

Combining Mutual Correspondence and Textual Variation in a Study of English GO and Norwegian GÅ

Signe Oksefjell Ebeling (University of Oslo)

Hilde Hasselgård (University of Oslo)

Abstract

This article expands and nuances the procedures of calculating translation bias (TB) and mutual correspondence (MC) in contrastive studies. Both measures are intended to show how similar, or mutually translatable, two linguistic items are across languages. Traditionally, these procedures have been carried out without attending to variation across individual texts in a corpus. In a contrastive case study of English GO and Norwegian GÅ, we calculate the TB and MC of the fully congruent instances, i.e., instances where the verbs correspond to each other in the same lexicogrammatical patterns, both in the traditional way and in a newly developed procedure where textual distribution is integrated into the measures. The study substantiates the initial hypothesis that the distributive variants of TB and MC—*Distributive Translation Bias* and *Distributive Mutual Correspondences*—are more informative and reliable, as the values more accurately show the range and distribution of correspondences. The study also suggests that the strict definition of congruence adopted (requiring lexicogrammatical correspondence) is better suited to capture details of cross-linguistic similarity than simply to require formal correspondence at the level of word class. This is made evident by the varying degrees of mutual correspondence in the different verb categories, e.g., the generally higher correspondence measures for intransitive GO and GÅ compared to the phrasal verb use.

Keywords: translation bias; mutual correspondence; textual variation; motion verbs; English/Norwegian

1. Introduction

Mutual correspondence (MC), as devised by Altenberg (1999), measures the degree of intertranslatability between items on the basis of bidirectional translation data, and is considered ‘a good measure which

Ebeling, Signe Oksefjell, and Hilde Hasselgård. 2024. ‘Combining Mutual Correspondence and Textual Variation.’ *Nordic Journal of English Studies* 23(2): 6–33. <https://doi.org/10.35360/njes.v23i2.39157>. Copyright (c) 2024 Signe Oksefjell Ebeling, Hilde Hasselgård. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

can be used to relate not just individual forms but also semantic categories and subsystems across languages' (Johansson 2007: 28). Despite the usefulness of reporting and interpreting MC values, as attested in several publications over the years,¹ a potential weakness is that the procedure has traditionally been carried out without attending to variation, i.e., dispersion (see, e.g., Brezina 2018: 11), across the different corpus texts. This paper therefore aims to expand and nuance the procedure of calculating mutual correspondence in contrastive studies by considering variation across texts. We believe that the measure will become more informative and reliable if textual variation is taken into account, while the conclusions based on them will be more robust.

As a case study, and to test the modified procedures for calculating and reporting translation correspondence, we analyse the lexemes *GO* and *GÅ* and their translations in the fiction part of the English-Norwegian Parallel Corpus (ENPC). *GO* and *GÅ* are cognate verbs (see entries in the *OED* and *NAOB*, respectively). They are also given as each other's translations in bilingual dictionaries (ordnett.no), albeit among several. Both *GO* and *GÅ* are syntactically and semantically highly versatile, and have been shown to have both overlapping and non-overlapping uses in English and Norwegian/Swedish (see, e.g., Viberg 1996 and 1999; Cej 2008). Due to their multifunctional nature, they are very frequent verbs in both languages, and thus suitable for an experimental study that relies on commonly used and widely dispersed items.

The analysis considers lexicogrammatical features such as verb form, syntactic pattern, and congruence in translation. The concept of congruence is particularly important when considering mutual correspondence as a measure of the degree of correspondence between items in translation. In the current study, we will introduce a potentially more adequate way of measuring congruence than the one put forth by Johansson (2007: 24). In addition to the formal criterion of identical word class we also require that items enter into the same syntactic patterns to be considered fully congruent (see further section 3.3.2). Example (1) illustrates a fully congruent translation of *gå*: intransitive *gå* in the infinitive, with a following place adjunct (*dit* 'there'), corresponds to intransitive *go* in the infinitive, with a following place adjunct (*there*). In

¹ See, e.g., Altenberg (1999, 2002 and 2007); Ebeling and Ebeling (2015); Hasselgård (2015); Thormodsæter (2021).

example (2), however, the phrasal verb *go on* is noncongruently translated into the Norwegian simplex verb *fortsette* ‘continue’.

- (1) Herman liker å gå dit etter skoletid ... (LSC1)²
Herman likes to go there after school ... (LSC1T)
- (2) He would go on now for weeks ... (MW1)
Han ville fortsette nå i ukesvis ... (MW1T)
‘He would continue now for weeks’

Against this backdrop, we will address the following questions:

- i. Can measures of translation correspondence be more informative when taking textual variation into account?
- ii. Can measures of translation correspondence be more useful if a stricter definition of congruence, based on lexicogrammatical features, is adopted?

The article is structured as follows: section 2 offers some background to the notion of mutual correspondence, including the concept of translation bias, while section 3 is devoted to the case study of GO-GÅ. The study starts with a brief introduction (3.1) before outlining some relevant previous (contrastive) research on the cognates (3.2) and introducing the corpus and the framework of analysis (3.3–3.4). The actual analysis of correspondences is reported in section 3.5. A discussion of the new procedure devised for calculating translation bias and mutual correspondence is offered in section 4. Finally, section 5 provides some concluding remarks as well as some thoughts on potential avenues for future research.

² The corpus ID refers to the author by initials (LSC) and to the text by that author (1). Translations are marked with a T (e.g., LSC1T). Literal translations of Norwegian examples are provided (in inverted commas) only when the translations in the corpus deviate (substantially) from the original (see, e.g., example (2)). For an overview of the texts in the ENPC, see Johansson (2007).

2. Background

The procedure of measuring the degree of translation correspondence between two items by means of a percentage was first introduced by Altenberg (1999). Using semantic classes of adverbial connectors in English and Swedish as his testbed, he proposed two measures, both of which rely on a bidirectional translation corpus. The first is known as *translation bias* (TB) and measures and compares ‘the degree of unidirectional correspondence displayed in the [...] translations’ (Altenberg 1999: 254–255), i.e., the number of times an item A is translated into an item B, calculated and expressed as a percentage. The other—*mutual correspondence* (MC)—takes the two TB values as input and reports on ‘the frequency with which different (grammatical, semantic, lexical) expressions are translated into each other’ (Altenberg 1999: 254), i.e., the number of times an item A is translated into an item B and vice versa, calculated and expressed as a percentage. Essentially, and as illustrated in Figure 1, TB reports on the number of times two items correspond to each other in each direction of translation (e.g. $A_{EO}=B_{NT}$ and $B_{NO}=A_{ET}$),³ divided by the total number of occurrences of each item (A and B) in the original texts (O = original texts), multiplied by 100.

Translation bias	
English → Norwegian	Norwegian → English
$\frac{A_{EO} = B_{NT} \times 100}{A_O} = n\%$	$\frac{B_{NO} = A_{ET} \times 100}{B_O} = n\%$

Figure 1: Translation bias formula (based on Altenberg 1999)

Mutual correspondence, then, merges the two directions of translation into a percentage representing the items’ intertranslatability. Altenberg’s (1999) original formula is given in Figure 2, where A_t and B_t are the number of times item A has been translated into B and B into A, respectively, and, as in Figure 1, A_o and B_o represent the overall number of occurrences of each of the items in the original texts in the corpus. Thus,

³ In Figure 1, English and Norwegian are used as example languages. A_{EO} refers to item A in English original texts, whereas B_{NT} refers to item B in Norwegian translations. Similarly, B_{NO} equals item B in Norwegian originals and A_{ET} equals item A in English translations.

‘the MC value of two cross-linguistic items is the ratio between the number of mutual translations in a bidirectional translation corpus and the sum of their frequencies in the original texts, expressed as a percentage’ (Altenberg 2002: 23).

Mutual correspondence
$\frac{(A_t + B_t) \times 100}{(A_o + B_o)} = n\%$

Figure 2: Mutual correspondence formula (based on Altenberg 1999: 254)

As an example, we can calculate the MC of the cognates HOPE and HÅP(E) (nouns and verbs) in English and Norwegian. With bidirectional data from the ENPC, we find that HOPE occurs 450 times in the English original texts (A_o), and 416 of these (A_t) are translated into HÅP(E) (i.e., $A_{EO}=B_{NT}$ in Figure 1), giving a percentage of 92.4%. The percentage is even higher in the other direction of translation (94.7%), as HÅP(E) in Norwegian originals is translated into HOPE in 356 out of 376 cases, yielding an MC of 93.5%. An MC value as close to 100% as this indicates a very strong cross-linguistic association between these two items.

As outlined above, the measures of TB and MC were originally devised to account for overall tendencies of translation correspondence in a bidirectional corpus, based on the total number of pairings between specific items in the data. In the following, we will carry out a contrastive analysis of the English-Norwegian cognates GO and GÅ in order to elaborate on these measures by adding the dimension of intra-corpus distribution of cross-linguistic items, thus moving away from a measure based on accumulated numbers.

3. Case Study: GO-GÅ

3.1 Introduction: Motion verbs

Motion verbs in general, and ‘go’ verbs in particular, have received considerable linguistic attention from both a monolingual and a multilingual (typological) perspective, and no attempt will be made here to do justice to the vast literature that exists.⁴ In the following, we will

⁴ The search tool for the major collections of Oxford University’s libraries—SOLO (Search Oxford Libraries Online)—returns approx. 8,000 results for the

outline a selection of previous studies that are directly relevant to the current one in the sense that they focus specifically on the cognates GO and GÅ, rather than on motion verbs in general.

3.2 Previous studies of ‘go’ verbs

Fillmore describes GO as a *deictic motion verb*, where ‘the directional complement of the verb ‘go’ indicates a place where the speaker (or encoder) is not located at coding time’ (1975: 50). In other words, the motion typically encoded by GO is directed ‘toward a location which is distinct from the speaker’s’ (Fillmore 1975: 54) at the time of the utterance, as in example (3).

(3) I’ll go there right away. (Fillmore 1975: 55)

Fillmore’s understanding of motion as ‘locomotion’, i.e., being ‘at the location at one time and at another location at another time’ (1975: 51), is clearly central to Viberg’s (1996, 1999) analysis of GO and GÅ in English and Swedish.

Viberg’s work is in turn an important backdrop for the present study of GO and GÅ since Swedish and Norwegian are similar to the extent that Viberg’s findings can be taken as hypotheses for our own comparison of English and Norwegian. Viberg’s studies (1996, 1999) are both based on bidirectional parallel corpora, but the former uses only a small set of texts from a preliminary version of the English-Swedish Parallel Corpus (ESPC, see Altenberg and Aijmer 2000). From the perspective of contrastive lexicology, Viberg classifies the instances of GO and GÅ into three major categories according to their meaning (1999: 90), namely (1) locomotion in concrete space by a human agent, (2) extended meanings (except grammaticalized ones), and (3) grammaticalized meanings.⁵ The first category is considered prototypical, but the languages differ in that the lexeme is specified for ‘on foot’ and ‘without vehicle’ in Swedish, but not in English (1999: 111). The second category includes motion with non-human subject, conventionalized and metaphorical uses (*go to school*, *go*

search strings ‘motion verbs’ and ‘verbs of motion’ (an open search for *motion verbs* returns more than 64,000 results).

⁵ Note that ‘non-human locomotion’ is a separate category in Viberg (1996) but subsumed under (2) ‘extended meanings’ in Viberg (1999).

12 Signe Oksefjell Ebeling and Hilde Hasselgård

behind sb's back = *gå i skolan, gå bakom någons rygg*), and mental uses (*go along with* = *gå med på*). Many of these are typically realized by multiword expressions. It is in the third category—grammaticalized meanings—that the languages differ most (Viberg 1999: 102). English has the future-referring *be going to*, a copular use (as in *go red*) and a continuative use (as in *go on talking*). Swedish GÅ has none of these, but has a modal meaning that English lacks, as in *Det går att öppna dörren* ‘it is possible to open the door’, Lit.: ‘It goes to open the door’ (Viberg 1999: 108). Despite differences in prototypical and extended meanings, GO and GÅ are translated into each other about a third of the time in both of Viberg’s studies, and can therefore be regarded as broadly equivalent (1999: 112).

An interesting finding by Storjohann (2003) is that the specific meanings of motion verbs such as English GO and German GEHEN have not been stable diachronically. For example, the Old English predecessor of GO was largely restricted to human locomotion on foot but extended its meaning after the Middle English period (2003: 167). The fact that two verbs, such as GO and GEHEN, are cognates does not necessarily make them follow the same path of development in different languages, and semantic changes can occur randomly (2003: 202).

The diverging polysemies of cognate verbs, as detailed in the studies cited above, are further highlighted in language-learning contexts. For example, Johansen (2011: 48) finds that most erroneous uses of the lexeme GÅ in L2 Norwegian are due to an overgeneralization of the verb meaning, such as using GÅ for movement not on foot. However, she also points out (2011: 44) that not all uses of GÅ are specified for manner of motion (i.e., ‘on foot’), e.g. *gå på restaurant, gå i kirken* (‘go to a restaurant’, ‘go to church’),⁶ which are idiomatic even in cases where someone travels in a vehicle.

In a predominantly experimental study, Stamenković et al. (2017) ask whether a verb’s perceived prototypicality influences the choice of translation equivalents. Through elicitation and evaluation experiments supported by corpus frequencies, Stamenković et al. arrived at a list of English motion verbs graded for frequency and prototypicality. In a third experiment, translators were asked to give single-word Serbian translations of each verb in the list. A significant correlation was found

⁶ Cf. Viberg’s example *go to school*.

‘between the level of prototypicality [...] and the number of different translation equivalents’ (Stamenković et al. 2017: 89): the more prototypical motion verbs were given fewer translation alternatives than the more peripheral ones.

Drawing on material from the Oslo Multilingual Corpus and inspired by Viberg (1996, 1999), Cej (2008) studies English GO and its translations into German and Norwegian. Based on a sample of 500 instances of the lemma GO, she analyses its uses syntactically (intransitive, copular, multi-word verb and grammaticalized) and semantically (locomotion, metaphorical (extended) and grammaticalized meanings) (2008: 32). Cej then correlates semantic categories with syntactic ones before presenting German and Norwegian translation correspondences accordingly. Overall, and regardless of syntactic and semantic category, she finds that the German and Norwegian cognates of GO—GEHEN and GÅ—are used as translations of GO in about one third of the cases (2008: 100). However, at a more detailed level it becomes clear that the syntactic and semantic categories play a role regarding the level of cognate correspondences. For example, the intransitive (locomotion) use triggers a cognate correspondence in around 50% of the cases in both German and Norwegian, the copular use rarely triggers a cognate correspondence, while the grammaticalized uses (BE *going to* and GO *on*) are never translated into either cognate.

It is expected that similar tendencies to those reported by Viberg (1999) and Cej (2008) will be found in the current case study of GO and GÅ, even if we adopt a slightly different (and stricter) framework of congruence (see section 3.3.2). Also, in our attempt at sharpening the notions of translation bias and mutual correspondence, we will be mainly concerned with the lexicogrammatical patterning of the lemmas. Thus, ‘intransitive locomotion’, as in example (4), and ‘intransitive metaphorical’, as in example (5), will both be subsumed under ‘intransitive’ as, syntactically, both instances of GO are followed by an adverbial realised by a PP.

(4) ... then she calmed herself and *went into the kitchen* ... (AB1)

(5) ... the old people have *gone to bed*. (ABR1)

Importantly, the grammaticalized use of GO in the future-marking semi-auxiliary BE *going to* (Hopper and Traugott 2003: 1–3; Quirk et al. 1985: 144, 214) is excluded from this study. This choice is substantiated by Huddleston and Pullum’s (2002: 210) observation that the construction has lost the meanings of both motion and progressivity. Moreover, previous studies (e.g., Viberg 1999) conclude that Scandinavian GÅ has not grammaticalized in the same way as GO; see Hasselgård (2015: 98) for Norwegian correspondences of BE *going to* as a future marker. However, we include the other grammaticalized use in Viberg’s (1999) classification, i.e., GO *on* as a marker of continuation (cf. Brinton 1985: 160), as it syntactically qualifies as a phrasal (lexical) verb rather than a semi-auxiliary (see Quirk et al. 1985: 1152).

3.3 *Material and data extraction*

The material for the case study is drawn from the fiction part of the English-Norwegian Parallel Corpus (ENPC).⁷ The ENPC is a bidirectional translation corpus, and the fiction component consists of 30 text extracts from the late 20th century in each of the two languages, aligned at sentence level with their published translations. The English and Norwegian sub-corpora are roughly equal in size, with texts originally written in English amounting to ca. 422,000 running words in total and texts originally written in Norwegian to ca. 402,000 words.⁸ The size of the individual text extracts varies from ca. 11,000 to ca. 17,000 words in the English material and from ca. 10,000 to ca. 16,000 in the Norwegian material.

To extract our data set we searched for each individual form of the lemmas through the Translation Corpus Explorer (Ebeling 1998) in the untagged version of the original texts in the corpus.⁹ The reason for this is that a tagger may miss some of the verb uses, due to some verb forms being homographs of nouns and adjectives (see examples 7–9 below). Thus, for example, the Norwegian participle *gående* ‘going’ is mostly tagged as an adjective, and the form *go* may receive a noun tag when it is in fact a verb, as in (6).

⁷ <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/>

⁸ <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/sub-corpora/>

⁹ The search terms were thus *go, goes, going, went, gone; gå, går, gående/gående, gikk/gjekk, gått*.

- (6) The first thing Natalie did after Hilary left in dudgeon was **go_n** round and see Arthur. (FW1)
- (7) *Go* as a noun: Give it a *go*, then. (DF1)
- (8) *Gone* as adjective: The Monster was *gone*, the Monster was dead. (SK1)
- (9) *Går* as a noun: I *går* svarte De ikke ... (KA1)
Yesterday you didn't answer ... (KA1T)

The (untagged) searches left us with a total of 1,603 instances of English GO and 1,840 of Norwegian GÅ. Following manual scrutiny, we excluded false positives, such as (7)–(9), and *going* as part of the semi-auxiliary BE *going to* as well as instances of GO and GÅ that occurred in clauses (i.e., *s*-units) with no translation (2 from English and 27 from Norwegian). This left us with 1,297 occurrences of GO (30.7 per 10,000 words) and 1,719 occurrences of GÅ (42.7 per 10,000 words).

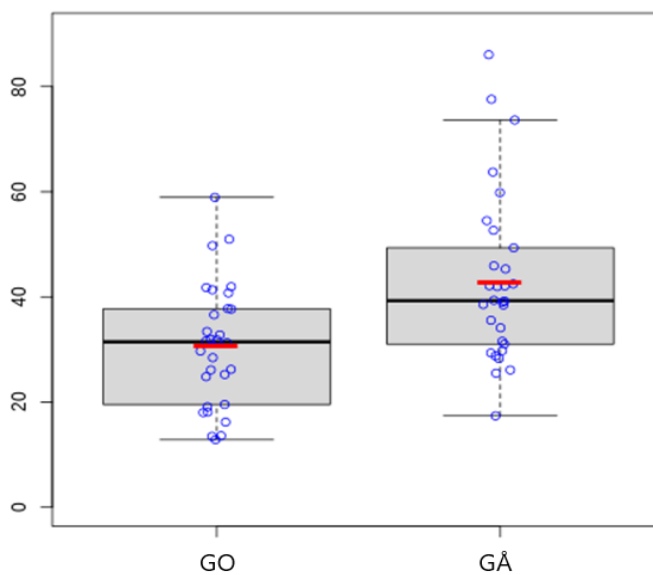


Figure 3: Dispersion of GO and GÅ across original corpus texts, each represented by a blue circle. The y-axis shows frequencies per 10,000 words.

The frequencies of GO and GÅ were also calculated and normalized per text (per 10,000 words) so that we can see the dispersion of these lexical items across corpus texts rather than treating the whole corpus as a bag of words. Figure 3 shows the dispersion of GO and GÅ across the original texts in ENPC fiction.

Norwegian GÅ is markedly more frequent than English GO as attested by the accumulated (normalized) frequencies per 10,000 words reported above. Figure 3 also shows that GÅ has a wider dispersion than GO, with higher median and mean values (the black and the red lines in the box, respectively), as well as a larger gap between the minimum and maximum frequencies per text (the lower and upper lines in the plot, i.e., the whiskers). This may suggest that GÅ has a wider area of use than its English counterpart and therefore potentially a wider range of English translation correspondences. It may be assumed that this will affect the extent of fully congruent correspondences in the two directions of translation (i.e., fewer instances of full congruence when going from Norwegian into English).

3.4 Framework of classification

3.4.1 Classification of the verbs

Rather than applying Viberg's (1996, 1999) semantic categories as the primary classification, this study takes a lexicogrammatical approach, mapping out the patterns in which GO and GÅ appear in the corpus. Each instance of GO/GÅ in original texts has been classified for the following features, of which the first four are most relevant in this study:

- i. **Type of lexeme**, particularly distinguishing simple from phrasal lexemes. The full lexeme was spelt out in the case of phrasal ones (e.g., *go through*).
- ii. **Verb category** (for simple lexemes): intransitive, transitive, copular, existential.
- iii. **Translation**: corresponding lexeme in the translation and aligned s-unit.
- iv. **Congruence** of translation (see distinctions below).
- v. Verb form (in original): infinitive, present, past tense, present participle, past participle, imperative.
- vi. Type of adverbial (place, manner, purpose, other) co-occurring with intransitive simplexes.

- vii. Animacy of subject (animate/inanimate).
- viii. Motion (+/-).

Examples (10) and (11) can serve as illustrations of the classification:

- (10) Hun begynte å gå mot huset. (THA1)
She began to go towards the house. (THA1T)
- (11) As it was, he was likely to go down for seven years. (RR1)
Slik sakene stod, ville han sannsynligvis få sju år. (RR1T)
'As the things stood, would he probably get seven years'

Example (10) has the simple lexeme (i) *gå* in the infinitive (v). *Gå* is intransitive (ii) and occurs with a place adverbial (vi) (*mot huset* 'towards the house'). It has an animate subject (vii) (*hun* 'she') and denotes motion (viii). The translation (iii) of the lexeme is *go*, which occurs in the same syntactic pattern and is therefore fully congruent (iv). In (11) the lexeme is *go down*, a phrasal verb in infinitive form. There is a time adverbial in the clause (*for seven years*), and the subject is animate (*he*). The translation of *go down* is the transitive simplex *få* ('get'), and thus noncongruent.

3.4.2 Classification of the correspondences

Our analysis of congruence is more fine-grained than that of Johansson (2007: 24–25), in which a correspondence (translation or original) belonging to the same formal class (i.e., word class) is considered *congruent*, and other overt correspondences are considered *divergent*. In a study of cognate verbs, however, it seems natural to give cognate correspondences a different status than other verbs. The lexicogrammatical pattern of the correspondence should also be taken into account (see also Ebeling 2017), not just its formal class or its status as a cognate. Hence, we arrived at a five-point scale of correspondence, as outlined in Figure 4.

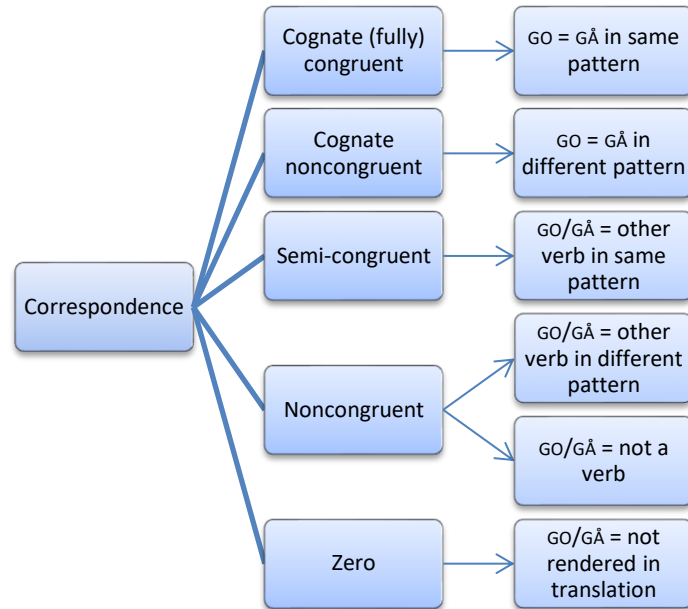


Figure 4: Five-point scale of correspondence

Only those correspondences that are cognate and occur in the same lexicogrammatical pattern in the source and the translation are labelled *fully congruent* (i.e., *cognate congruent*), as in (10) above. If the cognate is used in a different pattern, as in (12), where the phrasal verb *went out* corresponds to intransitive *gikk*, it is labelled *cognate noncongruent*.¹⁰ The use of a different verb in the same pattern, congruent according to Johansson (2007), is labelled *semi-congruent*, as illustrated by (13), where *går* corresponds to *walks*.

- (12) En gang *gikk* lyset. (THA1)
 ‘one time *went* the light’
 Once the lights *went out*. (THA1T)

¹⁰ We prefer the term ‘noncongruent’ to ‘divergent’; see Ebeling and Ebeling (2013: 33).

- (13) Rose *går* foran i taushet. (TB1)
 ‘Rose *goes* in-front in silence’
 Rose *walks* ahead in silence. (TB1T)

The correspondence is noncongruent if the correspondence is a verb with a different complementation pattern, as in (11) above, and similarly if the correspondence is not a verb, as in (14), where the participle *gone* corresponds to the adverb *borte*. Finally, the translation may not contain any trace of the source item, in which case the correspondence is zero, as in (15).

- (14) The blackbird had *gone*. (DL2)
 Svarttrosten var *borte*. (DL2T)
 ‘The blackbird was away’
- (15) Men ikke usårlig, som de *går rundt* og tror. (EHA1)
 ‘... as they *go* round and believe’
 But not as invulnerable as they ___ believe. (EHA1T)

3.5 Analysis of GO and GÅ and their correspondences

We begin by approaching the data in the traditional way by reporting some overall tendencies, in terms of distribution and correspondences of verb categories (3.4.1). Following this preliminary description of our data, we will move on to the core concern of this investigation, which is to further develop the procedure for calculating mutual correspondence in contrastive studies. As our reference point, we will calculate the overall TB and MC of GO and GÅ in our data, without considering dispersion or lexicogrammar, before proceeding to the proposed distributive versions of these measures (3.4.2).

3.5.1 Overall distribution and correspondences of verb categories

At the level of verb category, GO and GÅ seem to have similar distributions in the two languages, as shown in Figure 5.

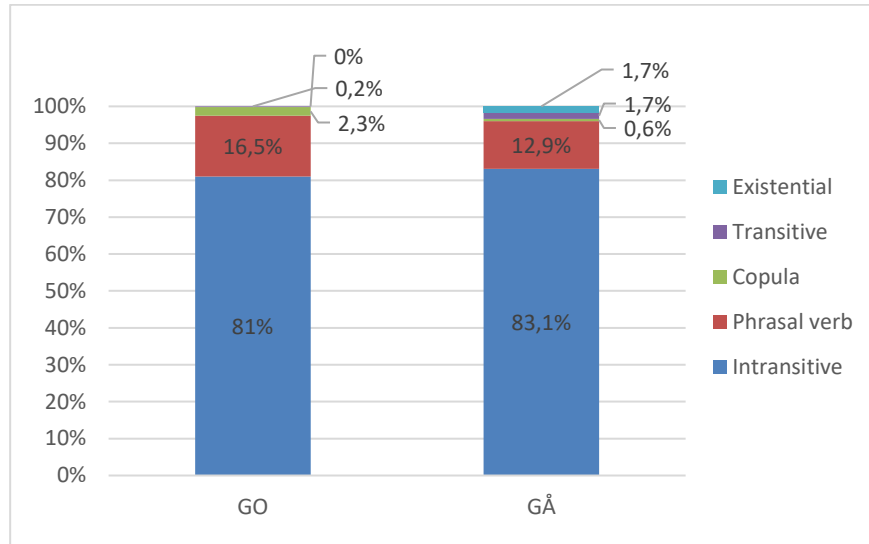


Figure 5: Proportions (%) of GO and GÅ according to verb category

The intransitive use of GO and GÅ is by far the most common one, and proportionally similar in the two languages, accounting for more than 80% of the occurrences. The second-largest category—phrasal verb—is also fairly similarly distributed in the two languages, although it is slightly more frequent for English GO, with 16.5% compared to around 13% for Norwegian GÅ. The remaining uses are at best marginal, although the copular use is slightly more frequently attested in the English data than in the Norwegian data.¹¹

Regarding the congruence of the translation correspondences of the two verbs, there are marked differences between the verb categories. Figure 6 shows that, in the translation direction from English into Norwegian (E-to-N), it is only the intransitive use that features a substantial number of fully congruent (i.e., cognate congruent)

¹¹ Viberg (1999) does not list copular among the uses of Swedish GÅ in his data. However, a search in the Swedish Language Bank (<https://spraakbanken.gu.se/korp/#>) returns examples such as ... *man riskerar att gå bankrutt* (SVT Nyheter 2014) ‘... one risks to go bankrupt’; *gå bankrutt* corresponds to similar instances in Norwegian, e.g.: *Ellers kunne alt gå fallitt* (HW2) *Otherwise their entire enterprise could go bankrupt.* (HM2T)

correspondences: N=452 (raw), or around 43% of the intransitive occurrences.

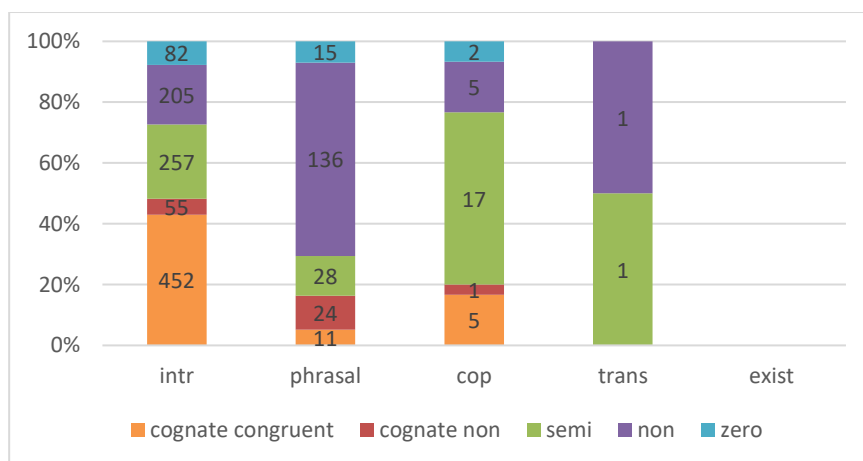


Figure 6: Degree and proportion of congruence across verb categories (E-to-N)

From Figure 6, it is also clear that the other relatively large verb category—phrasal verb—mainly has noncongruent correspondences (N=136, i.e., around 63% of all instances of phrasal GO). Only 11 occurrences (ca. 5%) in this category have a fully congruent phrasal translation with GÅ, one of which is shown in (16). The remaining categories have very few occurrences overall, and the only main trend is that copular GO typically receives a semi-congruent translation, i.e., a copular verb other than GÅ, as illustrated by (17).

- (16) The show *went off* as well as these things do. (JB1)
 Forestillingen *gikk* akkurat så greit *unna* som forventet.
 (JB1T)
 ‘The show *went* just as well *off* as expected’
- (17) Or am I *going* senile and imagining things? (MW1)
 Eller er jeg i ferd med å *bli* senil og innbiller meg ting?
 (MW1T)
 ‘Or am I in course of to *become* senile and imagine myself things?’

A relatively similar picture can be observed in translation from Norwegian into English (N-to-E), as shown in Figure 7. Intransitive GÅ has the highest proportion of fully congruent correspondences (N=550, ca. 46%), whereas phrasal GÅ has the highest proportion of noncongruent correspondences (N=155, ca. 70%).

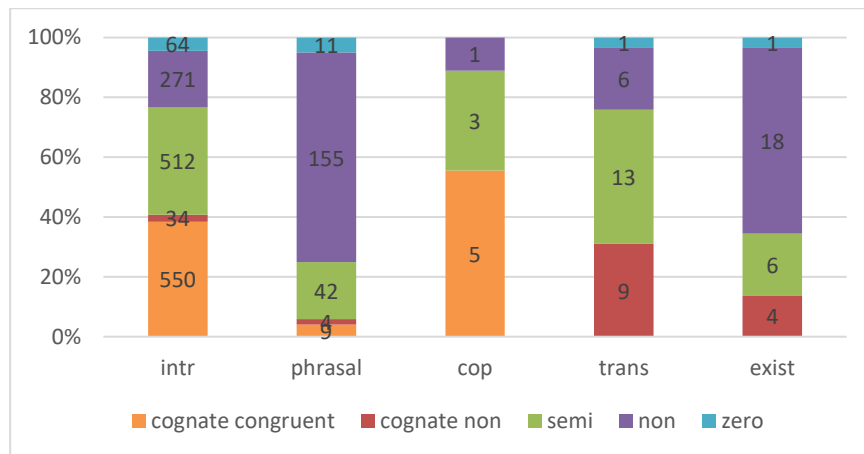


Figure 7: Degree and proportion of congruence across verb categories (N-to-E)

Semi-congruent correspondences of intransitive GÅ are almost as common as cognate congruent correspondences (512 vs. 550, respectively). The most recurrent lexical correspondence in these cases is WALK, which, in contrast to GO, specifies manner of motion as ‘on foot’. An example is given in (18).

- (18) Han *gikk* sakte, det ene beinet hang litt etter. (LSC2)
 ‘He went slowly, the one leg hung a little behind’
 He *walked* slowly, one leg hanging back a little. (LSC2T)

Due to the relatively infrequent use of the other categories, suffice it to say that transitive and existential GÅ are never rendered congruently, as expected from the (near) absence of these patterns for GO in English originals. However, translations sometimes have GO in another pattern than transitive or existential (i.e., cognate noncongruent), as well as semi-congruent correspondences of transitive GÅ and noncongruent ones in the case of existential GÅ.

3.5.2 Translation bias and mutual correspondence of GO and GÅ

As is clear from the data already presented, GO and GÅ are not always translated into each other. Presenting a traditional calculation of TB and MC, Table 1 shows that this happens in roughly one third of the cases when only fully congruent correspondences are considered. The frequency of fully congruent translations is slightly higher going from English to Norwegian than the other way, giving a mutual correspondence of 34.4%.

Table 1: Overall TB and MC of GO and GÅ (fully congruent, all uses)

Translation bias			
Eng → Nor	%	Nor → Eng	%
$\frac{470 \times 100}{1297}$	36.2%	$\frac{567 \times 100}{1719}$	33.0%
Mutual correspondence Eng ↔ Nor			
$\frac{(470 + 567) \times 100}{1297 + 1719}$		= 34.4%	

However, this way of reporting TB and MC masks potential intra-corpus variation: it cannot tell whether the measures are representative of all the corpus texts or of all the lexicogrammatical patterns. To better account for this, we propose a new procedure which will be exemplified with the present data. The first step is to calculate translation bias per corpus text, i.e., the percentage of instances where GO and GÅ correspond fully. The results are shown in Figure 8.¹²

¹² The same technique was used in Hasselgård (2024).

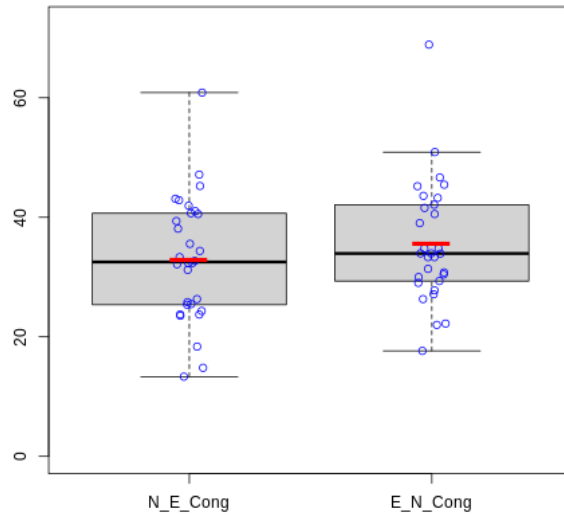


Figure 8: Text-distributed TB of GO and GÅ (cognate congruent, all uses)

Figure 8 shows that the percentage of full correspondence indeed varies quite substantially across texts. The variation is greater in N-to-E than in E-to-N. Because all uses, i.e., all verb categories, of GO and GÅ are included, Figure 8 still masks potential intra-corpus variation in terms of lexicogrammatical patterns. The degree and nature of the correspondence can be further refined by calculating TB per corpus text according to verb category. This aspect is accounted for in Figure 9. Only intransitive and phrasal uses are depicted since the other categories were too infrequent to make the method meaningful. For example, copular GÅ occurs only once or twice in seven of the Norwegian original texts. The TB in each text is either 0 or 100%. For such low frequencies a simple translation paradigm, i.e., the set of forms that correspond to an item in the original text, (Johansson 2007: 23; Levenston 1965) might be more informative; however, this level of detail lies outside the scope of the present study.

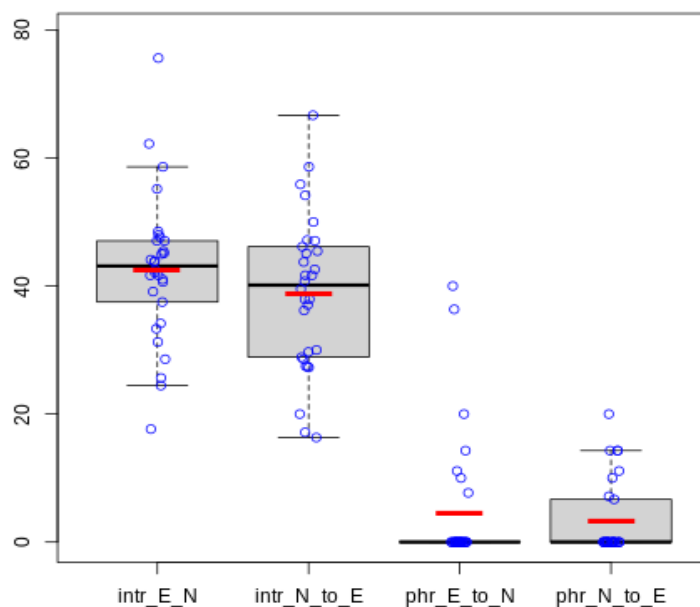


Figure 9: Text-distributed TB of intransitive and phrasal GO and GÅ (fully congruent)

As we have already gleaned from the overall correspondences in Figures 6 and 7, intransitive GÅ and GO are translated congruently much more often than phrasal ones. One reason for this may be that GO and GÅ as intransitive simple lexemes often represent a prototypical use (Viberg 1999: 93), which may be more likely to be translated congruently (cf. Stamenković et al. 2017). Figure 9 offers a more accurate and detailed picture of these main trends. Both the mean and the median values are higher in the E-to-N than in the N-to-E direction. The interquartile range, i.e., the box itself, representing 50% of the dataset, is narrower in E-to-N, indicating more uniform behaviour across texts and translators. Phrasal lexemes with GÅ and GO are hardly ever given a fully congruent translation and their median (TB) is 0% in both directions of translation, although the mean value is slightly higher. Again, the interquartile range is wider in N-to-E, but on the other hand, E-to-N has more outliers. Example (19) shows a congruent correspondence of the phrasal *go under*, which has exactly the

same composition in Norwegian. The more typical case of a noncongruent correspondence is found in (20), where *gå an* ('be possible') does not have an English counterpart with GO.¹³

- (19) If we lose the Sharp account, we're going to *go under* without a ripple. (SK1)
 Mister vi Sharp-jobben, kommer vi til å *gå under* uten å etterlate så mye som en krusning på overflaten. (SK1T)
- (20) *Går det an* å få se ham? (LSC2)
 'Goes it [particle] to get to see him?'
 Would it *be possible* to see him? (LSC2T)

So far only the translation bias has been shown according to dispersion across individual texts. An important aim of this paper was to apply a similar technique for calculating mutual correspondence without losing sight of textual variation. Since MC relies on bidirectional data comparison, the critical question is which texts should be paired with each other. After some discussion and experimenting (including random ordering of texts and pairing according to the frequency of the construction) we decided to use translation bias as the criterion for pairing texts. That is, the highest-scoring text in E-to-N is paired with the highest-scoring text in N-to-E, the two second-highest scoring texts are paired with each other, etc. This, in our opinion, gives the best representation of the range of variation. Since the intransitive uses of GO and GÅ are most frequent, as well as most frequently congruent, we selected these as a testbed for the model. The result is shown in Figure 10.

The plot in Figure 10 shows that the MC scores of text pairs vary between 17% and 60.4% (with an outlier at 71.1%). The mean value in the plot, 40.7%, is practically identical to the MC for intransitive GO and GÅ when calculated in the traditional way, namely 40.4%. At the same time the box plot accounts for textual variation, which makes it more informative. The fact that the texts are paired according to their ranked TB scores enables the plot to display the highest and lowest MCs possible in the corpus. Thus, the full range of variation is shown. A more detailed

¹³ This is similar to the modal use of Swedish GÅ mentioned by Viberg (1999: 108); see 3.2.

qualitative study that will look into possible reasons for abnormally high or low degrees of congruence is underway.

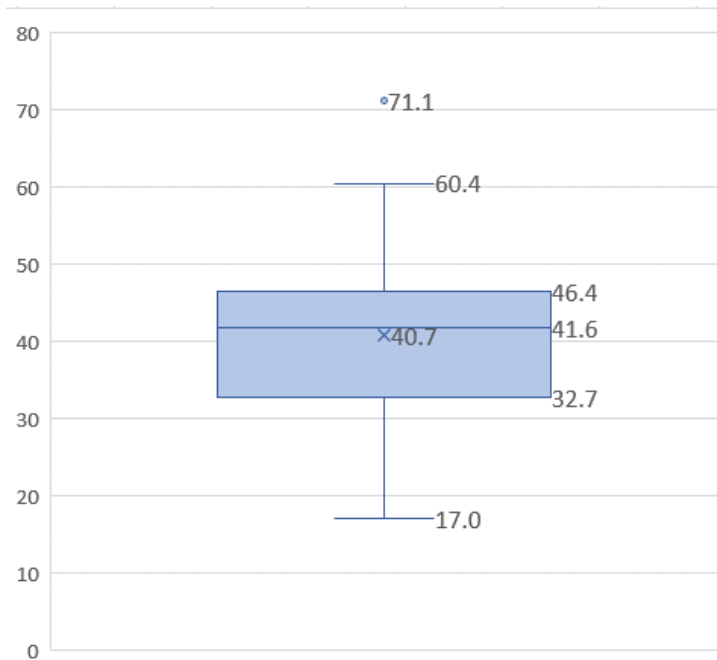


Figure 10: Text-distributed MC of intransitive GO and GÅ (fully congruent), with texts paired according to their TB

4. Discussion

The fact that the traditionally calculated MC value and the mean of the text-distributed MC are virtually the same suggests that both ways of calculating MC give a fair representation of the overall degree of correspondence between items in two languages. There is, however, no doubt as to the added value of representing TB and MC in the latter way. The case study presented in section 3 goes some way towards strengthening the validity of these measures, and by extension, of cross-linguistic findings emerging from a bidirectional corpus investigation. The method of measuring TB and MC on the basis of correspondences in individual texts, rather than on the basis of the corpus as a whole,

highlights variation in congruence across the corpus. Thus, the homogenous picture resulting from a traditional MC score is challenged, in the sense that a text-distributed MC shows that the degree of congruent correspondences is not evenly (or homogeneously) distributed across the data. The width of the interquartile range in a box plot, representing 50% of the data set (e.g., from 32.7% to 46.4% in Figure 10), is an important indicator of the relationship between the items: the narrower the range, the stronger the consensus among translators as to the degree of intertranslatability between the items compared. Conversely, a wider range indicates that there is more room for translators to select either congruent or noncongruent correspondences. Thus, with reference to Figure 9, the narrower box for the English-to-Norwegian translation direction suggests that the translators' judgement of the equivalence between intransitive GO and GÅ is more consistent than in the other translation direction (from GÅ to GO).

The case study also shows the importance of lexicogrammatical considerations in the presentation and understanding of congruence and MC. It is clearly the case that even cognates, such as GO and GÅ, have developed somewhat diverging lexicogrammatical features (cf. Viberg 1999; Storjohann 2003). This, we believe, has implications for how relevant it is to talk about an overarching MC for these verbs. The way forward rather seems to be to apply a more comprehensive framework of correspondence that can capture congruence, or lack thereof, at several linguistic levels. The framework proposed in this study takes the lexicogrammatical features of verbs into consideration. As a result, our definition of full congruence is relatively strict: it applies only to instances of GO and GÅ in the same lexicogrammatical pattern. This classification highlights the fact that some uses of the cognates have a tighter cross-linguistic relationship than others. In order to reveal uses that may be even more strongly connected across the two languages, further refinement of congruence and the degree of correspondence will be considered in the future. This will include a framework of congruence which is more fine-grained, in the sense that it will take a fuller set of the verbs' lexicogrammatical features into account, as indicated in section 3.4.1.

As pointed out in section 3.5.2, there was some discussion as to how the text-distributed TB scores for each direction of translation could best be combined to represent these measures in a joint MC. To randomly combine texts and their TBs in the data set would seem to offer an

inaccurate picture if the aim were to display the range of intertranslatability, from one extreme to the other, with 50% of the correspondences located in between (i.e., the interquartile range). The solution was to pair texts according to their individual TBs from lowest to highest in each translation direction, resulting in MC values ranging from the lowest to the highest possible in the data set. It should be noted that the proposed text-distributed model appears to work best on relatively frequent and well-dispersed linguistic items, as indicated by our attempts to represent the TB and MC of less common uses of GO and GÅ. Further studies are needed to address this issue, but it was suggested above that simple translation paradigms, possibly annotated for text distribution, may be more useful than a more abstract representation of infrequent items.

5. Concluding remarks

This study has shown that the traditional measure of mutual correspondence, following Altenberg (1999), serves as a good starting point for measuring cross-linguistic correspondence/equivalence. However, the main conclusion to be drawn from this investigation is that the expansion of the measures of translation bias and mutual correspondence into what we might call *Distributive Translation Bias* (DTB) and *Distributive Mutual Correspondence* (DMC) offers a more nuanced picture of translation correspondence.

Guided by two research questions, the case study of the cognate verbs GO and GÅ quite clearly demonstrated that 1) measures of translation correspondence are more informative when taking textual variation into account and that 2) measures of translation correspondence are more useful if based on lexicogrammatical features. With reference to 1), given the fact that texts vary substantially in the frequency with which the verbs are translated congruently, the traditional way of calculating TB and MC gives a (too) simplified picture of the correspondence between GO and GÅ. A more nuanced and accurate measure of TB and MC can be achieved by taking textual variation into account, showing the range and distribution of (congruent) correspondences within a given data set. Thus, when plotting the correspondences according to corpus texts, it was shown that the smaller the (interquartile) range, the more reliable the relationship reported by the TB and MC.

With regard to the second research question, the study showed that different verb categories yield varying degrees of correspondence, as

reflected in the generally higher DTB and DMC measures for the intransitive use compared to the phrasal verb use, for example.

The study also suggests that a stricter definition of congruent correspondence (requiring correspondence at a lexicogrammatical level, and not just at a formal, word-class level) better captures details of (non-) correspondence in translation. As for the advantages mentioned above of taking textual variation into account, this is equally important when lexicogrammatical features are part of the analysis.

On a more qualitative note, and beyond the strength of the cross-linguistic relationship between the verbs expressed through descriptive statistical measures, it should be noted that the study of GO and GÅ to a large extent corroborates previous findings regarding these verbs and their uses. In Viberg's (1999) terms, the two verbs are most similar in the prototypical use of 'locomotion in concrete space by a human agent', and to some degree also in 'extended meanings', such as *go to school/gå på skolen*. However, there is generally little overlap between GO and GÅ as part of phrasal verbs (which typically appear in Viberg's categories of mental meanings and grammaticalized uses). These observations will serve as the starting point for a future qualitative study of these verbs and their correspondence patterns, where more attention will be given to the full translation paradigms, as well as to (more specific) lexicogrammatical features that may shed more light on the degree of correspondence between them.

Finally, this study has demonstrated that bidirectional data are essential in determining the degree of cross-linguistic correspondence between items. It has also highlighted some of the other benefits of using parallel corpora in contrastive analysis. Importantly, and as pointed out by Johansson (2007: 5),

[t]he use of bilingual and multilingual corpora, with a variety of texts and a range of translators represented, increases the validity and reliability of the comparison. It can be regarded as the systematic exploitation of the bilingual intuition of translators, as it is reflected in the pairing of source and target language expressions in the corpus texts.

Having a range of translators represented makes it possible to track and measure each individual translator's output. Contrary to previous practice, we have combined the measures for individual translations and expressed them as a Distributive Translation Bias or Distributive Mutual

Correspondence. We believe that this further increases the validity and reliability of cross-linguistic comparison based on bidirectional data.

Corpus

English-Norwegian Parallel Corpus (1994–1997), Dept. of British and American Studies, University of Oslo. Compiled by Stig Johansson (project leader), Knut Hofland (project leader), Jarle Ebeling (research assistant), Signe Oksefjell (research assistant). <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/>

References

- Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In *Out of corpora. Studies in honour of Stig Johansson*, edited by Hilde Hasselgård, and Signe Oksefjell, 249–268. Amsterdam: Rodopi.
- Altenberg, Bengt. 2002. Concessive connectors in English and Swedish. In *Information structure in a cross-linguistic perspective*, edited by Hilde Hasselgård, Stig Johansson, Bergljot Behrens, and Cathrine Fabricius-Hansen, 21–43. Amsterdam: Rodopi.
- Altenberg, Bengt. 2007. The correspondence of resultive connectors in English and Swedish. *Nordic Journal of English Studies* 6(1):1–26. doi: <http://doi.org/10.35360/njes.2>.
- Altenberg, Bengt, and Karin Aijmer. 2000. The English-Swedish Parallel Corpus: A resource for contrastive work. In *Corpus linguistics and linguistic theory*, edited by Christan Mair, and Marianne Hundt, 15–34. Amsterdam: Rodopi.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brinton, Laurel. 1985. Verb particles in English: Aspect or aktionsart? *Studia Linguistica* 39(2): 157–168.
- Cej, Alla. 2008. The polysemous cognates English *go*, German *gehen* and Norwegian *gå*: A corpus-based contrastive study. MA thesis, University of Oslo. <http://urn.nb.no/URN:NBN:no-21406>.
- Ebeling, Jarle. 1998. The Translation Corpus Explorer: A browser for parallel texts. In *Corpora and cross-linguistic research: Theory*,

- method and case studies*, edited by Stig Johansson, and Signe Oksefjell, 101–112. Amsterdam: Rodopi.
- Ebeling, Jarle, and Signe Oksefjell Ebeling. 2013. *Patterns in contrast*. Amsterdam: John Benjamins.
- Ebeling, Jarle, and Signe Oksefjell Ebeling. 2015. An English-Norwegian contrastive analysis of downtoners, more or less. *Nordic Journal of English Studies* 14(1): 62–89. doi: <http://doi.org/10.35360/njes.340>.
- Ebeling, Signe Oksefjell. 2017. Bringing home the bacon! A contrastive study of the cognates *bring/bringe* in English and Norwegian. *Kalbotyra* 70: 104–126. doi: 10.15388/Klbt.2017.11193.
- Fillmore, Charles. 1975. Coming and going. In *Santa Cruz lectures on deixis* 1971, by Charles J. Fillmore, 50–69. Reproduced by the Indiana University Linguistics Club.
- Hasselgård, Hilde. 2015. Coming and going to the future: Future-referring expressions in English and Norwegian. In *Cross-linguistic perspectives on verb constructions*, edited by Signe Oksefjell Ebeling, and Hilde Hasselgård, 88–115. Newcastle: Cambridge Scholars Publishing.
- Hasselgård, Hilde. 2024. Periphrastic genitive constructions in English and Norwegian. In *Contrastive corpus linguistics. Patterns in lexicogrammar and discourse*, edited by Anna Čermáková, Hilde Hasselgård, Markéta Malá, and Denisa Šebestová, 62–86. London: Bloomsbury.
- Hopper, Paul, and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Huddleston, Rodney, and Geoffrey Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Johansen, Stine C. 2011. Går på tur aldri sur: En korpusbasert studie av bevegelseshandlinger og morsmålstransfer i innlærertekster fra voksne informanter i norsk som andrespråk. MA thesis, University of Oslo. <https://www.duo.uio.no/handle/10852/26696>.
- Johansson, Stig. 2007. *Seeing through multilingual corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.
- Levenston, E. A. 1965. The 'translation-paradigm'. A technique for contrastive syntax. *International Review of Applied Linguistics in Language Teaching* 3(3): 221–226. doi: <https://doi.org/10.1515/iral.1965.3.3.221>.

- NAOB. *Det Norske Akademis Ordbok*. Det Norske Akademi for Språk og Litteratur. <https://naob.no/>.
- OED (*Oxford English Dictionary*) Online. 2023. 'go, v.'. Oxford University Press.
- Ordnett.no. 2023. *Stor engelsk-norsk / norsk-engelsk ordbok* (2021). Oslo: Gyldendal Norsk Forlag.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Stamenković, Dušan, Miloš Tasić, and Vladan Pavlović. 2017. Prototype theory and translation equivalent selection: The case of motion verbs. *Studia Neophilologica*, 89(1): 81–94. doi: <https://doi.org/10.1080/00393274.2015.1132179>.
- Storjohann, Petra. 2003. *A diachronic contrastive lexical field analysis of verbs of human locomotion in German and English*. (European University Studies.) Frankfurt: Peter Lang.
- Thormodsæter, Øyvind. 2021. The idiomaticity of emotion in English and Norwegian: A corpus-based contrastive investigation of the phraseology of the three English-Norwegian verb pairs *enjoy-nyte*, *love-elske* and *like-like*. PhD dissertation, University of Oslo.
- Viberg, Åke. 1996. Cross-linguistic lexicology. The case of English *go* and Swedish *gå*. In *Languages in contrast. Papers from a symposium on text-based cross-linguistic studies. Lund 4-5 March 1994*, edited by Karin Aijmer, Bengt Altenberg, and Mats Johansson, 153–184. Lund: Lund University Press.
- Viberg, Åke. 1999. The polysemous cognates Swedish *gå* and English *go*. Universal and language-specific characteristics. *Languages in Contrast* 2(1): 89–115.