‘Be’ Verbs in a Contrastive Perspective: The Case of BÝT, BE and VÆRE

Anna Čermáková (Lancaster University)

Jarle Ebeling (University of Oslo)

Signe Oksefjell Ebeling (University of Oslo)

Abstract

This paper reports on a cross-linguistic study of ‘be’ verbs in Czech, English and Norwegian, viz. BÝT, BE and VÆRE, drawing on data from the fiction part of the International Comparable Corpus (Čermáková et al. 2021). The study identifies two main uses of ‘be’ verbs: auxiliary and linking, plus an ‘other’ category which includes minor (often) language-specific uses. The study reveals marked proportional differences in how the three languages exploit the grammatical and functional potential of their respective ‘be’ verbs: notably, there is a marked preference for linking uses in English and Norwegian and a more even distribution between auxiliary and linking uses in Czech. In a case study of the linking use, the languages are shown to behave similarly, but with some minor differences regarding choice of adjective to describe fictional subjects. The methodology highlights the importance of a carefully crafted *tertium comparationis* at several levels, not only in relation to datasets and linguistic phenomena investigated, but also as regards terminology and grammatical traditions of description for the languages compared.

Keywords: ‘be’ verbs; corpus-based contrastive analysis; International Comparable Corpus; Czech-English-Norwegian; adjective classification; linking verbs

1. Introduction and aims

This paper reports on a cross-linguistic investigation of ‘be’ verbs in Czech, English and Norwegian, i.e., BÝT, BE and VÆRE. BE and its closest counterparts in other languages have been researched extensively (e.g., Verhaar 1967–1972; Bybee and Dahl 1989); however, to our knowledge,

Čermáková, Anna, Jarle Ebeling, and Signe Oksefjell Ebeling. 2023. ‘‘Be’ Verbs in a Contrastive Perspective: The Case of BÝT, BE and VÆRE.’ *Nordic Journal of English Studies* 22(2): 27–58.

a detailed corpus-based contrastive mapping of the uses of the verbs in these three languages has not been performed before. The study draws on data from the International Comparable Corpus (ICC) (https://korpus.cz/icc; see Čermáková et al. 2021; Kirk et al. 2018), a corpus that facilitates contrastive studies across many languages and many registers. Although some work still remains to complete the different sub-corpora of the ICC, some sections are completed and available for research. The material for the current study is culled from one register only (creative writing, i.e., fiction) across the three languages: Czech, English, and Norwegian. These register-specific sub-corpora are relatively small, amounting to around 40,000 words per language; however, the word frequency lists reveal that the prototypical verbs of 'being' are the most frequent ones in all three languages: BÝT with 1,845 occurrences, BE with 1,720, and VÆRE with 1,292.[1]

Etymologically the three verbs are related in a somewhat complex web of partly overlapping origins (Rejzek 2015; OED; Bjorvand and Lindeman 2019). It is therefore not surprising that, synchronically, there are both overlapping and non-overlapping uses, functions, and meanings. For example, all three verbs are used to express existence and they all function as auxiliaries to form the passive voice, but only English BE can be used as an auxiliary marking the progressive aspect,[2] only Czech BÝT functions as an auxiliary to mark the past and future tense and the conditional mood, while Norwegian VÆRE may function as an auxiliary to form the perfect aspect. All three 'be' verbs are, however, extensively used as linking verbs (copulas), as in examples (1) a–c with an adjectival complement.[3]

(1) a. Facilities *were* rather spartan (ICC-EN)

b. Vlak *je* zrezivělý, a tohle nástupiště *je* teď pusté. (ICC-CZ) ['The train *is* rusty, and this platform *is* now deserted.']

[1] The second most frequent verb in all three languages is 'have' with much lower frequencies: MÍT (334), HAVE (637), HA (704).

[2] Norwegian has an *-ende* form of the verb corresponding to the English *-ing* form, but it rarely combines with auxiliary VÆRE to form the progressive and we do not expect it to be attested in the material.

[3] All the examples come from the ICC, ICC-EN refers to the English sub-corpus, ICC-CZ and ICC-NO to the Czech and Norwegian ones, respectively.

c. Hun visste at hunden *var* adskillig reddere. (ICC-NO)
['She knew that the dog *was* considerably more scared']

This brief introduction to some of the established similarities and differences between the three verbs serves as an incentive to investigate how the various uses are distributed within each language. Following an initial overview of the various uses of these verbs in the ICC material, the study moves on to an in-depth analysis of the three verbs when they have a linking function, which, in fact, is the most frequent use in all three languages (see Table 1 in section 4).

Against this backdrop, and by pinpointing the preferred uses of the three verbs, we wish to address the following research questions:

1. To what extent do these verbs overlap in meaning and use?
2. In the linking use with an adjectival complement, what kind of relationship does each of the verbs typically establish between the elements that are linked?
3. In relation to these verbs, what are the methodological challenges of a detailed cross-linguistic corpus-based comparable study of (typologically) different languages?

The first two questions are cross-linguistically (and cross-culturally) interesting in that they may shed light on similarities and differences in the linguistic make-up of fictional texts and fictional descriptions of people, objects, and events across the three languages. Answers to these two questions will feed into the third question, which may be seen as an overarching question with the aim of testing the potential of a comparable corpus such as the ICC.

Previous contrastive studies of 'be' verbs in the relevant languages typically focus on some specific patterns and uses, for example, clefts in English and Norwegian (e.g., Gundel 2002), existential constructions in English and Norwegian (e.g., Ebeling 2000), information structure in English and Norwegian and English and Czech (e.g., Gundel 2002; Dušková 2004, 2005).[4] The current study thus contributes to this type of research focus on one particular 'be' pattern (see research question 2). At

[4] See also Malá (2014) who investigates copular verbs other than BE in a contrastive perspective.

the same time, it adds another dimension in also being concerned with the larger picture of outlining and comparing the distribution of the various uses of 'be' verbs in these three languages (see research question 1). A detailed cross-linguistic analysis of (typologically) different languages also raises numerous concrete and more general methodological and theoretical issues, which we also wish to address (see research question 3).

The paper has the following structure. Section 2 introduces our contrastive framework and provides some background to the three 'be' verbs. Section 3 introduces the International Comparable Corpus, our methodology and our classification scheme. Section 4 provides a quantitative overview of the main functions of the 'be' verbs, and the different linking patterns in which they occur. Section 5 presents a more qualitative case study and analyses the pattern 'NP+ BE|VÆRE|BÝT+ADJP'. Section 6 offers conclusions and suggestions for further studies.

2. Contrastive framework of 'be' verbs

As mentioned in section 1, the three 'be' verbs are linked to each other cross-linguistically in various ways: etymologically, syntactically, semantically, and functionally. Before we move on to the cross-linguistic comparison proper, we will briefly discuss each verb in turn for two reasons in particular: 1) different traditions of grammatical description (and terminology) in the three languages, and 2) each verb seems to have uses that do not apply, or only marginally apply, to the other verbs.

Different traditions of grammatical description, and terminology in particular, may, in contrastive studies of languages as different as Czech vs. English and Norwegian, be misleading. It is therefore important to identify terminology that covers common ground to make sure we compare like with like. Or, as Stassen (1997: 4, 9) puts it, we need a 'cross-linguistic definition of the domain of inquiry', 'so that we will not end up working with a database which contains incomparable items'. Indeed, grammars and other relevant publications on the three 'be' verbs suggest that there is no consensus either across, or within, the languages regarding the terminology used to describe these verbs. For example, the term 'copular verb' may evoke different conceptualisations both within and across the three languages. In English, 'copular' may refer to both the SVsP ('subject+verb+subject predicative' as in *The country is independent*.) and SVA ('subject+verb+adverbial' as in *I have been in the*

garden.) patterns, or only the SVsP pattern (see section 2.1). The verb in the SVA pattern would then rather be labelled intransitive. For the purpose of this study, we have opted for the term 'linking' to refer to the SVsP and SVA patterns, reserving the term 'intransitive' for the SV pattern only.[5] In other words, we follow the terminology used by, for example, Quirk et al. (1985). In the following sub-sections, we will outline some language-specific background for each of the three 'be' verbs in turn.

2.1 BE: Some background

According to two of the major grammars of modern English—*A Comprehensive Grammar of the English Language* (Quirk et al. 1985) and *The Longman Grammar of Spoken and Written English* (Biber et al. 1999)—English BE primarily has auxiliary and copular uses. As an auxiliary verb, BE can be used to form the passive voice, as in (2), or the progressive aspect, as in (3). Moreover, according to the *Oxford English Dictionary* (OED), it is also possible for BE to be used as an auxiliary to form the perfect aspect, although it is 'now largely replaced by *have* following the pattern of transitive verbs' (OED). In addition, there are some multi-word uses with BE that may be aspectual (e.g., BE *about to*), quasi-modal[6] (e.g., BE *to*) or future-referring (e.g., BE *going to*) in nature.

(2) The autumn evenings *were* marked by the Listowel races (ICC-EN)

(3) Somewhere a baby *was* crying. (ICC-EN)

As far as the copular use of BE is concerned, the following patterns are recognized: 'S+BE' followed by either a subject predicative (sP) or an adverbial (A) (Quirk et al. 1985: 1171; Biber et al. 1999: 435). The subject predicative associates some attribute with the subject of the clause 'stating of what sort or what something is' (OED), as illustrated in (4), while the adverbial in an 'S+BE+A' pattern indicates 'the relationship of the subject in place, state, time, etc., to another thing or person' (OED), as the locative *downstairs* in (5). As pointed out by Biber et al. (1999: 141), some

[5] Yet other frameworks subsume all three under the intransitive label, notably so Stassen (1997) in his typological work on intransitive predication.
[6] Huddleston and Pullum (2002:113–114).

grammars would call the 'S+BE+A' pattern intransitive (e.g., Huddleston and Pullum 2002), along with the 'S+BE' pattern without a required complement when it has the meaning of 'to have or take place in the world of fact, to exist, occur happen' (OED), as in (6).

(4) My mother *was* furious. (ICC-EN)

(5) Every morning he *was* downstairs first (ICC-EN)

(6) When all things began, the Word already was. (*Bible (New Eng.)* John i.; OED)

Finally, there are a number of special constructions with dummy or empty subjects (*it* or *there*) that bear resemblance to copular 'S+BE+sP' uses, including anticipatory *it* as in (7), clefts as in (8), empty *it* as in (9) and existential '*there*+BE' constructions as in (10) (see also Quirk et al. 1985 for a discussion of these).

(7) *It was* sad to see so many of one's school pals gathering at the corner (ICC-EN)

(8) *It was* the woman who showed understanding (ICC-EN)

(9) *It was* after midnight. (ICC-EN)

(10) *There's* sherry and port wine for each of the ladies (ICC-EN)

In their discussion of copular verbs in English, Biber et al. (1999: 448) note that '[t]he typical subjects and complements occurring with *be* differ in important ways across registers.' For example, conversation and academic prose are found to differ proportionally in their use of adjectives, nouns, or prepositional phrases as complements (Biber et al. 1999: 449). As will become evident, this is an interesting observation in the current context since, like registers, languages may also be seen to differ in this regard, even within the 'same' register of fiction.

2.2 VÆRE: Some background

VÆRE is very similar to English BE in most of its uses and is described as a copula verb in the Norwegian reference grammar, *Norsk referansegrammatikk* (Faarlund et al. 1997: 733ff.), when it links together two constituents. The most frequent use of this kind is when the post-verbal constituent, the predicative, is an adjective (Faarlund et al. 1997: 734). The following examples, (11)–(14), show cases where the post-verbal element is an adjective phrase, a noun phrase, an adverb (phrase) and a clause, respectively.

(11) Hun *er* fortsatt for dårlig, men har visst våknet. (ICC-NO)
['She is still too ill, but has gained consciousness.']

(12) De skal vite at fadese *er* en mulighet. (ICC-NO)
['They should know that a faux pas is a possibility.']

(13) For en uke siden lå han på sofaen, nå *er* han her. (ICC-NO)
['A week ago he was on the sofa, now he is here.']

(14) Regelen vår hadde *vært* at jeg ringte. (ICC-NO)
['Our rule had been that I called.']

VÆRE is used as an auxiliary to form the passive voice and the perfect aspect, as in (15) and (16). In contrast to English, the passive voice with VÆRE is comparatively infrequent,[7] whereas the perfect aspect with VÆRE is still productive in Norwegian.

(15) brua *var* festet med store bolter. (ICC-NO)
['the bridge *was* secured with large bolts']

(16) Det *er* blitt sent på natten. (ICC-NO)
['It *is* become late at night.']

[7] In Norwegian, the passive voice is more commonly formed with the auxiliary BLI 'become' (*bli brukt* 'become used'). There is also an inflectionally marked passive called the *-s* passive (*brukes* 'use_PASS'). See Faarlund et al. (1997: 512ff., 523ff.).

Similar to English BE, Norwegian VÆRE is used together with the formal (anticipatory) subject *det* ('it') as in (17), cleft constructions as in (18) and *det* ('there') in presentative (existential) constructions, as in (19).

(17) *det* hadde ikke *vært* nødvendig å gå til legen. (ICC-NO)
['*it* had not *been* necessary to go to the doctor.']

(18) *Er det* dette du virkelig mener, *er det* dette du faktisk vil? (ICC-NO)
['*Is it* this you really mean, *is it* this you actually want?']

(19) Nei, *det* skal vel egentlig ikke *være* folk her nå. (ICC-NO)
['No, *there* should not really *be* people here now.']

In example (20), the formal subject is empty in the sense that it does not refer to a preceding or following unit in the co-text or a notional subject as in the case of extraposition and existential constructions (cf. the English examples (7) and (10) in section 2.1).

(20) *Det* har *vært* mørkt lenge. (ICC-NO)
['*It* has *been* dark a long time.']

VÆRE forms part of literally hundreds of more or less fixed phrases (see VÆRE in NAOB, 'The Norwegian Academy Dictionary'). Two special uses could be mentioned: when VÆRE incorporates a modal meaning as in (21) and when it combines with another (main) verb to instantiate distance as in (22). We do not expect to find many instances of such uses in our material.

(21) Han *er* ikke lenger å se. (Faarlund et al. 1997: 530)
['He can no longer be seen' (lit. He is not longer to see)]

(22) Han *var* og reparerte bilen i byen (Faarlund et al. 1997: 536)
['He was in town to get his car fixed' (lit. He was and repaired the car in town)]

2.3 BÝT: Some background

Czech grammars and dictionaries (Daneš et al. 1987, Cvrček et al. 2015, Havránek et al. 1989) describe three functions of the verb BÝT as: i) 'autosemantic', which refers to its existential meaning, ii) copular, and iii) auxiliary. In its auxiliary function, BÝT participates in the formation of the past tense, as in (23), the passive voice, as in (24), the conditional mood and the future tenses, as in examples (25) and (26), respectively.

(23) Rozsvítil *jsem* modrou lampičku (ICC-CZ)
['I turned on the blue lamp' (lit. Turned on *am* the blue lamp_DAT)]

(24) Je původní, potvrzuje dnešní majitel, který sem *byl* také před třiceti roky přenesen z Evropy. (ICC-CZ)
['It is original, confirms the contemporary owner, which *was* also thirty years ago brought over here from Europe']

(25) táta *by* asi nepřijel (ICC-CZ)
['Dad *would* probably not come']

(26) Pak jste pozoroval, co se *bude* dít (ICC-CZ)
['And then you observed, what *is going to* happen']

In its copular use, BÝT can participate in several patterns—with or without an overtly expressed subject. The patterns with unexpressed subject, where the subject is a 'general actant' (Cvrček et al. 2015: 361), include 'BÝT+PRED.' and 'BÝT+PRED.+OBJ.' (as in *Je mi smutno.* 'I am sad' [lit. Is me_DAT. sadness]). In 'BÝT+PRED.', the predicative can be either a noun or an adverb (as in *Je podzim.* 'It's autumn.' [lit. Is autumn]) or a prepositional phrase (as in *Bylo pod mrakem.* 'It was cloudy' [lit. Was under a cloud_INSTR.]). All other copular patterns have a subject; although this may be elided, it is inferred from the context and the verb form indicates person and number (and gender in third person past tense), i.e., pro drop, as in (27), where the subject 'you' is implicitly part of the verb form *jste* ('are', the polite formal sg. form). In addition, there are instances such as (28), where the subject, *černovláska* ('black-haired girl'), is recoverable from the context. The two main copular patterns with expressed subject are 'S+BÝT+PRED.' and 'S+ BÝT+PRED.+OBJ.', where

predicatives are most frequently adjectives and nouns (for a detailed discussion of the numerous patterns, see Daneš et al. 1987: 3.1). Traditionally, copula constructions in Czech do not include the SVA pattern.

(27) Zřejmě *jste* hodně sečtělý. (ICC-CZ)
['Evidently (you) *are* well-read']

(28) hezká černovláska, trochu vlnitý vlasy a měla na sobě bordó šaty na ramínka, *byla* hrozně sexy (ICC-CZ)
['pretty black-haired girl, a bit wavy hair and had on (her_REFLEX.) burgundy strappy dress, (she) *was*_FEM. really sexy']

BÝT is also used in a number of phraseological constructions that are more or less syntactically irregular (see Havránek et al. 1989). Its existential meaning ('autosemantic use') is conceptualised differently to that of English (and Norwegian) and there is no formal correspondence to the English existential construction with *there* (Dušková 2006: 12.21.4). Similarly, the frequent English (and Norwegian) dummy and empty *it* constructions have no direct counterparts in Czech. The English dummy/empty *it* is sometimes comparable to the pronoun *to*. However, unlike *it* in English, *to* preserves strong deictic and co-referential functions in many of its uses (Klimešová et al. 2015; Adamec 1998); Čermák (2010) notes its universal deictic function, while others also stress its pragmatic function (Havránek and Jedlička 1960: 381). Klimešová et al. (2015) highlight the frequent use of *to* in spoken as opposed to written language, *to* being the most frequent word form in informal spoken Czech. Their study confirms that *to* preserves strong referential semantics in about 70% of the cases in conversation. It should be noted in this context that, unlike English, Czech does not use determiners and therefore does not grammatically distinguish between known (definite article *the*) versus unknown (indefinite article *a/an*), which may be one of the reasons why *to* is strongly felt to be referential, especially in conversation.

2.4 Synthesising the descriptive frameworks

On the basis of grammars and previous research, the 'be' verbs in the three languages can be said to primarily have linking and auxiliary uses, in

addition to some other minor uses. This, as we shall see, is also reflected in the data from the ICC. What the ICC material offers, in addition to confirming these uses, is an overview of the proportions with which the respective 'be' verbs are used in the different functions, thus potentially pointing to similar or different preferred areas of use for BÝT, BE and VÆRE.

The discussion above illustrates that, in a contrastive study of items in languages as different as Czech vs. English and Norwegian, it is important to define a common linguistic ground and a common descriptive framework to ensure objective comparability between the items in the three languages. We have therefore opted for a common taxonomy and terminology that is wide, but precise enough to cater for all three languages. The principal verb categories we operate with are as follows:

1. Auxiliary verb
2. Linking verb
   - Linking subject and complement (NP, ADJP, ADVP, PP, NUM, clause)
   - Special linking expressions (existential, extraposition, cleft, empty S, *to* subject, general subject)
3. Idiomatic (phrasal) uses
4. Intransitive

The main contrastive analysis in section 5 focuses on the Linking category, and more specifically BE, VÆRE, BÝT with an adjective (phrase) as their complement. As we have seen, this is one of the patterns that many grammars label 'copular' (see sections 2.1 to 2.3). As the term 'copular verb' seems to evoke different conceptualisations both within and across the three languages, we will refer to this type of use as 'Linking'. A further restriction on the pattern selected for the more detailed analysis is that the subject should be (recoverable) in the form of a noun phrase. This first part of the pattern, i.e., the NP, will be further specified in terms of reference (human/non-human), whereas the second part (ADJP) will undergo a semantic classification of the head adjective, according to the framework outlined in section 3.2.

Such a detailed delimitation and classification is deemed necessary in order to be able to fully capture potential differences between the languages. The subject NP is often elided in Czech (see section 2.3), but

indirectly present through the verb and thus contextually identifiable. This level of description is needed to enable a direct comparison with English and Norwegian, both of which typically operate with expressed subjects.[8] This may also shed light on how the subject correlates with the 'complement' category. In other words, the semantic analysis of the adjectival head of the complement may establish similarities and differences in how subject NPs are characterized by means of adjectives across the languages and ultimately contribute cross-linguistic insights into the linguistic make-up of descriptions in fiction.

3. Data and methodology

3.1 The International Comparable Corpus

We use data from the International Comparable Corpus (ICC), which is still under construction, and, in fact, this is the first linguistic study based on parts of this corpus. There are currently twelve different national teams involved in the project representing twelve different languages. The teams are at different stages in the compilation process, but the aim is to create a 1-million-word corpus for each language, structured according to the design criteria of the International Corpus of English (ICE) in terms of text types, sampling and size (Greenbaum 1991). Regarding the corpus compilation itself, the idea is to reuse material from already existing corpora wherever possible (for more detail, see the project website https://korpus.cz/icc and Čermáková et al. 2021).

For the purpose of this study, we have chosen to focus on a small, but finished portion of the ICC for three languages, namely the fiction sections (called 'creative writing' in the ICE scheme) in Czech, English and Norwegian. The ICC-CZ fiction component contains twenty 2,000-word extracts from fiction published between 2000 and 2014 with an equal distribution between male and female writers.[9] Similarly, the ICC-NO material consists of extracts of books published between 2000 and 2011,

[8] Although not pro-drop languages, English and Norwegian may also elide the Subject, as long as it is recoverable from the context, e.g., in the infinitive clause *to be lenient* in the following example, where the understood subject is *she*: ***She*** *became softer, kinder, like a mother who has been impatient with her bold child but has suddenly decided* ***to be lenient*** (ICC-EN).

[9] The fiction part in ICC-CZ was compiled using texts from the Czech National Corpus (www.korpus.cz).

represented by ten female and ten male writers.[10] The fiction part of ICC-EN overlaps in its entirety with the creative writing part of ICE-Ireland (Kallen and Kirk 2008). Thus, the twenty text extracts are equally divided between the North and the South (i.e., Northern Ireland and the Republic of Ireland). While the ten extracts from the South are balanced between male and female writers, they are skewed nine to one in favour of male writers in those from the North. Moreover, the fiction texts included in ICC-EN were published between 1990 and 2000.[11] Despite the slightly earlier publication dates of the ICC-EN texts, we believe that comparability of our object of study within this short contemporary time frame is not jeopardised. In terms of size, then, each of the ICC components used in this investigation amounts to around 40,000 words, which by today's standards is very small, and which in turn meant that we opted for a high-frequency linguistic item as our object of study.

As mentioned in the Introduction, when we extracted a word frequency list, it became clear that the prototypical verb of 'being' is the most frequent verb lemma in fiction in all three languages, ranging from ca. 1,300 to 1,800 occurrences in the respective sub-corpora.[12] Based on this observation, we decided to investigate the behaviour of the three most common verb lemmas: BÝT, BE and VÆRE. A cross-linguistic comparison, including frequencies and proportions of different uses in the three languages, has not been performed before and will potentially result in valuable contrastive insights. Moreover, the contrastive analysis will serve as a useful testbed for future studies based on the ICC.

Before we move on to the study itself, a few words on using comparable corpora for contrastive research are in order. In any contrastive study it is important to establish criteria of comparability that ensure a sound and objective *tertium comparationis*, i.e., an objective frame of reference for the comparison. The comparable data in our three ICC components ensure this through text type, the time period the material is taken from and the object of study: as established above, the three 'be'

[10] The fiction part of ICC-NO consists of texts from the Norwegian 100-million-word corpus Leksikografisk bokmålskorpus (Fjeld et al. 2020).

[11] See further Kallen and Kirk (2008) for a detailed description of ICE-Ireland, and by extension the creative writing section of ICC-EN used here.

[12] These figures include all forms of the 'be' verbs; for Czech, it also includes negative forms of BÝT, that have a different root.

verbs are etymologically, syntactically, semantically and functionally related. Thus, there is an underlying perceived similarity that may serve as a starting point for a contrastive analysis (see, e.g., Ebeling and Ebeling 2020).

3.2 Methodology and data classification

As a first step, we classify all the instances of BE, VÆRE and BÝT in order to distinguish between the different uses and functions, i.e., between the linking, auxiliary and other uses (see section 2.4). In the next step, we focus on the linking use, identifying the existing patterns of this use, most of which are attested in all three languages. These patterns are:

- 'NP/Ø+V+NP'
- 'NP/Ø+V+ADJP'[13]
- 'NP/Ø+V+ADVP/PP/NUM/INF/CLAUSE'
- Dummy S constructions (i.e., special linking constructions: see section 2.4)
- Other

The subject position in the first three patterns, NP/Ø, is meant to illustrate the typological differences between the languages, allowing for both an expressed and elided subject as discussed above, where Ø refers to the latter.

As outlined in section 2.4, the special linking constructions include in English and Norwegian empty *it/det* constructions, existential *there/det* constructions, anticipatory *it/det* constructions and cleft constructions: see examples (29) with anticipatory *it* and (30) with existential *det*. In Czech, this category includes all constructions with the pronoun *to*(/*tohle*) as subject and all cases with unexpressed general subject (see section 2.3 and example 31). The 'other' category primarily includes intransitive uses and cases that did not easily fit the above categories, mainly idiomatic uses, e.g., *father to be*.

---

[13] The pattern 'NP/Ø+V+ADJP' subsumes the Czech construction 'NP/Ø+V+NP+ADJP' (see section 2.3), in which there is a NP in the dative case inserted between the adjective complement and the verb as in *Jsem ti dobrá jenom na* ('I am good enough for you only for' [lit. Am you_DAT. good only for]) (ICC-CZ).

(29) It *was* sad to see so many of one's school pals gathering at the corner (ICC-EN)

(30) det *er* for mange detaljer som bærer på en uhørt og urimelig mening (ICC-NO)
['there *are* too many details that carry an outrageous and unreasonable meaning']

(31) *je* tu horko (ICC-CZ) [lit. *is* here hot]

In the next step we focus on the linking pattern 'NP/Ø+V+ADJP'. We use a binary classification of all subjects, human or non-human, whereas their adjectival complements are classified according to the following semantic classes (primarily based on Biber et al. 1999 and Lorenz 1999; see also Dixon 2010: 73ff.), exemplified with English adjectives:

- Certainty: e.g., *clear*, *likely*, *obvious*, *true*
- Affective (affective psychological states and personal affective stance): e.g., *anxious*, *friendly*, *scared*
- Evaluative (evaluation of animate beings, situations, events, etc.): e.g., *awful*, *beautiful*, *surprising*
- Time (age, chronology, frequency): e.g., *early*, *old*, *quick*
- Colour: e.g., *black*, *bright*, *red*
- Physical property: e.g., *big*, *cold*, *wrinkled*
- Miscellaneous: e.g., *professional*, *missing*, *racist*

3.3 Reflection on some methodological issues

There are numerous issues to consider when comparing (typologically different) languages, and we discussed some of those relating to terminology and theoretical framework in section 2. Furthermore, typological differences raise questions regarding the comparability of categories and lexical items. A case in point is the special linking category that features 'dummy S constructions'. While the English and Norwegian constructions largely overlap, a comparable conceptualisation does not exist in Czech (see section 2.3). This category includes cases where the dummy or anticipatory subject is either *it* or *there* in English and *det* in Norwegian. In Czech, the mechanism of extraposition is typically handled by its flexible word order. The 'dummy S constructions' are mostly

impersonal constructions, which in Czech are often expressed with an empty subject ('general actant') or the pronoun/particle *to*. It is generally very difficult to distinguish between the various uses of *to* (see Adamec 1998) and specifically to distinguish between the referential and non-referential meanings: see example (32) for a case of referential use, that could have been included in the 'NP/Ø+V+ADJP' pattern.

(32) to *by bylo* zbytečný. (ICC-CZ) ['it *would be* useless.']

Since we are working with fictional texts, there is a fairly large proportion of constructions beginning with the pronoun *to* (66% of the 'dummy S' category). The use of *to*, in many cases, resembles English *it* or Norwegian *det* in function. These constructions are typical of spoken Czech (Klimešová et al. 2015), and based on our results, seem to be frequent in fiction as well. Therefore we have opted for a solution that relies on the surface syntactic similarity and included the constructions with *to* (together with the unexpressed general subject) in this category. A proper contrastive alignment of this category would require parallel (translation) data and is thus outside the scope of this study. At the same time, it is worth noting that the cases where the subject remains unexpressed in Czech (the general actant, see section 2.3), but also some of the phraseological constructions, may actually correspond to English /Norwegian constructions with dummy *it/det.* There is no scope in this study for a detailed category mapping, if such is indeed possible, and our frequency overview needs to be interpreted in the light of these existing, and not easily matched differences. Admittedly, this allows for a slight mismatch between the classifications, and therefore we will comment briefly on this in section 4.[14]

Another feature that may pose a challenge is the actual content of the corpus texts. As mentioned in section 3.1, the ICC is designed according to the guidelines for the ICE family of corpora in that it is a balanced corpus including equal proportions of a number of text type samples. Even if all sub-corpora have been carefully compiled, it is inevitable that the text extracts within each register or text type will focus on slightly different topics. In the case of the register used in this study, there may be further

[14] We would like to thank the two anonymous reviewers for helpful comments and suggestions regarding this point.

differences related to fictional sub-genre (general fiction, crime fiction, etc.) and the amount of dialogue vs. narrative in each extract and language. For practical compilation purposes, these issues can only be controlled for to a certain extent, and differences resulting from these may have some impact on our results.

4. Quantitative overview

Let us now turn to our first research question: To what extent do the three verbs overlap in meaning and use? Table 1 shows some clear differences in the proportion of linking uses on the one hand and auxiliary uses on the other, especially between Czech and the other two languages, but to some extent also between English and Norwegian. These findings are not that surprising since we know that Czech uses BÝT to form the past tense (67% of the auxiliary uses), which is the tense of choice in much of prose fiction. A substantial proportion of the auxiliary uses in Czech also represents the conditional mood (23%) with its own set of BÝT forms. Moreover, English uses BE to form the progressive aspect, which Czech and Norwegian do not, and Norwegian does not use VÆRE as a passive auxiliary to the same extent as English and Czech use BE and BÝT.

Table 1. Proportional distribution of BÝT, BE and VÆRE in their linking (including 'dummy S' constructions), auxiliary and other uses (raw figures and percentages of total number of each 'be' verb)

Verb	Linking	Auxiliary	Other	Total
BÝT	906 (49%)	864 (47%)	72 (4%)	1,842
BE	1,251 (72%)	454 (26%)	38 (2%)	1,742
VÆRE	1,158 (90%)	84 (6%)	50 (4%)	1,292

There are quite a few interesting observations to make about the auxiliary and other uses of the three verbs, including the fact that there are few overlaps between the languages, apart from their uses as passive auxiliaries (see sections 2.1–2.3). The relatively high proportion of the Czech auxiliary BÝT may also be due to the fact that present-day Czech has only one auxiliary verb available for the expression of grammatical meanings (as opposed to, e.g., English, which has several). The many cross-linguistic differences in, e.g., auxiliary uses, however, will have to be left for another study, as our focus will rather be on one of the linking uses.

Narrowing the scope, Table 2 shows the distribution of linking uses in Czech, English and Norwegian. 'Dummy S' constructions are included in the table, although they may only partly (functionally) overlap in the three languages (for a discussion see section 3.3).

Table 2. Main linking uses with BÝT, BE and VÆRE

Verb	**NP/Ø+V+NP**	**NP/Ø+V+ADJP**	**NP/Ø+V+ADVP/PP/ NUM/INF/CLAUSE**	**Dummy S constructions**
BÝT	153 (16.9%)	239 (26.4%)	197 (21.7%)	317 (35.0%)
BE	362 (29.3%)	436 (35.3%)	219 (17.7%)	217 (17.5%)
VÆRE	294 (26,0%)	335 (29.5%)	180 (15.9%)	326 (28.7%)

Table 2 uncovers some interesting cross-linguistic differences and similarities in terms of preferred uses. Leaving the category of 'Dummy S construction' aside, the 'NP/Ø+V+ADJP' pattern (the shaded column in Table 2), which will be the focus of our case study, is the most frequent linking pattern in all three languages. The relatively high proportion of the pattern 'NP/Ø+V+ADJP' in English (compared to Czech, in particular) may also reflect that what English tends to express in the pattern 'be+ADJP' will often be expressed by a lexical verb in Czech (*be slow/quick at doing something* vs. *dělat něco pomalu/rychle* ['do/make something slowly/quickly'] (see Dušková et al. 2006).[15] Moreover, if we include the referential uses of the *to*-pattern in the 'NP/Ø+V+ADJP' pattern, the number of cases for Czech increases to 273 (30%) (and correspondingly the 'Dummy S construction' decreases to 283 (31.2%).[16] English and Norwegian also show a strong preference for the 'NP+V+NP' pattern, while Czech shows greater variety of distribution across the various linking patterns.

Finally, the numbers for the 'NP/Ø+V+ADJP' pattern in Table 2 need to be adjusted since they do not take account of the fact that the ADJP in some instances contains more than one adjective, e.g., *The bus was <u>empty</u>*

[15] We would like to thank one of the reviewers for pointing this out.

[16] We re-classifed all *to*-patterns into referential and non-referential uses with the above caveat in mind that the distinction is not unambiguous: 61.5% (128) of the occurrences were classified as referential (cf. 70% in conversation in the study by Klimešová et al. 2015), and, out of these, 34 had ADJPs as complements. For the remainder of this investigation we will, however, operate with the original number of instances for the 'NP/Ø + BÝT +ADJ' pattern (N=254), thus leaving the 34 more or less ambiguous instances of (referential) *to* out of the study.

and warm. If we count all the (co-ordinated) adjectives, the number of instances included in our study is 254 for Czech, 456 for English and 364 for Norwegian (see Table A in the Appendix). It can be noted that this phenomenon is proportionally more frequently attested in English than in the other two languages (i.e., proportions calculated against the actual (raw) number of instances of this pattern in each language).

5. Case study of the '(NP) + BÝT/BE/VÆRE + ADJ' pattern

As reflected in our second research question, we want to take a closer look at the instances where a subject is described by means of an adjectival complement, as this is the most frequent linking pattern in all three languages. Since we are working with fiction texts, we assume that this pattern may reveal differences and similarities in the way authors in the three languages describe characters, objects, events, and places by means of predicative adjectives. Therefore we have classified the adjectives into semantic classes (see section 3.2). Figure 1 gives an overview of the distribution of these across the three languages (see also Table A in the Appendix for raw figures). In addition, we have categorised the (understood) subject in each instance of the pattern as human or non-human. Figure 2, further below, illustrates the correlation between the semantic category of the adjectives and their subjects in all three languages (see also Table B in the Appendix for raw figures).

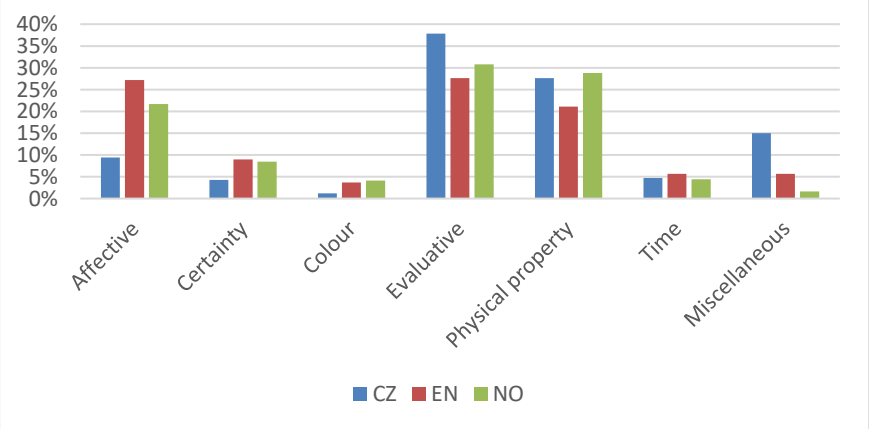

Figure 1. Proportions of the semantic classes of adjectives in Czech, English and Norwegian (as percent of the total number of adjectives in the pattern within each language)

As shown in Figure 1, the predominant semantic categories in all three languages are evaluative and physical property, and to some extent affective. Nevertheless, there are some proportional differences between the languages with regard to these preferred categories.

- Affective (CZ: 9.4%, EN: 27.2%, NO: 21.7%)
- Evaluative (CZ: 37.8%, EN: 27.6%, NO: 30.8%)
- Physical property (CZ: 27.6%, EN: 21.1%, NO: 28.8%)

Most notable is the difference in the use of affective adjectives in Czech compared to the other two languages, accounting for only 9.4% cases compared to 27.2% in English and 21.7% in Norwegian. Moreover, it can be noted that while Czech clearly favours evaluative adjectives, there is more of a division of labour between affective and evaluative in English, and between evaluative and physical property in Norwegian. However, it is also the case that, compared to the other languages, English makes relatively more use of affective adjectives, whereas Czech and Norwegian make more use of physical property adjectives.

Another observation worth making is that Czech differs from the other two languages in most of the other categories, namely certainty, colour, and miscellaneous. English and Norwegian, on the other hand, are very similar to each other, with the exception of the miscellaneous category which is markedly more frequent in English. To illustrate each of the semantic classes, one example from the language in which the category is proportionally most frequently attested is given in (33)–(39).

(33) Affective: Mary Louise herself *was* terrified. (ICC-EN)

(34) Certainty: Maybe the General *is* right. (ICC-EN)

(35) Colour: nå gjør det ingenting at de *er* gule. (ICC-NO)
['now it doesn't matter that they *are* yellow']

(36) Evaluative: Ten pirát *byl* hezkej, co? (ICC-CZ)
['The pirate *was* handsome, wasn't he?']

(37) Physical Property: Luften *var* svært tørr. (ICC-NO)
['The air *was* very dry']

(38) Time: I had *been* much <u>older</u> then. (ICC-EN)

(39) Miscellaneous: Věděla, že *je* <u>rozvedený</u>. (ICC-CZ)
['She knew he *was* <u>divorced</u>.']

As noted above, regarding the type of subjects, we recorded the distribution of human vs. non-human subject in the pattern; Table 3 gives an overview of this.

Table 3. Distribution of human (HS) vs. non-human (NHS) NP-subjects in the pattern

CZ/HS	CZ/NHS	EN/HS	EN/NHS	NO/HS	NO/NHS
150 (59.1%)	104 (40.9%)	261 (57.2%)	195 (42.8%)	179 (49.2%)	185 (50.8%)
254		456		364	

As we can observe from Table 3, Czech and English are very similar in the types of subjects they tend to describe in this pattern, with roughly a 60-40 split in favour of human subjects. Norwegian, however, has more of a 50-50 split, very marginally in favour of non-human subjects. It is not entirely clear how to interpret this, as it may have more to do with the content of the sub-corpora than with linguistic preferences. It will perhaps be more revealing to see what kind of adjectives are used to describe human vs. non-human subjects in the material.

After establishing the overall picture of the distribution of adjective classes, as outlined above, we examined the instances more carefully, and, perhaps not unexpectedly, it becomes clear that the type of subject the adjectives describe clearly has an impact on the semantic class. This is shown in Figure 2, where the correlation is calculated separately per language per semantic category. For example, in Norwegian, 73 out of 79 (92.4%) adjectives in the affective category are used to describe a human being. The remaining six adjectives describe a non-human subject (i.e., 7.6% of the cases). This is illustrated by the black vs. grey bars (NO/HS vs. NO/NHS) for the affective category in Figure 2 (see also Table B in the Appendix for raw figures). In Figure 2, human subject is represented by the dark colours and non-human subject by the light colours in each language (blue scale = CZ, red scale = EN and grey scale = NO).

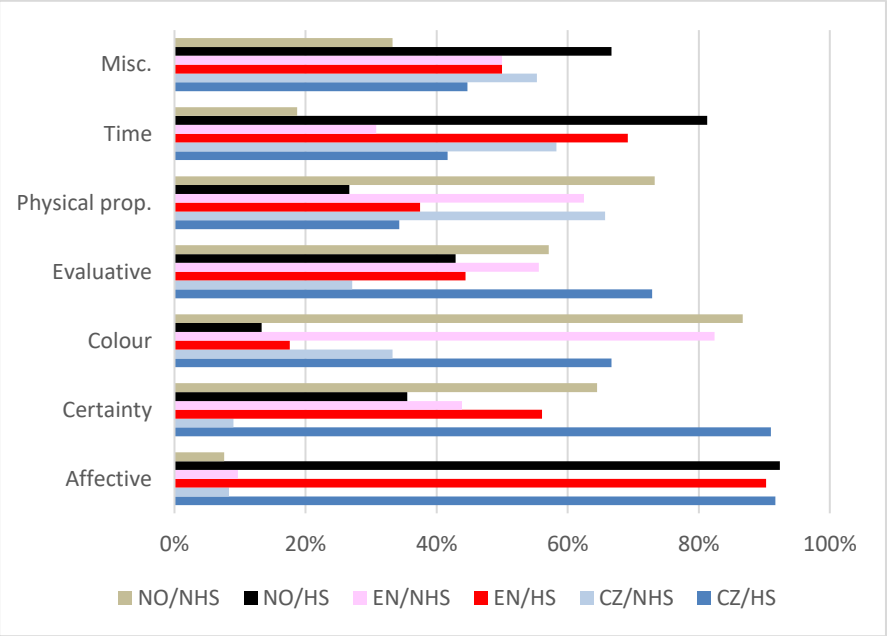

Figure 2. Correlating the semantic classes of adjective with human/non-human subject in the three languages

When correlating type of subject with the different adjective classes, we can observe the following:

- A human subject overwhelmingly attracts affective adjective in all three languages: see example (40):

(40) We all used to *be* much more scared. (ICC-EN)

- A human subject overwhelmingly attracts time adjective in English and Norwegian, whereas Czech prefers a non-human subject, as in (41)–(42):

(41) Han *var* eldre enn dem [‘he *was* older than them’] (ICC-NO)

(42) Mnohá jeho poranění *jsou* stará [‘many of his wounds *are* old’] (ICC-CZ)

- A non-human subject typically attracts colour adjective in English and Norwegian, as in (43), while in Czech a human subject is preferred (but there are very few occurrences overall):

(43) His skin *was* white. (ICC-EN)

- A non-human subject typically attracts evaluative adjective in English and Norwegian, as in (44), whereas in Czech evaluative adjectives typically occur with human subjects: see (45):

(44) This *is* so boring. (ICC-EN)

(45) Ale sestra *je* nudná ['but my sister *is* boring'] (ICC-CZ)

- A non-human subject typically attracts physical property adjective in all three languages, but most frequently so in Norwegian:

(46) Luften *var* svært tørr ['the air *was* very dry'] (ICC-NO).

- A human subject overwhelmingly attracts certainty adjectives in Czech, and to some extent also in English, whereas Norwegian prefers a non-human subject in this category:

(47) Ale *nejsem* si jistý. ['But I *am not* sure.'] (ICC-CZ)

(48) men det siste *er* minst sannsynlig. ['but the last *is* least probable'] (ICC-NO)

To summarise, our analysis suggests that the three languages resort to similar ways of describing fictional subjects by means of predicative adjectives. It seems that fiction is a relatively homogeneous register in this respect (at least fiction in these three European languages). However, there are some differences in what the different types of adjectives tend to describe in the individual languages.

The detailed analysis of the 'NP/Ø+V+ADJP' pattern did not reveal particular uses that can explain its proportionally more frequent use in English fiction. In other words, English fiction simply seems to have a stronger preference for this pattern than Czech and Norwegian fiction, and

potential reasons for the proportional discrepancy between English and the other two languages seem to lie at a higher level of description, i.e., syntactic choice rather than semantic.[17]

A very general and tentative cross-linguistic conclusion regarding the nature of fictional characterization using this pattern is that Czech and Norwegian are the furthest apart in terms of what kind of adjective they prefer to describe a human or non-human subject. There is more overlap between English and Norwegian than there is between English and Czech.

6. Concluding remarks

In this contrastive study based on material drawn from the fiction component of the International Comparable Corpus (ICC), we compared the use of the most frequent verbs, i.e., the 'be' verbs (BÝT, BE and VÆRE) in Czech, English and Norwegian. We established a simple, but cross-linguistically comparable, taxonomy of the uses of the three verbs, namely auxiliary, linking and 'other' uses. Following a quantitative analysis of these uses, we moved on to look at linking uses and, in particular, the linking use where the complement is made up of an adjective phrase. Based on the material and earlier research into the semantics and use of adjectives, we classified our adjective phrases into seven categories: affective, certainty, colour, evaluative, physical property, time and miscellaneous.

This analysis provided some tentative answers to our research questions. Our first research question was: To what extent do BÝT, BE and VÆRE overlap in meaning and use? As shown in Table 1 (see section 4), the proportional distribution of the various uses of the verbs is somewhat different between the languages. English and Norwegian show an overwhelming preference for the linking uses, whereas Czech has a higher proportion of auxiliary uses (which may be due to the text type investigated), with only a marginal preference for linking uses. Moreover, it was confirmed that there are few overlapping auxiliary uses of the verbs;

[17] Alternatively, it could be related to the proportions of dialogic vs. narrative passages in the corpus texts, as a recent study (Ebeling forthc.) suggests that predicative adjectives are more frequently used in dialogue than in narrative in both English and Norwegian. To investigate this further lies outside the scope of the current paper, as the ICC texts have not been marked up for narrative vs. dialogic passages.

the only exception is their use as a passive auxiliary. The use of BÝT in Czech to mark the past tense, the use of BE to form the progressive aspect in English and VÆRE to form the perfect aspect in Norwegian, show the great potential of these 'be' verbs as carriers of tense and aspect, in addition to voice.

A different picture emerged when we examined linking uses, as the three languages showed great overlap in their use of these verbs to link subject and complement. However, we also identified a number of linking uses that, due to typological differences, i.e., fixed (English/Norwegian) vs. free (Czech) word order, are less straightforward to compare: these were primarily the frequent use of 'dummy S constructions', where 'dummy S' is typically realised by *it* and *det* in English and Norwegian respectively, while a direct counterpart does not exist in Czech. An interesting finding in this respect is the many occurrences of *to*/*tohle* in Czech, which on the surface seem similar to the many dummy subject constructions in English and Norwegian.

To answer our second research question, we homed in on one of the linking uses, namely those with an adjective complement, and we asked what kind of relationship each of the verbs typically establishes between the elements that are linked. The short answer is that verbs behave similarly in how subject NPs are described, with some cross-linguistic differences or preferences, and that the use of predicative adjectives seem to be a more defining feature of English fiction than of Czech and Norwegian (see Figure 1 in section 5). Since we are dealing with the same text type in all three languages, i.e., prose fiction, we could perhaps not expect to find any great differences in what is described by adjectives and how subjects, human and non-human, are described. There are, however, some proportional tendencies that distinguish between the languages and that would be fruitful to investigate further based on a larger dataset and other text types or genres. These include the following: is Czech more concerned with evaluating human subjects than English and Norwegian, and why?; are English and Norwegian more concerned with describing human subjects in terms of 'time', e.g., age, and why?

Our third and final research question was: In a cross-linguistic comparison of 'be' verbs in typologically different languages, what are the methodological challenges? We can report on various challenges that are important to have in mind when embarking on grammatical studies using comparable corpora:

1. Quantitatively, the material for the in-depth, qualitative analysis became relatively small when classified into the various adjective categories and due to syntactic and morphological differences between the languages, e.g., pro-drop, definiteness and compounding, more data is needed for these aspects to be fully fleshed out and explored.
2. Qualitatively, we encountered challenges regarding terminology and grammatical descriptive frameworks, apparatus or tradition for the three languages.
3. Another challenge relates to the actual classification of the adjectives encountered. Since the material is limited, many adjectives only occurred once and it was difficult to see a pattern or pin down the exact meaning based on the material at hand.
4. Finally, the subject matter of the text extracts in the sub-corpora should perhaps not be underestimated, as these may have a bearing on whether, e.g., people or objects are most frequently described, and how they are described or characterised.

This investigation also serves as a pilot study to test the potential of small, comparable, datasets for cross-linguistic research, moving from the more traditional contrastive research of language pairs to a more ambitious comparison of three (typologically different) languages. In this way we were able to point to marked proportional differences in how the three languages exploit the grammatical and functional potential of their respective 'be' verbs. In addition, the linking use was found to behave similarly across the languages, but with some minor and potentially language-specific preferences of use. In this way, the cross-linguistic comparison has given more insight into the preferred behaviour of 'be' verbs in three individual languages and when and how their behaviour overlaps across the languages. This, in turn, lends empirical evidence to how the verbs have taken different or similar paths in their development. The methodology highlights the importance of a carefully crafted *tertium comparationis* at several levels, not only in relation to datasets and linguistic phenomena investigated, but also terminology and grammatical traditions of description for the individual languages.

Our case-study points to some potential limitations of comparable data in this type of cross-linguistic analysis. While we were able to identify comparable categories and establish quantitative trends, the qualitative

analysis is somewhat limited in the sense that only one genre was compared. To complete the picture of predicative complementation across the three languages, a detailed scrutiny of other predicative elements needs to be conducted in the future. Also, there may be some limitations related to the interpretation of one of our findings in particular, namely the fact that English makes more use of the 'NP/Ø+V+ADJP' pattern than Norwegian and Czech.[18] Here are some hypotheses and questions for further research:

- Norwegian/Czech fiction makes more frequent use of attributive adjectives to convey the same message? (e.g., *The moon is blue* vs. *Den blå månen* 'the blue moon')
- Norwegian/Czech fiction makes use of other verbs than BÝT and VÆRE to get the same message across? (e.g., *The moon is blue* vs. *Månen ser blå ut* 'the moon looks blue')
- English fiction is generally more concerned with characterising or describing the Subject in terms of feelings, evaluations and other properties?

To investigate, and be able to test, these hypotheses, we need a broader scope, as well as a bigger and more representative corpus and perhaps, ideally, a tri-directional, parallel translation corpus, that will offer comparable data alongside translation paradigms that may point to alternative and idiomatic ways of expressing a similar relationship to that of the 'NP/Ø+V+ADJP' in Norwegian and Czech. Such a corpus was envisaged by Johansson in the late 1990s and its adaptation for our purposes is illustrated in Figure 3.

[18] A similar tendency is noted for the language pair English and Spanish in Pérez Blanco's (2016) study of negative evaluative adjectives in newspaper opinion pieces: English uses the predicative function more often than Spanish. In this case, the explanation for the discrepancy seems to lie in the fact that Spanish has a strong preference for adjectives as postmodifiers of nouns, e.g., *Luisa es una jefa mala* ['Luisa is a boss bad'].

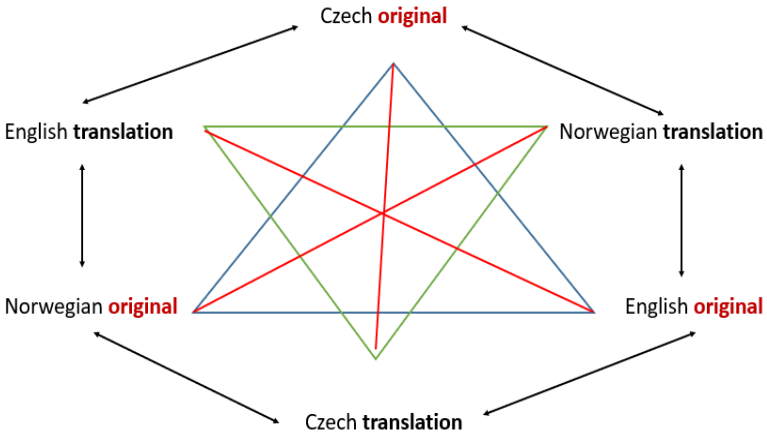

Figure 3. Tri-directional translation corpus model with the languages Czech-English-Norwegian (based on Johansson 2000)

With reference to Figure 3, the present study works along the blue triangle (comparable corpus model), whereas a tri-directional translation corpus would add the green triangle, arguably enabling more robust contrastive studies between the three languages in the sense that other similar expressions, including non-congruent ones, would come to light through translation correspondence.[19]

Appendix

Table A. The frequency of the semantic classes of adjectives in Czech, English and Norwegian (raw figures and percentages of total number of adjectives in the pattern)

Semantic class of adjective	**CZ**	**EN**	**NO**
Affective	24 (9.4%)	124 (27.2%)	79 (21.6%)
Certainty	11 (4.3%)	41 (9%)	31 (8.5%)
Colour	3 (1.2%)	17 (3.7%)	15 (4.1%)
Evaluative	96 (37.8%)	126 (27.6%)	112 (30.7%)
Physical property	70 (27.6%)	96 (21.1%)	105 (28,8%)
Time	12 (4.7%)]	26 (5.7%)	16 (4.4%)
Misc.	38 (15%)	26 (5.7%)	6 (1.6%)
Total	254	456	365

[19] Three unidirectional translation corpora, operating along the outer black arrows of Figure 3, would add similar, albeit not as detailed, insights.

Table B. The correlation between type of subject and adjective class in raw figures and percentages

	CZ / HS	CZ / NHS	CZ (total)	EN / HS	EN / NHS	EN (total)	NO / HS	NO / NHS	NO (total)
Affective	22 (91.7%)	2 (8.3%)	24 (100%)	112 (90.3%)	12 (9.7%)	124 (100%)	73 (92.4%)	6 (7.6%)	79 (100%)
Certainty	10 (91%)	1 (9%)	11 (100%)	23 (56.1%)	18 (43.9%)	41 (100%)	11 (35.5%)	20 (64.5%)	31 (100%)
Colour	2 (66.7%)	1 (33.3%)	3 (100%)	3 (17.6%)	14 (82.4%)	17 (100%)	2 (13.3%)	13 (86.7%)	15 (100%)
Evaluative	70 (72.9%)	26 (27.1%)	96 (100%)	56 (44.4%)	70 (55.6%)	126 (100%)	48 (42.9%)	64 (57.1%)	112 (100%)
Phys. prop.	24 (34.3%)	46 (65.7%)	70 (100%)	36 (37.5%)	60 (62.5%)	96 (100%)	28 (26.7%)	77 (73.3%)	105 (100%)
Time	5 (41.7%)	7 (58.3%)	12 (100%)	18 (69.2%)	8 (30.8%)	26 (100%)	13 (81.3%)	3 (18.7%)	16 (100%)
Misc.	17 (44.7%)	21 (55.3%)	38 (100%)	13 (50%)	13 (50%)	26 (100%)	4 (66.7%)	2 (33.3%)	6 (100%)

References

Adamec, Přemysl. 1998. K syntaxi a sémantice „malých slov“: slůvko *to*. In *Přednášky z XLI. běhu LŠSS*: 55–58. Prague: Charles University.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.

Bjorvand, Harald, and Fredrik Otto Lindeman. 2019. *Våre arveord. Etymologisk ordbok*. 3rd ed. Oslo: Novus.

Bybee, Joan L., and Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1): 51–103.

Čermák, František. 2010. *Lexikon a sémantika*. Prague: Nakladatelství Lidové Noviny.

Čermáková, Anna, Jarmo Jantunen, Tommi Jauhiainen, John Kirk, Michal Křen, Marc Kupietz, and Elaine Uí Dhonnchadha. 2021. International Comparable Corpus: Challenges in building multilingual spoken and written comparable corpora. *Research in Corpus Linguistics* 9(1): 89–103. https://ricl.aelinco.es/index.php/ricl/article/view/155.

Cvrček, Václav, Vilém Kodýtek, Marie Kopřivová, Dominika Kováříková, Petr Sgall, Michal Šulc, Jan Táborský, Jan Volín, and Martina Waclawičová. 2015. *Mluvnice současné češtiny.* Prague: Karolinum.

Daneš, František, Zdeněk Hlavsa, and Miroslav Grepl. 1987. *Mluvnice češtiny* 3. *Skladba.* Prague: Academia.

Dixon, Robert M. W. 2010. *Basic linguistic theory. Volume 2: Grammatical topics*. Oxford: Oxford University Press.

Dušková, Libuše. 2004. Syntactic constancy of the subject complement. Part 1: A comparison between Czech and English. *Linguistica Pragensia* 14(2): 1–17.

Dušková, Libuše. 2005. Syntactic constancy of the subject complement. Part 2: A comparison between English and Czech. *Linguistica Pragensia* 15(1): 1–17.

Dušková, Libuše, Markéta Malá, Pavlína Šaldová, and Gabriela Brůhová 2006. *Mluvnice současné angličtiny na pozadí češtiny.* Prague: Academia. https://mluvniceanglictiny.cz/.

Ebeling, Jarle. 2000. *Presentative constructions in English and Norwegian. A corpus-based contrastive study*. Oslo: Acta Humaniora.

Ebeling, Signe Oksefjell, and Jarle Ebeling. 2020. Contrastive analysis, *tertium comparationis* and corpora. *Nordic Journal of English Studies* 19(1): 97–117.

Ebeling, Signe Oksefjell. Forthc. Structural and semantic features of adjectives across languages and registers. To appear in *Languages in Contrast*, special issue edited by Magnus Levin and Jenny Ström Herold.

Faarlund, Jan Terje, Sven Lie, and Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Fjeld, Ruth E. V., Anders Nøklestad, and Kristin Hagen. 2020. Leksikografisk Bokmålskorpus (LBK): bakgrunn og bruk. *Oslo Studies in Language* 11(1): 101–124. https://journals.uio.no/osla/article/view/8176.

Greenbaum, Sidney. 1991. The development of the international corpus of English. In *English corpus linguistics*, edited by Karin Aijmer and Bengt Altenberg, 83–91. London: Longman.

Gundel, Jeanette. 2002. Information structure and the use of cleft sentences in English and Norwegian. In *Information structure in a cross-linguistic perspective*, edited by Hilde Hasselgård, Stig Johansson, Bergljot Behrens, and Cathrine Fabricius-Hansen, 113–128. Amsterdam: Rodopi.

Havránek, Bohumil (ed.). 1989. *Slovník spisovného jazyka českého*. 2nd ed. Prague: Academia.

Havránek, Bohumil, and Alois Jedlička. 1960. *Česká mluvnice*. Prague: SPN.

Huddleston, Rodney, and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Johansson, Stig. 2000. Towards a multilingual corpus for contrastive analysis and translation studies. SPRIKreports 1 (reports from the project Språk i kontrast [Languages in Contrast]). University of Oslo.

Kallen, Jeffrey L., and John M. Kirk. 2008. *ICE-Ireland: A user's guide. Documentation to accompany the Ireland component of the International Corpus of English* (ICE-Ireland). Queen's University Belfast, Trinity College Dublin, and Cló Ollscoil na Banríona. https://johnmkirk.etinu.net/johnmkirk/documents/003647.pdf.

Kirk, John, Anna Čermáková, Signe Oksefjell Ebeling, Jarle Ebeling, Michal Křen, Karin Aijmer, Vladimir Benko, Radovan Garabík, Rafał

Górski, Jarmo Jantunen, Marc Kupietz, Mária Šimková, Thomas Schmidt, and Oliver Wicher. 2018. Poster presentation at the 2018 Using Corpora in Contrastive and Translation Studies (UCCTS) conference in Louvain-la-Neuve, Belgium.

Klimešová, Petra, Zuzana Komrsková, Marie Kopřivová, and David Lukeš. 2015. Slovo *to* v mluvených korpusech ČNK, jeho prefixace a reduplikace. *Časopis pro moderní filologii* 97(1): 21–30.

Lorenz, Gunter R. 1999. *Adjective intensification – Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodopi.

Malá, Markéta 2014. *English copular verbs. A contrastive corpus-supported view*. Prague: FFUK.

NAOB. *Det Norske Akademis Ordbok*. Det Norske Akademi for Språk og Litteratur. https://naob.no/.

OED = *Oxford English Dictionary* online. https://www.oed.com/.

Pérez Blanco, María. 2016. The construction of attitudinal stance: A corpus-based contrastive study of negative evaluative adjectives in English and Spanish opinion discourse. *Languages in Contrast* 16(1): 31–53.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Rejzek, Jiří 2015. *Český etymologický slovník*. 3rd ed. Prague: Leda.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.

Verhaar, J. W. M. (ed.). 1967–1972. *The verb 'be' and its synonyms*: *Philosophical and grammatical Studies* Parts I–V. Dordrecht: D. Reidel Publishing.