

BE verbs in a contrastive perspective: The case of BÝT, BE and VÆRE

Anna Čermáková, Jarle Ebeling, Signe Oksefjell Ebeling



FACULTY OF ARTS
Charles University



UNIVERSITY
OF OSLO

Outline

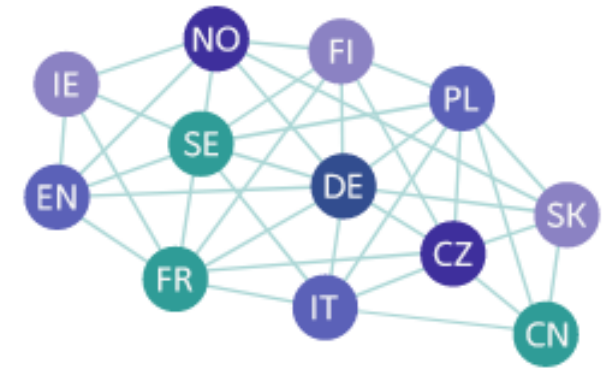
- Aims and RQs
- Material and method
 - The International Comparable Corpus (ICC)
 - Contrastive comparable method
- Background
 - BE verbs: classification
 - BE verbs: uses in ICC-CZ, ICC-EN, ICC-NO (creative writing part)
- Contrastive analysis of the linking function (NP/Ø+V+ADJ)
- Conclusion and further research

Aims and RQs

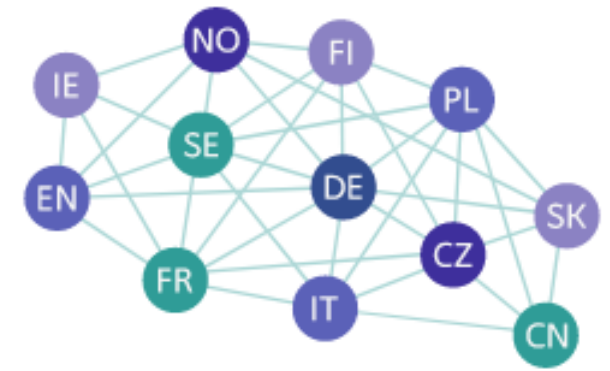
1. To what extent do BÝT, BE and VÆRE overlap in meaning and use?
2. In the **linking use with an adjectival complement**, what kind of relationship does each of the verbs typically establish between the elements that are linked?
 - E.g. He **was** never **judgemental** ...
3. In a cross-linguistic comparison of BE verbs in typologically different languages, what are the **methodological challenges**?



International Comparable Corpus



- 12 teams / languages
- 1 million words per ICC component
- 60% spoken data, 40% written data in each sub-corpus
 - following the design of the International Corpus of English in terms of text types, sampling and size
 - Reuse of already existing corpus material where possible
 - <https://korpus.cz/icc>



This study:

- **Creative writing** section of ICC-CZ, ICC-EN, ICC-NO
- 40,000 words from each language (20 text extracts of 2,000 words in each ICC component)
- Word frequency list: the prototypical **verb of being** is the most frequent verb in all three languages: 1,842 (CZ), 1,742 (EN), 1,292 (NO).

Contrastive comparable method

- Criteria of comparability that ensure a sound *tertium comparationis*:
 - **Text type** = creative writing (novels and short stories);
 - **Time period** = contemporary (1990s-2010s);
 - **Object of study** = the three BE verbs are etymologically and functionally related; thus, there is an **underlying perceived similarity** that may serve as a starting point for a CA

Background: BE verbs

In a contrastive study of items in languages as different as Czech vs. English and Norwegian, it is important to find a common ground (terminology) to make sure we compare like with like.

- E.g. "copular verb" may evoke different conceptualisations both within and across the three languages.
 - Quirk et al. (1985):
 - Copular = SVsP and SVA
 - Intransitive = SV
 - Huddleston & Pullum (2002):
 - Copular = SVsP
 - Intransitive = SV and SVA

"Linking"

SVsP: The country **is** independent.
SVA: I have **been** in the garden

Background: BE verbs

Linking

- (S)+BE+sP
- (S)+BE+A
- Dummy/empty subject constructions

Auxiliary

- Passive voice
- (Czech): past & future tense / conditional
- (English): progressive aspect
- (Norw.): perfect aspect

Other uses:

- Intransitive (*the father to be*)
- Multi-word uses (*e.g. be going to, be bound to*)

Case study: BÝT, BE and VÆRE in the creative writing section of ICC

Table 1. Propotional distribution of BÝT, BE and VÆRE as linking, aux. and other uses

Verb	Linking	Aux.	Other	Tot.
BÝT	894 / 49%	882 / 48%	66 / 3%	1,842
BE	1,251 / 72%	454 / 26%	37 / 2%	1,742
VÆRE	1,159 / 90%	86 / 7%	47 / 3%	1,292

'Other' includes phrasal uses, idioms and intransitives

Linking: SVsP (sP = ADJP)

- (1a) Vlak **je** zrezivělý, a tohle nástupiště **je** teď pusté. (ICC-CZ)
[‘The train is rusty and this platform is now deserted.’]
- (1b) Facilities **were** rather spartan ... (ICC-EN)
- (1c) Hun visste at hunden **var** adskillig reddere. (ICC-NO)
[‘She knew that the dog was considerably more scared’]

Linking: SVA

- (2a) Auto zaburácelo a **bylo** pryč. (ICC-CZ)
[‘The car roared and was away.’]
- (2b) ..., and still you **are** not here. (ICC-EN)
- (2c) Vis oss at du heller vil **være** her enn ... (ICC-NO)
[‘Show us that you would rather be here than ...’]

Linking: dummy S constructions, e.g. constructions with *to* ‘it’, extraposition *it*, existential *det* ‘there’

- (3a) ... to **by** **bylo** zbytečný. (ICC-CZ) [‘... it would be useless.’]
- (3b) It **was** sad to see so many of one's school pals gathering at the corner ... (ICC-EN)
- (3c) ... 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’]

Auxiliary: Passive voice

- (4a) Je původní, potvrzuje dnešní majitel, který sem **byl** také před třiceti roky přenesen z Evropy. (ICC-CZ)
[‘It is original, confirms the contemporary owner, it *was brought* over here thirty years ago from Europe’]
- (4b) The autumn evenings **were** marked by the Listowel races... (ICC-EN)
- (4c) ... brua **var** festet med store bolter. (ICC-NO)
[‘the bridge *was secured* with large bolts’]

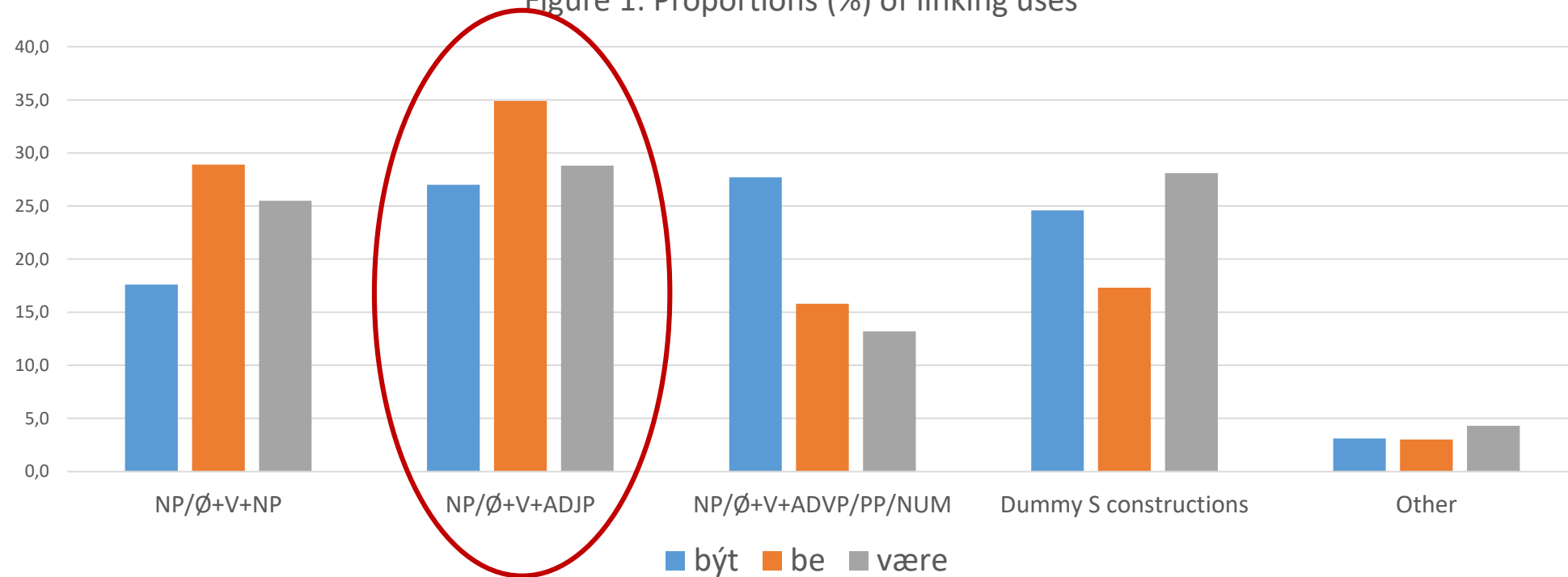
Auxiliary: past tense (CZ), progressive aspect (EN), perfect aspect (NO)

- (5a) Rozsvítil **jsem** modrou lampičku a posadil se na posteli. (ICC-CZ)
[‘I turned on the blue lamp and sat up on my bed’]
- (5b) Somewhere a baby **was** crying. (ICC-EN)
- (5c) Det **er** blitt sent på natten. (ICC-NO)
[‘It is become late at night.’]

Table 3. Linking uses of BÝT, BE and VÆRE

Verb	NP/∅+V+NP	NP/∅+V+ADJP	NP/∅+V+ADVP/PP/NUM	Dummy S constructions	Other	Total
BÝT	157	241	248	220	28	894
BE	362	436	198	217	38	1,251
VÆRE	296	334	153	326	50	1,159

Figure 1: Proportions (%) of linking uses



Classification of the S+V+ADJP pattern

- Subject: human vs. non-human (NP or understood S)
- ADJP: head adjective classified into the following classes (based on Biber et al. (1999) and Lorenz (1999)):
 - Certainty: e.g. *clear, likely, obvious, true*
 - Affective (affective psychological states & 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*
 - Size/Amount: e.g. *big, low, short*
 - Physical property: e.g. *cold, flushed, wrinkled*
 - Miscellaneous: e.g. *professional, missing, racist*

Adjectival categories

The most frequent semantic categories in all three languages are affective, evaluative and physical property

Proportionally:

- Affective (**CZ: 16.2%**, EN: 26.1%, NO: 23.2%)
- Evaluative (CZ: 29.5%, EN: 25.5%, NO: 28%)
- Physical property (CZ: 19.9%, EN: 17.4%, NO: 21.7%)

Type of subject: human/non-human + ADJ

- Human subject overwhelmingly attracts affective adjective in EN and NO
 - *We all used to be much more scared.*
- Non-human subject typically attracts colour adjective in EN and NO – but few occurrences overall
 - *His skin was white.*
- Non-human subject typically attracts evaluative adjective in EN and NO
 - *This is so boring.*
- Human subject more typically attracts evaluative adjective in CZ
 - *Ale sestra je nudná ...* ['but my sister is boring...']
- Non-human subject typically attracts physical property adjective in all three languages, but more frequently so in CZ and NO
 - *My tunic was wrinkled.*

Concluding remarks on case study

- The detailed analysis of the S+V+ADJP pattern did not reveal particular uses that can explain its proportionally more frequent use in English fiction.
- The analysis suggests that the three languages resort to similar ways of describing fictional subjects by means of adjectives.
 - Fiction is a homogeneous register in this respect (at least fiction in these three European languages)
- A potential reason for the proportional discrepancy between English and the other two languages seems to lie at a higher level of description, i.e. syntactic choice rather than semantic.
 - In other words, English fiction simply has a stronger preference for the S+V+ADJP pattern than Czech and Norwegian fiction

Conclusion

RQ1: Degree of overlap between BE verbs

- Few overlapping auxiliary uses (the passive an exception)
- Similar linking uses, but different proportions of preferred uses

RQ2: Relationship established by the verbs in their Linking use

- Behave similarly in how Subjects are described with some minor differences/tendencies
- Predicative adjectives a more defining feature of English fiction than of Czech and Norwegian (cf. Figure 1)

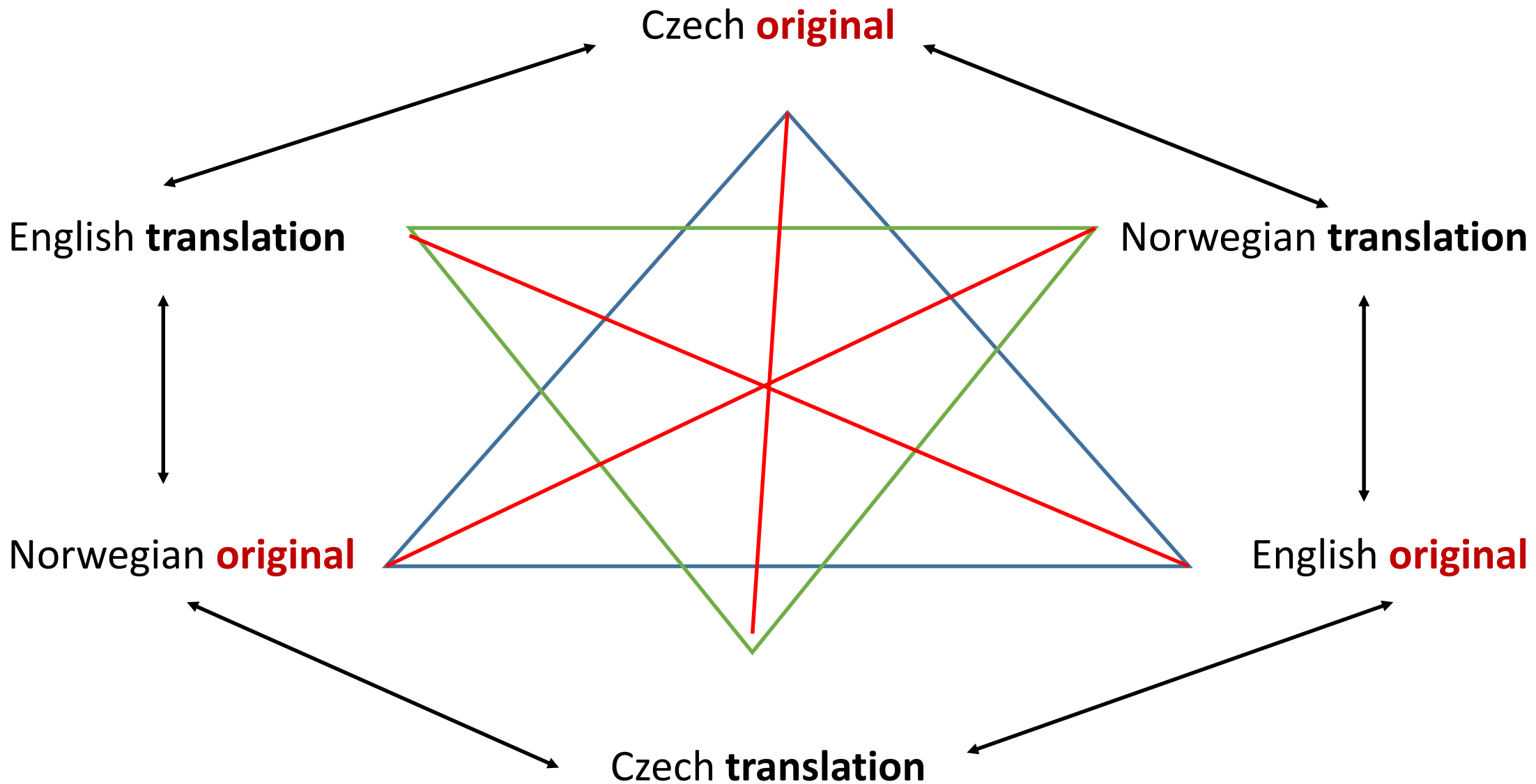
RQ3: Methodological challenges

- Quantitatively: e.g. syntactic/morphological differences between the languages (pro-drop, definiteness, compounding)
- Qualitatively: terminology and grammatical descriptive framework/apparatus/tradition and (native-like) knowledge of all languages (i.e. none of us knows all three languages).
- Other issues: annotation practices

Further research/Hypotheses

Why does English make use of the NP/Ø+V+ADJP pattern more often than Norwegian and Czech:

- Norwegian/Czech fiction makes more frequent use of attributive adjectives to convey the same message?
- Norwegian/Czech fiction makes use of other verbs than BÝT and VÆRE to get the same message across?
- English fiction is generally more concerned with characterising/describing the Subject in terms of feelings, evaluations and other properties?



Based on Johansson (2000)

References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bjorvand, H. & F.O. Lindeman. 2019. *Våre arveord. Etymologisk ordbok*. [3rd ed.] Oslo: Novus.
- Bybee, J. L. & Ö. Dahl. 1989. The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language* 13 (1), 51–103.
- Čermáková, A., Jantunen, J., Jauhiainen, T., Kirk, J., Kren, M., Kupietz, M. & Uí Dhonnchadha, E. 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>
- Huddleston, R. & G. K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Johansson, S. 2000. Towards a multilingual corpus for contrastive analysis and translation studies. SPRIKreports, number 1 (reports from the project Languages in Contrast (**Språk i kontrast**)). University of Oslo.
- Kirk, J., A. Čermáková, S.O. Ebeling, J. Ebeling, M. Křen, K. Aijmer, V. Benko, R. Garabík, R. Górski, J. Jantunen, M. Kupietz, M. Šimková, T. Schmidt & O. Wicher. 2018. Poster presentation at the *UCCTS 2018 (UCCTS 2018 'Using Corpora in Contrastive and Translation Studies')* in Louvain-la-Neuve.
- Lorenz, Gunter R. 1999. *Adjective Intensification – Learners Versus Native Speakers: A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- OED Online. 2021. "be, v.". Oxford University Press. <https://www-oed-com.ezproxy.uio.no/view/Entry/16441?rskey=n7iZuu&result=4&isAdvanced=false> (accessed February 06, 2020).
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rejzek, J. 2015. *Český etymologický slovník*. [3rd ed.] Praha: Leda.