* 13(2): 170–192.

Martinková, M. 2014. K metodologii využití paralelních korpusů v kontrastivní lingvistice. *Naše řeč* 97(4-5): 270-285.

Mathesius, V. 1936. On some problems of the systematic analysis of grammar. *Travaux du Cercle Linguistique de Prague* VI: 95–107.

Partington, A. 2015. Evaluative Prosody. In *Corpus Pragmatics: A Handbook.* K. Aijmer and C. Rühlemann, (eds), 279‑303. Cambridge: Cambridge University Press.

Peprník, J. 1984. *Angličtina pro pokročilé I.* Prague: SPN.

Poldauf, I. 1964.The Third Syntactical Plan. *Travaux linguistiques de Prague 1, Lʹécole de Prague aujourdʹhui*, 241–55.

Šebestová, D. and Malá, M. 2016. Anglické překladové protějšky českých vět s částicemi *copak* a *jestlipak*. *Časopis pro moderní filologii* 98(1): 228–237.

Šmilauer, V. 1969. *Novočeská skladba.* Prague: SPN.

Štícha, F., Vondráček, M., Kolářová, I., Hoffmannová, J., Bílková, J. and Svobodová, I. 2013. *Akademická gramatika spisovné češtiny*. Prague: Academia.

Zouharová, M. 2008. *Slovník frekventovaných syntaktických termínů*. Available at `http://clanky.rvp.cz/clanek/c/Z/1940/slovnik-frekventovanych-syntaktickych-terminu.html/` [last accessed 30 October 2017].

Primary sources

Czech National Corpus - InterCorp 9. Institute of the Czech National Corpus, Faculty of Arts, Charles University, Prague. Available at `http://www.korpus.cz` [last accessed 5 October 2017].

Czech National Corpus - ORAL2013. Institute of the Czech National Corpus, Faculty of Arts, Charles University, Prague. Available at `http://www.korpus.cz` [last accessed 5 October 2017].

Czech National Corpus - SYN2015. Institute of the Czech National Corpus, Faculty of Arts, Charles University, Prague. Available at `http://www.korpus.cz` [last accessed 5 October 2017].

Corresponding author's address

Denisa Šebestová
Department of English Language and ELT Methodology, Faculty of Arts
Charles University, Prague
nám. Jana Palacha 2
116 38 Praha 1
Czech Republic

sebestovadenisa@gmail.com