Prepositional phraseological patterns in Czech and English

Towards a contrastive study resource[1]

Denisa Šebestová

Charles University, Prague (Czech Republic)

This pilot study aims to identify differences in native and non-native phraseologies, focussing on prepositional patterns. Previous research suggests L2 users' limited phraseological choices may hinder the accuracy of their language production, and prepositions can pose a particular challenge to Czech learners of English, given the lack of correspondence between translation equivalents. Further, prepositional patterns contribute to text structuring, making them an important part of learners´ competence. Using representative corpora of English and Czech, 3- to 5-grams containing the equivalent preposition pair *in*/*v* are extracted. The identified patterns are classified by their semantics and textual functions. While *in/v* patterns mostly fulfil corresponding functions in the languages compared, the distribution of these functions differs. Specifically, some pattern types are only found in English, highlighting its analytic nature as opposed to inflectional Czech.

Keywords: n-grams, prepositions, native and non-native phraseology, typologically distant language pair, Czech/English

1. Introduction

This study is based in cross-linguistic distributional (Granger and Paquot, 2008) or data-driven (Granger and Meunier (eds), 2008) phraseology, i.e. examining recurrent word combinations through corpora. It was prompted by earlier findings provided by research into non-native phraseology (Ebeling and Hasselgård, 2015; Granger, 2017; Granger and Bestgen, 2014; Hasselgård, 2019; Vašků, Brůhová, and Šebestová, 2019), as well as by the interest in – and need for – teaching materials reflecting those findings (Reppen, 2011). It is conceived as a pilot study, aiming to contrast a selected pattern group – prepositional patterns – between the typologically distant language pair of Czech and English. The results of this contrastive analysis can then be used as a springboard towards suggesting how n-gram based studies of phraseology can inform foreign language instruction.

[1] This research was funded by the Faculty of Arts, Charles University, within the project 'Specifický vysokoškolský výzkum - Jazyk a nástroje pro jeho zkoumání' (2020).

CROSSING THE BORDERS: ANALYSING COMPLEX CONTRASTIVE DATA. Edited by Anna Čermáková, Signe Oksefjell Ebeling, Magnus Levin and Jenny Ström Herold. *BeLLS* Vol 11, No 1 (2021), DOI: 10.15845/bells.v11i1.3437.

Section 2 introduces the theoretical background and motivation for the study. Section 3 introduces the material and methods employed in the study. Section 4 presents the textual functions conveyed by prepositional patterns in the English and Czech data. Results are described for each language separately. Section 5 reports on differences in pattern usage between the two languages. Finally, Section 6 summarizes the results and suggests potential avenues for further research.

2. Background and motivation

Phraseology (in the sense of the use of recurrent word combinations, cf. Gray and Biber, 2015: 125; Ebeling and Hasselgård, 2015: 207) has been shown to "unmistakably [distinguish] native speakers of a language from L2 learners" including advanced learners (Granger and Bestgen, 2014:,229). It has been suggested that L2 learners have a limited repertoire of phraseological sequences, and employ these in ways which differ considerably from native usage (Granger, 2017). As a result, L2 learners tend to overuse a restricted set of phraseological sequences which they have mastered and feel confident using. Hasselgård (2019) terms these 'phraseological teddy bears', referring back to Hasselgren's (1994) idea of 'lexical teddy bears'.

These limitations have a serious bearing on the learner's communicative skills: they pose a potential hindrance to language production, since phraseological competence forms a crucial part of a learner's overall language proficiency (Howarth, 1998; Hyland, 2008; Paquot, 2018; Paquot and Granger, 2012). The degree of phraseological competence is also an important criterion in determining L2 fluency, distinguishing native speakers from non-native learners (Granger and Bestgen, 2014; Hasselgård, 2019). Moreover, becoming acquainted with recurrent word combinations is important as they form a major component of everyday language use (Biber *et al*., 2004; Erman and Warren, 2000).

One way to address this issue is to contrast the phraseologies of the target and source languages, using the results to inform language instruction. For instance, Granger (2018) combines contrastive analysis (comparing different languages) with a translation studies perspective and learner corpus data. The resulting 'Contrastive Translation Analysis' approach allows for comparing original language to translated, as well as learner language to native, and by extension "to tease out developmental vs. L1-specific features of interlanguage" (Granger, 2018: 4). This suggests that a contrastive corpus analysis can produce valuable insights into how a speaker's knowledge of their L1 can be reflected in their L2 production. Granger also points out the value of phraseology for examining the influence of one language on another, including L1 transfer in learner language (ibid.). She concludes that frequent phraseological combinations, which can be efficiently unveiled through n-gram extraction, are of great relevance to L2 learners (ibid.: 5), in line with studies of phraseological competence (Paquot, 2018 among others).

Aiming to contribute to the contrastive description of phraseology, the present study compares the use of patterns containing the equivalent preposition pair *in – v* between English and Czech. The results should ultimately inform a study resource developing the phraseological competence in advanced Czech students of English, primarily at university level. Further, the phraseological contrastive analysis of this language pair is potentially valuable from the typological perspective. Previous cross-linguistic phraseological studies indicate that the n-gram method can efficiently identify recurrent sequences and point out cross-linguistic differences in their use. However, n-grams pose methodological difficulties when dealing with typologically distant language pairs, such as English and Spanish, French, or Czech (Čermáková and Chlumská, 2016, 2017; Cortes, 2008; Granger, 2014; Šebestová and Malá,

2019). In the case of Czech, the challenges are due to the highly inflectional nature of Czech, as opposed to predominantly analytical English. A further obstacle is posed by the greater variability of Czech word-order compared to English. Both these factors influence the delimitation of a recurrent multi-word unit in Czech, and have motivated the development of new software capable of identifying patterns with partial lemmatisation and positional mobility (cf. Section 3).

2.1 Prepositional patterns

As pointed out by Hunston (2008), focus on phraseological patterns containing grammatical words ('small words', ibid.) can be beneficial because such patterns contribute to shaping the structure of texts. Fulfilling important textual functions, on a larger scale these grammatical patterns also help reveal pervasive discourse patterning. Discourse-organizing functions are frequently fulfilled by phraseological combinations (Granger, 2018:6), which further indicates that the n-gram method is a suitable means to this end. Moreover, discourse organizing and text structuring is a crucial skill for advanced learners (Granger, 2018), especially for university students, required to produce complex written assignments. Hence, a 'small words' approach seems suitable for this study. Another argument in favour of using grammatical words as the starting point is their extensive frequency and dispersion throughout discourse (Groom, 2010; Sinclair, 1991), making them an efficient tool to provide a comprehensive portrait of the phraseological characteristics of a corpus, to identify a variety of pattern types fulfilling different textual functions and manifesting varying degrees of formulaicity (Groom, 2010:71). For these reasons, function words seem a valid starting point for this study.

Specifically, prepositions were selected as the basis for the identification of phraseological patterns. Prepositions are a valuable starting point from the contrastive and pedagogical perspective since they are a frequent source of errors in EFL students, including advanced learners; apart from their polysemy and polyfunctionality, this is possibly due to a large degree of translation non-correspondences, and inaccurate/oversimplified representation in translation dictionaries (Klégr and Malá, 2009; Peřestá, 2017). In this pilot study, I focus on the preposition pair *in – v*, ranking among the most frequent prepositions in both languages. To summarize, this study aims to identify prepositional patterns involving the translation equivalent preposition pair *in – v* in representative corpora of English and Czech, respectively. These patterns will be described in terms of their textual functions and compared across the two languages.

Although *in* and *v* are translation equivalents, their senses and contexts of use do not entirely correspond across languages (Klégr and Malá, 2009; Peřestá, 2017). The polysemic nature of prepositions seems an important factor, as different senses of a preposition will often be translated by different equivalents (Klégr and Malá, 2009). Consequently, both *in* and *v* are likely to fulfil a range of textual functions. However, the functions carried out by each preposition are expected to differ between the two languages. My aim is to inquire into the nature and extent of these cross-linguistic differences.

2.2 Corpus methods in language teaching

Interest in corpus-informed teaching materials has been growing and influencing approaches to foreign language instruction (Huang, 2011; Reppen, 2011). Corpus material can help learners become acquainted with authentic language, presenting them with a variety of contexts of use (Reppen, 2011:35). Reppen outlines three techniques of employing corpora in language instruction: learning aids prepared by the instructor based on corpus data; interactive practice with students using corpora in class; and using (available or custom-made) specialized corpora

(2011: 36), enabling learners "to explore the patterns found in the writing of their discipline" (2011: 44). Likewise, Hyland (2008: 5) points out the importance of advanced learners knowing discipline-specific phraseological expressions, since "their very 'naturalness' [signals] competent participation in a given community". His analysis shows that scientific disciplines are distinguished by their use of patterns. These patterns are not only content-oriented (or referential lexical bundles, to use Biber *et al.*'s (2004) term); disciplines may use different functional types of lexical bundles, e.g. stance bundles used as hedges are often found in social sciences, while hard sciences employ more reader-oriented bundles (Hyland, 2008). Mastering such bundles is therefore crucial to ESP or EAP learners.

In a related vein, Vašků et al. (2019) compared phraseological of-sequences in English essays by Czech novice academics, with professional academic writing. Differences in pattern use were most prominent in prepositional patterns, where novice writers overused semantically transparent patterns. Similarly, Rankin and Schiftner (2011) investigated the use of English complex prepositions by German learners. In native English, some complex prepositions have specific collocational and contextual preferences, of which the learners seemed unaware.

To conclude, corpus-informed teaching materials are potentially valuable as they contribute to learners' phraseological competence and their mastery of recurrent phraseological sequences, including discipline-specific ones. Even advanced learners tend to have a limited knowledge of phraseological sequences. Since phraseological tendencies (cf. Sinclair, 1991) pervade all levels of language, learners' insufficient phraseological competence pertains also to function word patterns such as prepositional ones. This evidence makes a case for the relevance of corpus-informed teaching materials dedicated to the phraseology of function words.

3. Material and method

The data employed in this study were drawn from corpora roughly comparable in terms of design and size: representative national corpora of English (British National Corpus, 2007) and Czech (SYN2015, Křen *et al.*, 2015, 2016), each around 100 million words. Both contain a variety of written texts; they do not entirely match as regards the time of publication. The BNC, compiled in the early 1990s, contains texts from the late 20th century (Burnard, 2009), mostly between the 1960s-1990s. The SYN2015 covers fiction and non-fiction published between 1990—2015, and journalism from 2010—2015, most texts falling under the span 2010—2014 (Cvrček and Richterová, 2020).

The composition of the English and Czech corpus roughly corresponds: the BNC represents British English and comprises 90% of written texts (fiction, journalism, academic texts, letters, essays etc.); the remaining 10% is spoken informal conversation (Burnard, 2009). By contrast, SYN2015 is written only; it contains a variety of printed and published fiction, non-fiction and journalism (Cvrček and Richterová, 2020). While aware of the two corpora not being a perfect match, their comparable size and overall nature (general representative national corpora) was the criterion for their choice.

As an initial step, a list of the ten most frequent prepositions was compiled for either language. Top ten frequent prepositions were identified manually within the frequency lists available for each corpus (Křen *et al.*, 2016 for SYN2015; and Kilgarriff, n.d. for the BNC).[2]

[2] Kilgarriff: BNC database and word frequency lists. Available from http://www.kilgarriff.co.uk/bnc-readme.html Czech National Corpus: Reference frequency lists (Srovnávací frekvenční seznamy). Available from <https://wiki.korpus.cz/doku.php/seznamy:srovnavaci_seznamy> (in Czech)

Only unambiguous prepositions were selected in Czech.[3] The choice of the English prepositions warrants a comment. Kilgarriff's wordlists were used since they are based on the entire BNC, and thus informative as to the prepositions' frequencies relative to the whole collection, showing that prepositions rank among the most frequent words in the corpus. However, the lists do not include normalised frequency information. Moreover they are based on the BNC World Edition (2001), which is no longer available, hence the frequencies differ slightly from the currently accessible XML version. On the other hand, searching for the frequencies of all prepositions in BNC XML Edition (2007), the frequency breakdown is limited to a random sample of 250,000 hits. However, the lemmatised top ten prepositions match those based on Kilgarriff, only their ranking is slightly different. Cf. Table 1.

Table 1. Top 10 English prepositions in the BNC World – as per lemmatised wordlists (Kilgarriff n.d.); compared to top 10 of a random retrievable 250,000 hit sample of preposition lemmata (BNC XML Edition, 2007).

Rank	**Prepositions in BNC World**	**Rank in whole wordlist**	**Raw freq in BNC World**	**Preposition in sample**	**Raw freq in prep sample**	**Raw freq – whole BNC XML**	**ipm – whole BNC XML**
1	of	3	3,093,444	of	59,085	3,040,670	30,928
2	in	6	1,924,315	to	50,779	2,593,740	26,382
3	to	10	1,039,323	in	36,341	1,937,966	19,712
4	for	11	887,877	for	16,925	878,741	8,938
5	on	16	680,739	with	12,748	658,584	6,698
6	with	17	675,027	on	12,524	729,558	7,420
7	at	19	534,162	at	10,092	521,697	5,306
8	by	20	517,171	by	9,893	512,215	5,210
9	from	24	434,532	from	8,393	424,972	4,322
10	as	48	201,968	as	4,304	653,610	6,648

To confirm the translation equivalence of *in* and *v*, in line with the corpus-driven (Tognini-Bonelli, 2001) approach adopted in this study, equivalents were extracted from the InterCorp v. 12 parallel corpus (Čermák and Rosen, 2012; Rosen *et al*., 2020) via the Treq application, 2.1 (Vavřín and Rosen, 2015; Škrabal and Vavřín, 2017).[4] This confirms that the prevalent English equivalent of Czech v is indeed in, see Table 2.

Table 2. English translation equivalents in InterCorp 12 as per Treq (Vavřín and Rosen, 2015).

Czech preposition	**prevalent English equivalent (*Treq*)**	**rank in SYN2015 lemmatised wordlist**	**raw freq in SYN2015**	**ipm in SYN2015**
v	in	4	2,296,562	19,075

[3] The preposition *se* (homonymous with a reflexive pronoun) was excluded. In fact, *se* ranks third in the SYN2015 wordlist (raw frequency = 3,070,434). However, a search in SYN2015 (Křen, *et al*. 2015) reveals that merely 155,508 of those instances are prepositional, the vast majority (2,306,916 hits) being the reflexive pronominal uses.

[4] The direction of translation was Czech to English, the query was lemmatised and case-insensitive. The search was performed within the entire corpus, i.e. not limited to any specific subcorpora.

As mentioned earlier, the preposition pair *in – v* was chosen due to their frequency: both *in* and *v* rank among the most frequent prepositions, as well as the most frequent words in the corpus overall (cf. Tables 1 and 2).

3.1 N-gram method – state of the art

N-gram methodology has proven a useful starting point for cross-linguistic studies working with related languages. When contrasting typologically distant language pairs such as English and Spanish, French, or Czech (Čermáková and Chlumská, 2016, 2017; Cortes, 2008; Granger, 2014; Šebestová and Malá, 2019) the methodology poses problems.

For instance, Granger (2014) compared lexical bundles in English and French across two genres (parliamentary debates and newspaper editorials), focusing on stems, i.e. combinations of subject and verb with optional pre-verbal elements (Altenberg, 1998). French was expected to employ more bundles overall. This tendency was apparent in editorials, but inconclusive in debates (ibid.: 64), indicating that phraseological tendencies may differ markedly across languages as well as registers.

Hasselgård (2017) on the other hand compared English and Norwegian 2-4-grams expressing temporal meanings. This study illustrates how n-gram methodology highlights typological differences which would be difficult to identify otherwise. The Norwegian data contained fewer recurrent n-grams overall, indicating English may have a stronger tendency towards recurrence. Yet in Norwegian, temporal n-grams formed a larger part of all the n-grams identified. Also, Norwegian n-grams corresponded to (fragments of) clauses more often (ibid.: 86). Hence, while some languages display more recurrence than others (i.e. typological properties are an important factor shaping phraseology), a language may employ phraseological means of expression to varying degrees in different semantic or functional areas, pointing towards a register-dependent distribution. Hasselgård´s study also hints towards n-gram methodology being potentially challenging even when applied to typologically related languages.

N-grams applied to the English-Czech language pair pose methodological challenges due to the typological non-correspondences. In Čermáková and Chlumská's (2016) n-gram analysis of Czech and English children's literature, English datasets yielded hundreds of n-grams, whilst in the Czech data of comparable size, only tens of n-grams were identified. This suggests that the results for each language are best examined separately as cross-linguistic comparability may be limited. In summary, previous cross-linguistic n-gram-based research indicates that typological properties and the register factor enter into a complex interplay. Further, depending on corpus design, the validity of the results is likely limited to the particular registers explored. These findings were used to inform the choice of data for the present study, namely large representative corpora, to ensure a variety of registers were represented.

In the following analysis, I use *n-gram* to refer to recurrent sequences of n words identified mechanically in corpus data, which may or may not correspond to structural units such as phrases; sometimes an n-gram comprises a complete phrase along with fragments of adjacent phrases or other structures (e.g. *of fall in love* or *fall in love with*, or *fall in love and*, where the conjunction implies a following clause; cf. Figure 1 in Section 3.2).

3.2 Engrammer software description

The data in this study was processed using the custom-made Engrammer freeware (Milička, 2019).[5] Engrammer enables searches for sequences of words of different lengths at once,

[5] *Engrammer*, available from <http://www.milicka.cz/en/engrammer/>

collapsing overlapping n-grams, e.g. *in order + in order to = in order to*. The frequencies of the individual overlapping variants can still be displayed. Figure 1 shows the Engrammer interface. The n-gram search results are in the left column. Clicking the n-gram, all variants subsumed under it are displayed in the right-hand column, together with their collocation strength and frequency. Optionally, collapsing is also available for similar n-grams ('similar' defined as differing in one position only). In Figure 1, lemmatised n-grams *fall in love with, have fall in love, to fall in love, I fall in love* etc. were collapsed. Henceforth I will be referring to the collapsed n-grams as *n-gram types* (e.g. *bear in mind, in spite of, fall in love* in Figure 1 are three different n-gram types); and individual n-gram occurrences as *n-gram tokens*.

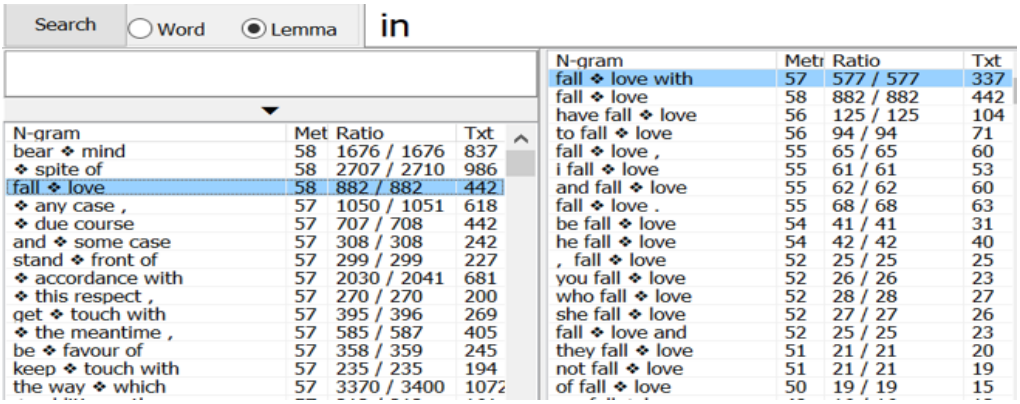

Figure 1. Engrammer interface displaying n-grams containing *in*.

3.3 N-gram search

Full text lemmatised versions of the corpora were plugged into Engrammer, one at a time. For each corpus, lemmatised 3- to 5-grams were extracted (all lengths at once), containing the preposition *in/v* in any slot (cf. Table 3). Variable word order was allowed within n-grams because Czech word order is highly flexible (cf. Čermáková and Chlumská, 2016; 2017). Given that grammatical word patterns contribute to linking, they can be expected to occur near syntactic boundaries: hence punctuation was included. The search was set so that similar n-grams (differing in one lemma only) were collapsed (cf. 3.1). The search retrieved a total of 398 n-gram types, 55,790 tokens for English; 431 n-gram types and 21,660 n-gram tokens for Czech.

Next, I analysed the collapsed n-grams manually, searching for "meaningful, linguistically structured" (Lindquist and Levin, 2008: 144) units within them, which I term *patterns*. For practical reasons, the dataset for each language was limited to the top frequent 250 (collapsed) n-gram types. Table 3 illustrates the process of identifying a pattern within lemmatised n-grams: the pattern *in front of* was abstracted from the individual n-gram types.

Table 3. Breakdown of the collapsed pattern *in front of* (span: 3-5-grams).

N-gram (lemmatised)	**N-gram token freq.**
in front of i ,	75
right in front of	89
in front of he ,	162
just in front of	70
in front of the television	69
in front of they ,	57
Total n-gram tokens	**522**
Total n-gram types	**7**

The resulting sequences were ordered by the 'risk of n-gram' rubric, using the risk ratio metric. Generally, risk ratio is based on comparing the probability of a particular item occurring in a context A as opposed to occurring in another context B (Březina, 2018: 115–16). The 'risk of n-gram' measures the strength of association between the node word (*in/v*) and each n-gram. The frequency of *in/v* in a given n-gram is compared to the frequency of *in/v* alone, and the corpus size is taken into account. E.g. *in* alone occurs 2,593,740 times in the BNC XML edition (cf. Table 1); the sequence *in front of the television* (cf. Table 3) occurs 69 times, and the corpus length is 96,986,707 tokens. This results in a risk of n-gram value of 2.1 (confidence interval = 1.8—2.2), i.e. *in front of the television* occurs at least 1.8 times more often than can be expected by chance.

While a comparable number of n-grams was extracted from both languages, English n-grams exhibited higher 'risk of n-gram' values overall than Czech (cf. Table 4), suggesting a greater degree of fixedness in English. However, this tendency may be enhanced by the analytical nature of English.

Table 4. Cross-linguistic differences in node-n-gram association strength.

English *in*		**Czech *v***	
Risk of n-gram	**No. of n-gram types**	**Risk of n-gram**	**No. of n-gram types**
57	23	52	7
56	68	51	86
55	133	50	138
54	171	49	200
Total	**395**	**Total**	**431**

3.4 Classification of *in* and *v* patterns

The prepositional patterns were sorted into functional-semantic groups in an inductive, bottom-up manner. This approach was adopted with regard to potential pedagogical applications: the most frequent patterns containing a given word can serve as the starting point for identifying the common contexts of usage of any selected word.

Where applicable, patterns were grouped based on a semantic perspective. The criterion was the meaning conveyed by lexical words in the pattern. This resulted in 6 groups of patterns, 5 of these conveying adverbial meanings. Apart from these, the *body/mind* group was singled out, since patterns referring to body parts (e.g. *go hand in hand with*) or the mind (*bear in mind*) were frequent in both corpora.

Since not all patterns lend themselves to semantic classification, the semantic perspective was complemented with a formal-structural one wherever no overarching semantic feature was identified, but multiple patterns shared a grammatical structure or part of speech: e.g. complex preposition patterns (*in front of, v rámci* 'in the framework of'[6]), or patterns comprising a 'copula + complement' (*be in charge, být v pořádku* 'be in order').

Finally, two groups of patterns stood out: patterns conveying emphasis (*in the first place, v prvé řadě* 'in the first place') and hedging patterns (*in a sense, v jistém smyslu* 'in a sense'). Both were subsumed under the broadly conceived 'pragmatic' patterns, defined by fulfilling a discourse function, rather than by semantics or formal characteristics.

Some patterns could be classified by more than one of the three types of criteria (semantic, formal-structural, pragmatic): e.g. *v žádném případě* 'in no case/by no means' or *v mnoha ohledech* 'in many respects' could be classified semantically as adverbial patterns of regard, or pragmatically as emphasizers. The semantic criterion was prioritised and the patterns were classified as adverbial, since the adverbial group was considered broader and able to

[6] Henceforth, all verbatim translations from Czech into English, given in single quotation marks, are mine.

encompass the pragmatically specialized usages. Similarly, wherever a pattern conveyed adverbial meaning but also contained a distinctive structural element, e.g. a complex preposition or phrasal verb (e.g. *ve srovnání s rokem X* 'in comparison with the year X'), it was classified under the corresponding structural pattern type rather than the semantic adverbial type, in line with the focus on phraseological patterning centred around function words.

3.5 Idiomaticity as an additional criterion

Independently of the classification based on semantic/formal/functional criteria, I annotated the patterns for idiomaticity, defined broadly as being lexically (at least partly) fixed: either a given word cannot be replaced with its (near) synonym: e.g. *be in short supply* not **be in brief/abbreviated supply*; or the choice of acceptable synonyms is limited: *be not in a position;* possibly also *be not in a place*[7], but not **be not in a location*.[8]

Idiomatic patterns occurred across the pattern groups and will be discussed in Section 4.6. The decision to add this perspective was prompted by the occurrence of potentially metaphorical patterns among the *body/mind* pattern group, e.g. *hand in hand* (cf. 6.2). Next, idiomatic patterns were assessed in terms of semantic transparency/opacity. Patterns were considered opaque if the whole pattern conveyed a meaning which was not a sum of the meanings of its parts (e.g. *in the light of these*), their meaning was perceived as figurative rather than literal (*keep in touch with*), or they contained a limited-collocability item (*in the nick of*). As shown by Table 5, the proportion of transparent and opaque patterns is even in both languages; idiomatic patterns were more frequent in English overall.[9]

Table 5. Idiomatic patterns in English and Czech.

Fixed patterns	**English**	**Czech**
Opaque	26	16
Transparent	22	15
Total – idiomatic patterns	**48**	**31**
Total – all patterns	**250**	**250**

A variety of meanings and functions is conveyed by *in* and *v* patterns. Table 5 outlines the pattern groups identified, ordered by frequency for each language corpus, listing pattern type frequencies for each group.[10] Most pattern groups were identified in both English and Czech. Pattern groups identified in one language only are addressed in Section 5.

Table 6 lists the pattern groups according to their respective defining criteria: structural, semantic or pragmatic. Section 4 goes on to discuss the attested pattern groups.

[7] One example was attested in the BNC: *Thee ain't in no place to talk about prying*; possibly informed by analogy with *it is not my place to*.

[8] This can be viewed as a manifestation of Sinclair's (1991:110) principle of idiom, i.e. "a large number of semi-preconstructed phrases that constitute single choices", or as Altenberg (1998: 115) puts it "more or less prefabricated or routinized building blocks".

[9] Admittedly it proved difficult to establish robust criteria for determining semantic opacity. A possible solution would be having the patterns evaluated by native speakers, followed by an inter-rater agreement analysis.

[10] E.g. complex preposition patterns comprised 44 pattern types, one of them being *in front of* (described in Table 3). Higher pattern type frequency indicates a greater formal variety within the particular pattern group. Contrarily, a low pattern type frequency points towards a greater degree of formal repetitiveness within that group.

Table 6. Pattern groups in English and Czech.

Pattern group	English		Czech	
	Example of pattern type	**Pattern type freq**	**Example of pattern type**	**Pattern type freq**
Structural				
Complex prep.	in front of	44	v rámci NP 'within the framework of NP'	28
Complex conj.	in order to	21	N/A	0
Copular/phasal verb	be in charge	20	být v pořádku 'be in order / all right' pokračovat v chůzi 'continue walking'	20
Phrasal/prep. verb	come in handy	10	spočívat v tom, že 'lie/consist in the fact that'	28
Valency	interested in	7	N/A	0
Semantic				
ADV place	in chapter…, in appendix…	33	pobyt v nemocnici 'a stay in hospital'	46
ADV regard	and in some case	23	v tomto ohledu 'in this respect'	19
ADV manner	way in which	22	ve zkratce 'in short'	1
ADV time	in the morning	18	aktivní v noci 'active at night	44
ADV circumstances/ state	in silence, in doubt	9	přednost v jízdě 'right of way' být v klidu 'be calm'	23
Body/mind	in a ADJ voice	4	sucho v ústech 'dryness in the mouth'	19
Pragmatic				
Emphasis	in the first place, in any case	35	v prvé/neposlední řadě 'in the first place'/'last but not least'	17
Hedge / approximation	in a sense	4	v jistém smyslu být 'in a sense be'	4
Other	N/A	0	minulý měsíc ubývat v 'last month decrease in'	1
TOTAL		**250**		**250**

4. Discussion of pattern uses

4.1 Semantically defined patterns: Adverbial patterns

This group includes patterns expressing adverbial meanings, as illustrated by examples (1) through (5).

(1) Place: *in* court / sedět *v* kuchyni 'be sitting in the kitchen'

(2) Time: *in* the morning / aktivní *v* noci ‘active at night’

(3) Manner: *in* short / *ve* zkratce ‘in short’

(4) Regard: *in* this respect / *v* tomto ohledu ‘in this respect’

(5) State: if *in* doubt / být *v* klidu ‘be calm’

Some state adverbial patterns could form part of copular constructions; yet the n-grams retrieved did not contain the copula, e.g. (*být*) *jako v bavlnce* ‘(to be) comfortable’.

4.2 Semantically defined patterns: Body/mind patterns

Patterns containing a noun referring to body parts or the mind, see example (6), were singled out; idiomaticity was taken into account as a result, since these expressions are frequent source domains for metaphors (Lindquist and Levin, 2008).

(6) hand in hand / říci si v duchu ‘say to oneself’

4.3 Structurally defined patterns: Verbal patterns

In verbal patterns, copular (example 7), phrasal and prepositional (8) verbs occurred. This was not surprising since all these verbs form part of phraseological sequences: copular verbs require complementation, while phrasal/prepositional verbs constitute multi-word units by definition.

(7) be in charge / být v pořádku ‘be in order/all right’

(8) come in handy/ spočívat v tom, že ‘consist in the fact that’

One verbal pattern group was limited to English: verbs with a valency complement, e.g. *interested in*. These are discussed in Section 5.

4.4 Structural: Complex prepositions and conjunctions

Another group of patterns was formed by complex prepositions (9) and conjunctions (10), the latter only attested in English.

(9) in front of /v rámci NP ‘within NP’

(10) in order to

4.5 Pragmatic patterns

Lastly, patterns with pragmatic functions were identified: emphasis (example 11) and hedging/approximation (12).

(11) Emphasis: in the first place / v neposlední řadě

(12) Hedge: in a sense / v jistém smyslu

While some functional-semantic pattern groups comprise a diverse set of expressions (e.g. place adverbials), the pragmatic group seemed limited to few recurrent patterns. This suggests that the pragmatic functions may favour more conventionalised forms of realization.

4.6 Idiomatic patterns

Examples of idiomatic patterns were found across a range of semantically/structurally/pragmatically defined pattern groups, as shown in Tables 7 and 8 in decreasing order of frequency for each language.

Table 7. English idiomatic patterns: distribution across pattern groups.

Group	Opaque	Transparent	Total types
Copular/phrasal	12	3	15
Emphasis	3	5	8
Complex prep.	4	3	7
Phrasal/prep verb	1	4	5
Time	3	2	5
Body/mind	2	2	4
Circumstances/state	0	3	3
Valency	1	0	1
Total types	**26**	**22**	**48**

In English, most idiomatic patterns occurred in the verbal type comprising a copular or phrasal verb (13), followed by patterns, serving to emphasize, structure and punctuate discourse (14); and complex prepositions, likewise means of text structuring (15).

(13) fall in love; get in touch with

(14) in any case; in the first place

(15) in spite of; in the wake of the

Table 8. Czech idiomatic patterns: distribution across pattern groups.

Group	Opaque	Transparent	Total types
Body/mind	3	8	11
Copular/phrasal	4	4	8
Circumstances/state	4	1	5
Place	3	0	3
Phrasal/prep. verb	2	1	3
Manner	0	1	1
Total types	**16**	**15**	**31**

Among Czech idiomatic patterns, especially those referring to body and mind were prominent (16), followed by copular verbal patterns (17).

(16) jít ruku v ruce ‘go hand in hand’

(17) být v sedmém nebi ‘be in seventh heaven’

Notably, some adverbial circumstances/state patterns potentially overlap with verbal ones: the pattern in ex. (18) would often occur with the copula *být* (‘be’). However, the copula was not included in the recurrent pattern since it can alternate with other verbs. Ex. (19) could be alternatively classified under *phrasal/prepositional verb* (cf. 21 below).

(18) (být) jako v bavlnce ‘be very comfortable’

(19) nechat ve štychu ‘leave in the lurch’

5. Cross-linguistic differences

The first major cross-linguistic difference lies in the distribution of pattern groups. Essentially, *in* and *v* patterns convey the same functions in both languages. However, pattern types are distributed differently: ranked by raw frequency, corresponding pattern groups differ in their position within the top-frequent ranking (see Table 9). In other words, Czech does not employ individual pattern types with the same frequency as English. To illustrate this, Table 9 lists the top frequent five pattern groups in both languages. The frequency was assessed by the n-gram token counts within each pattern group, i.e. by the number of all n-grams conveying this function. The patterns in *italics* (place adverbials, complex prepositions) rank among the top five in both languages.

Table 9. Top five functions for each language, ordered by n-gram token frequency.

English	**Token freq.**	**Czech**	**Token freq.**
Complex prep.	11,430	Time	8,373
Emphasis	7,183	Phrasal / prep. verb	5,722
Manner	4,375	*Complex prep.*	5,420
Place	4,239	*Place*	5,376
Copular	3,853	Regard	4,172

Secondly, two pattern groups were identified in English only, highlighting its analytic features: complex conjunctions, and verbs with valency complements. Below I discuss the language-specific features revealed by the pattern analysis for each language.

5.1 English *in* patterns

There were two extra pattern groups attested in English: complex conjunctions and prepositional verbs with valency complements. As regards complex conjunctions, the majority of this group was represented by *in order to* or variations thereof (18 out of 21 n-gram types). Either there is an adjectival head postmodified by an infinitival clause introduced by *in order to* (example 20); or the pattern captures the following verb (example 21). The other 2 conjunction pattern types were *in such a way as/that* and *except in so far*.

(20) necessary in order to

(21) in order to achieve/gain/understand/avoid

Since the English and Czech corpora did not entirely match in terms of the text types represented (cf. Section 2), the question arises whether complex conjunctions may be limited to English due to their distribution in specific text types, perhaps less represented in the Czech corpus. This was checked for the most frequent conjunction *in order to*. As apparent from Table 10, *in order to* is predominantly found in books; a closer inquiry into its distribution across text domains shows that it occurs predominantly in social sciences, followed by world affairs (i.e. newspapers). Interestingly, *in order to* is widely used in social sciences (215 ipm) while much less common in natural sciences (143 ipm). This evokes Hyland's (2008) findings about specialized discourses being marked by the usage of text-structuring patterns.

Table 10. *in order to* - distribution across text types in BNC.

Text type	No. of words	Freq .raw	Freq. ipm
Written books and periodicals	79,187,792	10,243	129.35
Written miscellaneous	7,437,161	1,292	173.72
Context-governed	6,175,896	485	78.53
Demographically sampled	4,233,962	16	3.78
Written-to-be-spoken	1,278,618	14	1.95
Total	**98,313,429**	**12,050**	**122.57**

Furthermore, complex prepositions are more frequent in English (44 n-gram types, 11,430 n-gram tokens) than in Czech (28 n-gram types, 5,420 n-gram tokens). Not only are complex preposition patterns almost twice as frequent in English overall (cf. n-gram-token counts), they are also formally more varied (= more n-gram types). In sum, the findings about complex conjunctions and prepositions indicate there may be more complex function patterns in English overall, in line with English being an analytic language with a rich and recurrent repertoire of function words.

The second type of pattern exclusive to English was a verb followed by its valency complement; a prepositional object (22) or adverbial prepositional phrase (23).

(22) interested in NP

(23) an increase in NP

Although this group was not attested in Czech, there were similar Czech patterns, namely a verb followed by a prepositional phrase, as in (24).

(24) vzít v potaz/úvahu ‘take into account/consideration’

However, in Czech, the noun phrase complementing the preposition is lexically fixed. Czech patterns such as *vzít v* + NP are idiomatic, hence the choice of the noun *potaz/úvahu* (‘take into account’)*;* while in English patterns as in *interested in*, the following slot is open and may contain any one of a range of nominal complements. Due to this collocational fixedness, Czech patterns of the type *vzít v úvahu* were labelled as phrasal/prepositional verb. (Admittedly such Czech constructions are not formally analogous to English phrasal verbs; yet they are characterised by lexical fixedness).

A glimpse at the n-gram type-token ratios of the attested patterns reveals that some pattern groups are formally repetitive; a qualitative look at the data confirms this. English adverbial patterns of manner consist almost exclusively of *the/a way in which* (18 of total 22 n-gram types). A similar tendency was observed in Czech adverbial patterns of regard (variations on *v tomto případě* ‘in this case’, *v tomto ohledu* ‘in this respect’). Complex prepositions are likewise repetitive in both languages, which can be expected given their formal fixedness.

Similarly, body/mind patterns comprise a mere four types – cf. Table 11 (or rather three, given the overlap *go hand in hand with*). Despite its repetitiveness, the body/mind pattern group is very frequent: it would warrant closer investigation to find out more about its common contexts of use.

Table 11. Body/mind pattern group.

N-gram	Collocation strength (risk of n-gram)	Freq.
bear in mind	57.012	1676
go hand in	56.7823	172
hand in hand with	56.3167	114
say in a low voice	55.1806	62

5.2 Czech *v* patterns

Similarly to English, an examination of type-token ratios provides some insights into Czech prepositional patterns. Place adverbial patterns are a diverse group comprising a number of given names, whose referents range from TV series (25) to institutions (26) or even topical events (27).

(25) Ordinace v růžové zahradě 'Surgery in the Rose Garden' [11]
Sex ve městě 'Sex and the City'

(26) fakulta UK v Praze 'faculty of Charles University in Prague'
krajský soud v Brně 'regional court in Brno'

(27) olympiáda v Soči 'Olympics in Sochi'

Other adverbial place patterns are register-specific, as in (28), typical of the language of advertising.

(28) info o ceně v obchodě – 'price information available in the shop'

Further, place patterns refer to a variety of locations (29). Indeed, *v* is one of the most common prepositions to combine with the locative (Cvrček *et al.*, 2015: 172).[12]

(29) v nemocnici / v kuchyni / ve vězení – 'in hospital/the kitchen/prison'

Lastly, idiomatic place patterns were found (30).

(30) viset ve vzduchu 'hang in the air'; praskat ve švech 'burst at the seams'

On the other end of the diversity cline are pragmatic patterns expressing emphasis, mostly variations on *v žádném/každém případě* 'by no/all means'. This may reflect the tendency of pragmaticalised patterns to become fixed with repeated usage. By contrast, adverbial place patterns may refer to a host of referents, reflecting speakers´diverse communicative needs.

Finally, more idiomatic patterns were attested in English than in Czech overall. A qualitative assessment of the idiomatic patterns seems to suggest that there is in fact a cline of semantic opacity, as illustrated by (31–33) (note: 31 and 32 are equivalents which occured in both languages).

(31) fully opaque, non-compositional: be in full swing - být v plném proudu

(32) abstract uses (e.g. personifications): go hand in hand – jít ruku v ruce s

(33) fully transparent: put in an appearance – být jako v transu 'be as if in a trance'

[11] A popular Czech soap opera.
[12] My thanks go to the anonymous reviewer for pointing this out.

6. Conclusion

This pilot study has examined prepositional patterns in English and Czech, classifying them into inductively defined groups based on semantic, structural or pragmatic criteria. Major pattern types represented in both languages included adverbial patterns, verbal patterns, complex prepositions, and conjunctions. Pragmatic patterns served as a means of emphasis or hedging. While the pattern groups generally corresponded between the two languages, they are distributed differently: e.g. complex prepositions occurred nearly twice as often in English than in Czech. The distribution may be influenced by text type or register – more research into this is needed. A potential application of this finding would present itself in the use of custom-made corpora of specialized texts in the classroom, enabling students to identify patterns and compare their uses in their L1 and L2, or to observe whether translation equivalent patterns are used in similar contexts or registers.

To some extent, patterns reflected the typological properties of the languages. Analytical English employs more complex prepositions and conjunctions, both in terms of n-gram type and token counts. As earlier reseach has indicated that even advanced EFL learners may tend to use fewer patterns and prefer less lexically sophisticated ones (Vašků *et al*., 2019), this is further evidence that the use of such complex text-structuring patterns deserves attention in class.

Finally, the pattern types display a varying degree of repetitiveness. This may be caused by some meanings being more closely associated with particular expressions (*v žádném případě* – 'under no circumstances'). Alternatively, it may simply reflect the high frequency of some patterns in the corpus (*in order to*). These hypotheses prompted by the pilot study findings provide an interesting impetus for further research; the reasons for the differences in individual patterns' frequencies could be investigated through a qualitative analysis of a larger dataset. At any rate, the observations regarding pattern idiomaticity suggest that this parameter warrants special attention in language instruction. Under an inductive teaching approach, similar observations about specific patterns and their usage can be made efficiently by students exploring corpus data.

To complement this study, patterns around other frequent prepositions should be compared to *in* and *v*. Lastly, bearing in mind that phraseological patterns can identify register-specific features (Biber *et al*., 2004), another follow-up possibility is a comparison of the prepositional patterns identified in large representative general corpora such as the BNC and SYN2015, to patterns found in specialized corpora of particular registers – building on research on register variation in Czech (Cvrček *et al*., 2020).

The results of the study have illustrated the potential value of viewing phraseological sequences through a cross-linguistic lens: contrasting prepositional patterns in two corpora of different languages reveals similarities in the pattern types employed by the languages, while also highlighting differences in the distribution, overall frequency, functional load and diversity of pattern types. Given the importance of phraseological competence for L2 proficiency (Paquot, 2018), contrastive phraseological analyses can provide advanced learners with valuable insight into their target language phraseologies. Further, patterns may illustrate the typological features of languages, as reflected in the greater frequency of complex prepositions and conjunctions in analytical English – such observations may help L2 learners better grasp the theoretical notion of language typology as well as to notice structural differences between languages.

Acknowledgements

My sincere thanks go to Jiří Milička for methodological assistance, as well as to the reviewers for their thorough reading and helpful critical comments.

References

Altenberg, B. 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In *Phraseology: Theory, Analysis and Applications*, A.P. Cowie (ed.), 101–22. Oxford: OUP.

Biber, D, Conrad, S. and Cortes, V. 2004. 'If You Look at…': Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25(3):371–405. doi: 10.1093/applin/25.3.371.

Březina, V. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: CUP.

British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Burnard, Lou. 2009. What Is the BNC? Oxford Text Archive, IT Services, University of Oxford. Available at http://www.natcorp.ox.ac.uk/corpus/index.xml [Last accessed 21 April 2021].

Čermák, F. and Rosen, A. 2012. The Case of InterCorp, a Multilingual Parallel Corpus. *International Journal of Corpus Linguistics* 13(3):411–27.

Čermáková, A. and Chlumská, L. 2016. Jazyk dětské literatury: kontrastivní srovnání angličtiny a češtiny. In *Jazykové paralely*, A. Čermáková, L. Chlumská and M. Malá (eds), 162–187. Prague: NLN.

Čermáková, A. and Chlumská, L. 2017. Expressing Place in Children's Literature. Testing the Limits of the N-Gram Method in Contrastive Linguistics. In *Cross-linguistic Correspondences: From Lexis to Genre*, *Studies in Language Companion Series*, T. Egan and H. Dirdal (eds), 75–95. Amsterdam: John Benjamins.

Cortes, V. 2008. A Comparative Analysis of Lexical Bundles in Academic History Writing in English and Spanish. *Corpora* 3(1):43–57.

Cvrček, V. *et al*. 2015. *Mluvnice současné češtiny 1: Jak se píše a jak se mluví*. Charles University, Karolinum.

Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A. and Zasina, A.J. 2020. *Registry v češtině*. Praha: NLN.

Cvrček, V. and Richterová, O. (eds) 2020. *En:Cnk:Syn2015*. Available at http://wiki.korpus.cz/doku.php?id=en:cnk:syn2015&rev=1598975168 [Last accessed 15 August 2021].

Czech National Corpus: Reference frequency lists (Srovnávací frekvenční seznamy). 2016. Institute of the Czech National Corpus. Available from: http://www.korpus.cz

Ebeling, S.O. and Hasselgård, H. 2015. Learner Corpora and Phraseology. In *The Cambridge Handbook of Learner Corpus Research*, S. Granger, G. Gilquin and F. Meunier (eds), 207–230. Cambridge: CUP.

Erman, B. and Warren, B. 2000. The Idiom Principle and the Open Choice Principle. *Text* 20(1):29–62.

Granger, S. 2014. A Lexical Bundle Approach to Comparing Languages: Stems in English and French. Special issue of *Languages in Contrast* 14(1), M.-A. Lefer and S. Vogeleer (eds), 58–72. doi: 10.1075/lic.14.1.04gra.

Granger, S. 2017. Academic Phraseology. A Key Ingredient in Successful L2 Academic Literacy. *Oslo Studies in Language* 9(3), *Academic Language in a Nordic Setting – Linguistic and Educational Perspectives*, R.V. Fjeld, K. Hagen, B. Henriksen, S. Johansson, S. Olsen and J. Prentice (eds), 9–27.

Granger, S. 2018. Tracking the Third Code. In *The Corpus Linguistic Discourse: In Honour of Wolfgang Teubert*, *Studies in Corpus Linguistics*, A. Čermáková and M. Mahlberg (eds), 185–204. Amsterdam: Benjamins.

Granger, S. and Bestgen, Y. 2014. The Use of Collocations by Intermediate vs. Advanced Nonnative Writers: A Bigram-Based Study. *International Review of Applied Linguistics in Language Teaching (IRAL)* 52(3):229–52.

Granger, S. and Meunier, F. (eds). 2008. *Phraseology. An Interdisciplinary Perspective*. Vol. 139. Amsterdam: Benjamins.

Granger, S. and Paquot, M. 2008. Disentangling the Phraseological Web. In *Phraseology:An Interdisciplinary Perspective*, S. Granger and F. Meunier (eds), 27–49. Amsterdam; Philadelphia: Benjamins.

Gray, B. and Biber, D. 2015. Phraseology. In *The Cambridge Handbook of English Corpus Linguistics*, D. Biber and R. Reppen (eds), 125–145. Cambridge: CUP.

Groom, N. 2010. Closed-Class Keywords and Corpus-Driven Discourse Analysis. In *Keyness in Texts*, M. Bondi and M. Scott (eds), 59–78. Amsterdam: Benjamins.

Hasselgård, H. 2017. Temporal Expressions in English and Norwegian. In *Contrasting English and other Languages through Corpora*, M. Janebová, E. Lapshinova-Koltunski and M. Martinková (eds), 75–101. Newcastle: Cambridge Scholars Publishing.

Hasselgård, H. 2019. Phraseological Teddy Bears: Frequent Lexical Bundles in Academic Writing by Norwegian Learners and Native Speakers of English. In *Corpus Linguistics, Context and Culture*, V. Wiegand and M. Mahlberg (eds), 339–362. Berlin, Boston: De Gruyter.

Hasselgren, A. 1994. Lexical Teddy Bears and Advanced Learners: A Study into the Ways Norwegian Students Cope with English Vocabulary. *International Journal of Applied Linguistics* 4(2):237–59. doi: https://doi.org/10.1111/j.1473-4192.1994.tb00065.x.

Howarth, P. 1998. Phraseology and Second Language Proficiency. *Applied Linguistics* 19(1):24–44. doi: https://doi.org/10.1093/applin/19.1.24.

Huang, L.-F. 2011. *Discourse Markers in Spoken English: A Corpus Study of Native Speakers and Chinese Non-Native Speakers*. Doctoral thesis, University of Birmingham, Birmingham, UK.

Hunston, S. 2008. Starting with the Small Words. Special issue of *International Journal of Corpus Linguistics* 13(3): *Patterns, Meaningful Units and Specialized Discourses*, U. Römer and R. Schulze, 71–95.

Hyland, K. 2008. 'As Can Be Seen': Lexical Bundles and Disciplinary Variation'. *English for Specific Purposes* 27:4–21.

Kilgarriff, A. n.d. BNC Database and Word Frequency Lists. Available at https://www.kilgarriff.co.uk/bnc-readme.html [Last accessed 25 June 2021].

Klégr, A. and Malá, M. 2009. English Equivalents of the Most Frequent Czech Prepositions: A Contrastive Corpus-Based Study. In *Proceedings of the Corpus Linguistics Conference, CL 2009, Conference in Liverpool, 20-23 July 2009*, M. Mahlberg, V. González-Díaz and C. Smith (eds). Available at http://ucrel.lancs.ac.uk/publications/cl2009/ [Last accessed 26 June 2021].

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kováříková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondřička, P. and Zasina, A. 2015. *SYN2015: reprezentativní korpus psané češtiny*.

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kováříková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondřička, P. and Zasina, A. 2016. SYN2015: Representative Corpus of Contemporary Written Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2522–2528. Portorož: ELRA.

Lindquist, H, and Levin, M. 2008. Foot and Mouth: The Phrasal Patterns of Two Frequent Nouns. In *Phraseology. An Interdisciplinary Perspective*, S. Granger and F. Meunier (eds), 143–158. Amsterdam: Benjamins.

Milička, J. 2019. *Engrammer*. Praha: Institute of the Czech National Corpus, Faculty of Arts, Charles University.

Paquot, M. 2018. Phraseological Competence: A Missing Component in University Entrance Language Tests? Insights From a Study of EFL Learners' Use of Statistical Collocations. *Language Assessment Quarterly* 15(1):29–43. doi: https://doi.org/10.1080/15434303.2017.1405421.

Paquot, M. and Granger, S. 2012. Formulaic Language in Learner Corpora. *Annual Review of Applied Linguistics* 32:130–49. doi: 10.1017/S0267190512000098.

Peřestá, G. 2017. *Akviziční interference angličtiny a češtiny v prepozicionálních konstrukcích / Acquisitional Interference of English and Czech in Prepositional Constructions*. Unpublished BA thesis, Charles University, Faculty of Arts, Prague.

Rankin, T. and Schiftner, B. 2011. Marginal Prepositions in Learner English: Applying Local Corpus Data. Special issue of *International Journal of Corpus Linguistics* 16(3), *Applying Corpus Linguistics*, F. Farr and A. O'Keeffe (eds), 412–434.

Reppen, R. 2011. Using Corpora in the Language Classroom. In *Materials Development in Language Teaching*, B. Tomlinson (ed.), 35–50. Cambridge: CUP.

Rosen, A., Vavřín, M. and A.J. Zasina. 2020. *The InterCorp Corpus*, Version 13 of 1 November 2020.

Šebestová, D., and Malá, M. 2019. Expressing Time in English and Czech Childrens Literature: A Contrastive N-Gram-Based Study of Typologically Distant Languages' In *Language Use and Linguistic Structure: Proceedings of the Olomouc Linguistics Colloquium 2018*, 469–483 Olomouc: Palacky University.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Škrabal, M. and Vavřín, M. 2017. Databáze překladových ekvivalentů Treq. *Časopis pro moderní filologii* 99(2):245–60.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.

Vašků, K., Brůhová, G. and Šebestová, D. 2019. Phraseological Sequences Ending in of in L2 Novice Academic Writing. In *Computational and Corpus-Based Phraseology. EUROPHRAS 2019, Lecture Notes in Computer Science*, G. Corpas Pastor and R. Mitkov (eds), 431–443. Cham: Springer.

Vavřín, M. and Rosen, A.. 2015. Treq (v. 2.1). *Treq*. https://treq.korpus.cz/ [Last accessed 28 April 2021].

Author's address

Denisa Šebestová
Faculty of Arts, Department of English Language and ELT Methodology
Charles University
nám. Jana Palacha 1/2
CZ-Prague 110 00
Czech Republic
denisa.sebestova@ff.cuni.cz