FROM CORPUS DATA TO CONSTRUCTIONAL NETWORKS:
ANALYZING LANGUAGE WITH THE USAGE-BASED CONSTRUCTION
GRAMMAR FRAMEWORK

JAKOB HORSCH
Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia
Faculty of Languages and Literatures, Catholic University of Eichstätt-Ingolstadt,
Eichstätt, Germany

HORSCH, Jakob: From corpus data to constructional networks: Analyzing language
with the Usage-based Construction Grammar framework. Journal of Linguistics, 2023,

Abstract: Construction Grammar (CxG) is an innovative approach to language that
has become increasingly popular in the Anglosphere over the last 30 years. In CxG, the
basic units of linguistic analysis are constructions: arbitrary and conventional form-meaning
pairings, reminiscent of Saussure’s linguistic sign, but applied to levels of linguistic analysis
beyond the lexicon. A large body of research has provided ample evidence in support of
CxG. However, the theory remains unknown to many colleagues outside the Anglosphere.

In this paper, I highlight a particularly interesting strand of CxG that is referred to as
‘usage-based’, an approach that assumes constructions are learned based on input frequency,
that is, through repeated exposure to and use of a linguistic structure (hence usage-
based).

The main aim of this paper is thus to demonstrate how corpus data can be analyzed to find
evidence for ‘entrenchment’ of linguistic structures and thus, the existence of constructions.
I will illustrate this procedure by applying so-called covarying-collexeme analyses to data
from the Slovak National Corpus (SNC) and the Slovak Web 2011 corpus from which
I extracted 785 tokens of the so-called Comparative Correlative (CC) construction (e.g. Čím
viac čítam, tým viac rozumiem).

Keywords: Usage-based Construction Grammar, Comparative Correlative, Slovak,
meso-construction, mental network

1. INTRODUCTION

1.1 Construction Grammar and usage-based approaches

Construction Grammar (commonly abbreviated as CxG) is a linguistic theory
that has been steadily gaining popularity in the Anglosphere over the last three

1 The paper has been written within the Slovak National Corpus project supported by the Slovak Academy
of Sciences, Ministry of Education, Science, Research and Sport of the Slovak Republic, Ministry of
Culture of the Slovak Republic and the Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences.

Corpus data annotation and analysis was carried out in the context of the research project Comparing
Comparative Correlatives: The more languages, the better (DFG-HO 3904/5-2), funded by the German
Research Foundation DFG.
decades but does not appear to be well-known in other parts of the linguistic community, including Czechia and Slovakia. (I admit that this is a very subjective assessment based on my personal experience from attending various conferences in Czechia and Slovakia. I do not at all intend to suggest a complete lack of awareness of constructional approaches in these countries). As a cognitive approach to language, CxG differs quite fundamentally from the ‘traditional’ and dominant approach that is commonly referred to as Mainstream Generative Grammar (MGG, cf. Culicover – Jackendoff 2005, p. 3). MGG presupposes that linguistic knowledge – that is, the potential to produce an infinite amount of grammatical utterances – can be accounted for with words that are combined according to maximally abstract rules, and that only words can carry arbitrary and conventional meaning. Hence this approach is also referred to as a ‘words-and-rules’, or ‘lexical’ approach.

CxG, in contrast, posits that arbitrary and conventional form-meaning pairings (i.e., linguistic signs) exist at all levels of linguistic analysis, that is, they can be located anywhere along a continuum that represents a “gradation from contentful/lexical to procedural/grammatical” (Traugott – Trousdale 2013, p. 73, see also Hilpert 2013, p. 5). This includes “morphemes, words, idioms, partially lexically filled and fully general linguistic patterns” (Goldberg 2003, p. 219), making CxG an explicitly “non-modular” approach (Traugott – Trousdale 2013, p. 73) because the concept of “constructions at all levels […] defies a strict separation between grammar and lexicon” (Fried 2017, p. 244). Thus, “[t]he notion of a construction in [CxG] is much more general than the traditional notion of a construction” (Croft – Cruse 2004, p. 256). In CxG, constructions are “the basic unit of grammatical analysis” (Fried 2017, p. 243), and since a strong focus has been on how phrasal patterns carry meaning, the theory is also referred to as a ‘phrasal’ approach (as opposed to the ‘lexical’ approach, see above).

To illustrate this concept, consider the Slovak idiom *mať plné zuby* ‘to be annoyed’ (lit. ‘to have full teeth’), as in example (1):

(1) *Mám plné zuby (z toho).*
I-have full teeth from that
‘I am annoyed (by this)’

As noted by Croft and Cruse, idioms are “grammatical units larger than a word which are idiosyncratic” (2004, p. 230) because their “meaning or use can’t be predicted” (Nunberg et al. 1994, p. 493, cited in Croft – Cruse 2004, p. 230). This is because idioms are non-compositional: The individual words and the rules according

---

2 For example, an anonymous reviewer noted on the abstract for my presentation at the SlaviCorp 2018 conference: “According to the title of the contribution the author will present the topic within the frame [sic] of construction grammar, but *this framework is neither presented, nor mentioned* through the whole abstract.” (Emphasis JH). This implies one of two things: Either (1) the reviewer was not familiar with CxG or (2) the reviewer wanted me to explain the theory in more detail because he/she assumed that the audience would not be familiar with CxG.
to which they have been combined do not ‘add up’ to their overall meaning. When someone utters the sentence in example (1), he obviously does not intend to express that his teeth are full (full of what, anyways?). Under a generative analysis, idioms are therefore commonly ‘explained away’ as exceptions, since arbitrary form-meaning pairings (‘words’) are only possible on the level of the lexicon. In this regard, Croft and Cruse note that “one of the crucial characteristics of [the generative] model is that there are no idiosyncratic properties of grammatical structures larger than a single word” (2004, p. 227, emphasis JH).

CxG, on the other hand, will treat the whole structure [M ať plné zuby] as a form-meaning pairing in its own right. That is, CxG assumes that the entire phrasal pattern (and not just the individual words, which need to be combined according to maximally abstract rules to generate meaning) generates the meaning ‘to be annoyed’. This illustrates the notion of constructions as “complex, multidimensional sign[s], not in the traditional sense of a syntagmatic string of words” (Fried 2017, p. 243). In the case of [M ať plné zuby], the pattern consists of lexically fixed, invariable material (plné zuby cannot be replaced with e.g. *plný zub without changing/losing the meaning) as well as a partially open slot (the verb mať ‘to have’ must be conjugated but cannot be substituted by near-synonyms such as vlastniť ‘to own’).

In analogy to the ‘traditional’ Saussurean sign – an arbitrary and conventional form-meaning pairing – under a CxG analysis the Slovak idiom [M ať plné zuby] is therefore also a linguistic sign, since it combines a distinct form with a distinct meaning:

<table>
<thead>
<tr>
<th>Form:</th>
<th>[M ať plné zuby]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning:</td>
<td>‘to be annoyed’</td>
</tr>
</tbody>
</table>

Any definition of ‘construction’ in the CxG sense will necessarily be based on the original definition of the Saussurean sign, but extend it to phrasal patterns that are more abstract than simple lexical units:

C is a ConSTRUCTION if and only if C is a form-meaning pair <F_i, S_j> such that some aspect of F_i or some aspect of S_j is not strictly predictable from C’s component parts or from other previously established constructions. (Goldberg 1995, p. 4)

While originally, CxG “grew out of a concern to find a place for idiomatic expressions in the speaker’s knowledge of a grammar of their language” (Croft – Cruse 2004, p. 225), the notion of construction is not restricted to idioms. Table 1 provides example to illustrate the continuum ranging from “contentful/lexical to procedural/grammatical” (Traugott – Trousdale 2013, p. 73). At the contentful/lexical end, we find

---

3 Of course, from a diachronic perspective there might be a plausible explanation for the words used in a given idiom. However, the average speaker will not be aware of this. Therefore, it is appropriate to treat idioms as non-compositional.
fully substantive constructions (i.e., phonologically/lexically invariable) and at the procedural/grammatical end, fully schematic constructions consisting entirely of open slots, where “all elements are lexically open” (Croft – Cruse 2004, p. 248).

Table 1: Examples of constructions of varying degrees of schematicity

<table>
<thead>
<tr>
<th>Form</th>
<th>Meaning</th>
<th>Example</th>
<th>Schematicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>krížom krážom</td>
<td>‘in all directions’, ‘all over’</td>
<td>Precestoval Slovensko krížom krážom. ‘He travelled all over Slovakia.’</td>
<td>fully substantive</td>
</tr>
<tr>
<td>[ne-ADJ1]</td>
<td>‘not Property1’</td>
<td>nešťastný ‘unhappy’</td>
<td>partly substantive, partly schematic</td>
</tr>
<tr>
<td>[MAŤ plně zuby]</td>
<td>‘to be annoyed’</td>
<td>Mám plné zuby z toho. ‘I am annoyed by this.’</td>
<td>partly substantive, partly schematic</td>
</tr>
<tr>
<td>[Čím […]comparative element 1 (clause 1)]&lt;sub&gt;C1&lt;/sub&gt; [tým […]comparative element 2 (clause 2)]&lt;sub&gt;C2&lt;/sub&gt;</td>
<td>‘comparative correlative’</td>
<td>Čím viac sa jazyku venujeme, tým rýchlejšie sa ho naučíme. ‘The more we dedicate ourselves to (a) language, the faster we learn it.’</td>
<td>partly substantive, partly schematic</td>
</tr>
<tr>
<td>[Obj1 NEMUSÍT]&lt;sup&gt;3&lt;/sup&gt;</td>
<td>‘to dislike patient,’</td>
<td>Karla Gotta nemusím. ‘I dislike Karel Gott.’</td>
<td>partly substantive, partly schematic</td>
</tr>
<tr>
<td>[Sbj1 V3 Obl2]&lt;sup&gt;4&lt;/sup&gt;</td>
<td>‘Theme&lt;sub&gt;1&lt;/sub&gt; moves Path&lt;sub&gt;2&lt;/sub&gt; by V&lt;sub&gt;3&lt;/sub&gt;-ing’</td>
<td>Peter cestuje do Košic. ‘Peter is travelling to Košice.’</td>
<td>fully schematic</td>
</tr>
</tbody>
</table>

Now, one of the main tenets of CxG is that the grammar of a given language does not simply consist of an inventory of constructions (unlike the lexical approach, where the lexicon consists of a list of words), but that these constructions are interconnected in a “network […] that captures our knowledge of language in toto” (Goldberg 2003, p. 223). That is, constructions are stored holistically in a speaker’s so-called “mental grammar” (Hoffmann – Trousdale 2013, p. 3) and linked based on formal similarity (vertical links) and semantic similarity (horizontal links). The complete network of constructions is referred to as the “construct-i-con” (Jurafsky 1992, p. 302; Goldberg 2003, p. 219), a term that is “a morphological blend of construction and lexicon” (Ziem – Flick 2019, p. 203).

Among the many offshoots of Construction Grammar, one of the most successful has been Usage-based Construction Grammar, an approach with a focus on input

---

<sup>3</sup> This intriguing construction was first discussed by Hansen et al. (2011).  
<sup>4</sup> In CxG-based literature, this pattern is commonly referred to as the “Intransitive Motion construction” (cf. e.g. Goldberg 1995, p. 109).
frequency\textsuperscript{6}. Proponents of Usage-based CxG assume that in addition to idiomatic expressions, fully compositional structures can also be stored as constructions, provided they are encountered and used often enough by speakers – hence the term *usage-based*, which was coined by Langacker (1988) “to emphasize the importance of usage […] for the analysis of linguistic structure” (Diessel 2015, p. 295). Under this approach the definition of construction (see above) is extended beyond non-compositional structures: “[Phrasal] patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency”\textsuperscript{7} (Goldberg 2006, p. 5). The storage of patterns due to their frequent use is commonly referred to as *entrenchment*: “with repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence” (Langacker 1987, p. 59).

It follows from this that “the mental grammar of speakers is shaped by the repeated exposure to specific utterances” (Hoffmann – Trousdale 2013, p. 4), or in other words, “knowledge of language emerges from language use” (Croft – Cruse 2004, p. 3) – hence the term *usage-based* CxG. Therefore, usage-based approaches assume that language is acquired based on so-called domain-general cognitive mechanisms such as schematization and generalization. That is, the ability of language learners to abstract and ‘store’ linguistic patterns from the repeated input of various tokens of the same type\textsuperscript{8}. For example, a language learner might be exposed to the following utterances (2a-c):

\begin{enumerate}
\item a. *Father is driving the car.*
\item b. *The uncle is driving the van.*
\item c. *The neighbor is driving the truck.*
\end{enumerate}

Due to schematization across these three utterances, slightly more abstract semantic roles will then arise: \{\textsc{driver} [is \textsc{driving}] \textsc{vehicle}\}. By means of generalization (‘analogy’), the learner has thus been able to detect similarities across constructions and a new syntactic construction arises in her mental grammar. Later on, she will compare this construction with other input, for example (3a-c):

\begin{enumerate}
\item a. *Mother is singing a song.*
\item b. *Grandma is baking a cake.*
\item c. *The man is mowing the lawn.*
\end{enumerate}

As a result, even more abstract semantic roles arise, and a new construction, or node in the network, emerges that expresses a force-dynamic relationship between

\textsuperscript{6} There are CxG-based approaches that do not assume frequency to play an important role, e.g. complete inheritance approaches.
\textsuperscript{7} For a discussion of what qualifies as ‘sufficient frequency’, the reader is referred to Section 1.2 below.
\textsuperscript{8} Actually, experiments have revealed that abstract constructions are best learned through skewed input – that is, a combination of high type frequency with one high-frequent token (Casenhiser – Goldberg 2005).
an Agent A and a Patient B (note how this requires the learner to generalize both formally and semantically):

<table>
<thead>
<tr>
<th>Form</th>
<th>[A is V-ing B]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>‘Agent is V-ing Patient’</td>
</tr>
</tbody>
</table>

Thus, schematization and generalization, combined with sufficient input frequency, will ultimately lead to the entrenchment of highly abstract constructions – in this case, the phrasal pattern that corresponds to the Subject-Verb-Object rule in English:

<table>
<thead>
<tr>
<th>Form</th>
<th>[Sbj₁ V₃ Obj₂]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>‘Agent₁ affects Patient₂ by V₃-ing’</td>
</tr>
</tbody>
</table>

Note that each of these constructions – [Driver is Driving Vehicle], [A is V-ing B], and [Sbj₁ V₃ Obj₂] – is stored (‘entrenched’) as a separate node in the mental network. Nevertheless, they are interconnected based on formal and semantic similarity. To distinguish between the different levels of abstractness, CxG uses the following terms:

- micro-construction, i.e., fully substantive constructions such as the actual utterances in (2) and (3),
- meso-construction, i.e., any constructions that are neither fully substantive nor fully schematic (e.g. [Driver is Driving Vehicle] and [A is V-ing B]),
- macro-construction, i.e., fully schematic constructions such as [Sbj₁ V₃ Obj₂].

Every time the learner later produces utterances, she will rely on multiple ‘nodes’ (i.e., meso-constructions) of the network, and not just the most abstract construction. Thus “[e]ach node inherits the properties of its dominating nodes” (Traugott – Trousdale 2013, p. 61). In CxG parlance, every micro-construction is a “specific, substantive instance of a construction” (Hoffmann et al. 2019, p. 26). As such it inherits from multiple ‘parent’ meso-constructions, which in turn inherit from even more abstract meso-constructions and ultimately, a macro-construction. This implies a network that is strictly hierarchical: In essence, inheritance links indicate taxonomic constraints; they “capture the fact that all nonconflicting information between two related constructions is shared” (Goldberg 1995, p. 74-75). Ultimately, then, “[t]he inheritance network lets us capture generalizations across constructions” (Goldberg 1995, p. 67).

Croft and Cruse (2004, p. 321) illustrate the notion of such taxonomic networks with a simple example, the English imperative construction (e.g. Stay awake! Don’t worry!). At the bottom of Figure 1 are examples of the various micro-constructions that a learner might be exposed to (and that a speaker might produce, for that matter).
The usage-based approach assumes that based on sufficient exposure to these micro-constructions, a learner will generalize and entrench more abstract meso-constructions such as the positive imperatives \([\text{VERB}]\) and \([\text{be ADJ}]\) as well as their negative counterparts \([\text{don't ADJ}]\) and \([\text{don't be ADJ}]\). After more input she will generalize even further to the more abstract \([\text{PREDPHRASE}]\) and \([\text{don't PREDPHRASE}]\) meso-constructions. Ultimately, she will store a maximally abstract imperative construction as a macro-construction.

![Figure 1: The taxonomy of English imperative constructions (from Croft – Cruse 2004, p. 321)](image)

When later producing micro-constructions, she will then draw on this part of her mental network as well as other meso-constructions (e.g. the schematic \([\text{ADJ}]\) construction). Thus, a micro-construction such as \(\text{Be happy!}\) is said to inherit from the \([\text{ADJ}]\) construction and the \([\text{be ADJ}]\) meso-construction, which in turn inherits from the \([\text{PREDPHRASE}]\) construction, and so on.

As Figure 1 shows, learning language in this way will result in a “language network that is rather baroque, involving massive redundancy and vastly rich detail” (Traugott – Trousdale 2013, p. 53). The usage-based CxG approach is therefore “maximalist, nonreductive, and bottom-up” (Langacker 1988, p. 264) as opposed to ‘top-down’ generativist approaches, in which “the grammar of a language is reduced to the smallest possible set of statements, with all redundancy avoided” (Langacker 1988, p. 262).

A large body of empirical evidence in support of the usage-based approach to language (i.e., the existence of a mental grammar consisting of interconnected constructions of varying degrees of abstractness) comes from various disciplines such as first- and second-language acquisition, psycholinguistics and neurolinguistics (Hoffmann – Trousdale 2013, p. 3). However, it appears that these studies are not well known in Czechia and Slovakia. One question that I have been asked many times is whether there is any evidence for the existence of constructions. It is this question that I wish to address in the following, using a small corpus study to
illustrate the process of finding evidence for the existence of constructions and how they are arranged in a constructional network.

In fact, corpus data play a crucial role in Usage-based CxG, a theory based on the assumption that grammar is shaped by frequency of use (see above). This follows from Schmid’s “from corpus-to-cognition principle” (2000, p. 39) which can be subdivided into two principles, as discussed in Hoffmann et al. (2019, p. 10):

- the **“corpus-as-output” principle** (Stefanowitsch – Flach 2017, p. 101-103) posits that corpus data can be used to make “inferences about the mental representations” (Stefanowitsch - Flach 2017, p. 102-103, as cited in Hoffmann 2019, p. 18).
- the **“corpus-as-input” principle** states that corpus data provides evidence of “the input that speakers […] are exposed to” (Hoffmann 2019, p. 18), thus influencing the form of their mental networks (Stefanowitsch – Flach 2017, p. 103).

As large collections of “(1) machine-readable (2) authentic texts […] which [are] (3) sampled to be (4) representative” (McEnery et al. 2006, p. 5), modern corpora allow researchers to quite effortlessly retrieve the frequencies of even the most abstract phrasal patterns (so-called types) thanks to POS tagging and regular expressions. The aim is to find out whether linguistic patterns “occur with sufficient frequency” (Goldberg 2006, p. 5) to be considered constructions.

Corpora come with several advantages, as noted by Hoffmann: Because they “easily allow hypotheses to be tested by different researchers the objectivity and falsifiability of this data source is […] uncontroversial” (2006, p. 168). It is therefore unsurprising that the majority of evidence presented in Usage-based CxG studies comes from corpora, although there are other sources as well, e.g. elicited data obtained by means of psycholinguistic experiments. Such experiments have been used to corroborate corpus study findings in previous studies (e.g. Gries et al. 2010, Bresnan 2007, Bresnan – Ford 2010, Horsch 2023).

In the following, instead of a comprehensive and long-winded overview of relevant literature, which would include an overwhelming amount of research based on English constructions, I will attempt to illustrate the Usage-based CxG approach using a case study featuring the Slovak Comparative Correlative (CC) construction (e.g. Čím viac čítam, tým viac rozumiem ‘The more I read, the more I understand’). The aim is to show how corpus data can provide evidence for the existence of constructions, specifically using **Covarying-collexeme analysis** (Gries 2007), a statistical method that is known as a “standard test in Construction Grammar” (Hoffmann et al. 2019, p. 11). I will also demonstrate how the results obtained by means of Covarying-collexeme analysis can help shed light on the nodes and links of the constructional network underlying constructions.

The paper is structured as follows: In Section 1.2 I discuss the Comparative Correlative Construction, providing a brief overview of its complex semantics and
formal features. In Section 2.1, I discuss the corpus data used in the case study, and in Section 2.2, I show how Covarying-collexeme analysis works. In Section 3, the results are presented, and in Section 4 I discuss them within the framework of Usage-based Construction Grammar. In Section 5, I present my conclusion.

1.2 The Slovak Comparative Correlative (CC)

The Comparative Correlative (CC) construction prototypically consists of two clauses that I will refer to as C1 and C2 in the following. It has attracted considerable attention over the last two or three decades (for a concise overview of studies on CCs in various languages I refer the reader to Horsch 2021 and 2023). Example (2), found in the Slovak National Corpus (SNC), illustrates what CCs typically look like in Slovak:

(4) [Čím  skôr   to  urobiš]_{C1}  [tým  lepšie  pre  teba.]_{C2}
čím sooner this do:you:2:SG TÝM better for you:2:SG
‘The sooner you do this, the better for you.’
<SNC public all version 6.0 TKem3>

I have described the Slovak CC construction in greater detail previously (Horsch 2019, Horsch 2020, Horsch 2021), so that in the following I will highlight only its most relevant features. Semantically, CCs are characterized by a “conditional, asymmetrical, effect-cause relationship between C1 and C2” as well as “parallel change over a certain time period” in C1 and C2 (Horsch 2021, p. 196). On the formal side, the Slovak CC construction (like many of its counterparts in other languages) is a phrasal pattern consisting of construction-specific invariant lexical/phonological elements (čím in C1 and tým in C2) and obligatory as well as optional slots. The obligatory slots can be filled with an adjective phrase (AdjP) (5), an adverb phrase (AdvP) (6), or a noun phrase (NP) (7). The clause slot is optional so that it is possible to distinguish between what I have referred to as “verbless clauses” (8) and “full clauses” (9) (Horsch 2021, p. 198), depending on whether the slot is filled or not.

(5) [čím bude  [mnohotvárnejšia]_{AdjP}  [tým bude  [užitočnejšia]_{AdjP}]_{C2}
čím will-be multifaceted TÝM will-be beneficial
‘The more multifaceted it will be, the more beneficial it will be.’
<SNC public all version 7.0 SME93/02>

(6) [čím  [váčšmi]_{AdvP} niekoho  ľúbíš]_{C1}
čím more someone love:you:2: sg
[tým  [váčšmi]_{AdvP}  t’  napokon  sklame.]_{C2}
TÝM more you:Acc ultimately disappoint
‘The more you love someone, the more he/she will ultimately disappoint you.’
<SNC public all version 7.0 BBra3>
The less people that understand a certain artwork, the better this is apparently.

The more emotive the better.

The better the ironworks are doing, the better off we will be as well.

Based on these features, I have suggested (Horsch 2019, p. 184) the following maximally abstract “constructional template” (10) (Hoffmann 2019, p. 184) for the Slovak CC construction:

Note that the phrasal pattern underlying the CC construction is “lexically irregular as well as syntactically and semantically irregular” (Croft – Cruse 2004, p. 235), a fact that appears to have caused considerable headache for generativist analyses: The CC has been relegated to the “periphery” (Borsley 2003); C1 and C2 have been described as “non-standard head-filler phrases” and the whole construction as “a non-standard head-adjunct phrase” (Borsley 2004, p. 71). Similarly, other generativist analyses have suggested that CCs “require special constructions” (Abeillé et al. 2006, p. 7), and that they are a “non standard type of subordinate construction” (Abeillé et al. 2006, p. 18).

Croft and Cruse have noted that English CCs are “schematic idioms” and are therefore composed of “unfamiliar pieces unfamiliarly arranged” (2004, p. 235). The observation also applies to the Slovak CC: For example, the clause-initial elements čím and tým resemble instrumental-case pronouns but do not function as such in the CC construction (cf. Horsch 2021, p. 197), which is why I refer to them as construction-specific (and not idiosyncratic, a term commonly employed in generativist analyses).
Such phenomena have led to the CC construction gaining a “reputation of being irremediably quirky” (den Dikken 2005, p. 509) in generativist analyses. I believe, however, that the Usage-based CxG approach can quite effortlessly accommodate the CC’s various idiosyncrasies. Addressing all of these, however, would be beyond the scope of this paper, whose aim is to show how evidence for the entrenchment of constructions can be found in corpus data. To highlight the advantages of the Usage-based CxG approach, in the following I will therefore focus on one interesting aspect of the CC: cross-clausal association phenomena.

Let us return to the maximally abstract CC template (10), which is certainly useful for accounting for the productivity of the construction and the varying schematicity of its constructional elements. However, it does remain unsatisfactory in several respects. For one, the order of constructional elements can vary in Slovak due to its more flexible word order (as described in Horsch 2021, p. 199-200). More importantly, a number of recent corpus studies on the English CC (e.g. Hoffmann 2019; Hoffmann et al. 2019, 2020; Horsch 2021, 2023) have revealed statistically significant cross-clausal associations across C1 and C2 using so-called covarying-collexeme analysis (cf. section 2.2). These cross-clausal associations suggest that the syntax of C1 and C2 is anything but context-free. For example, regarding what they call “truncation phenomena” (i.e. whether the optional clause slots in C1 and C2 are filled), Hoffmann et al. (2019) have demonstrated that in English CCs there is a statistically significant preference for C2 to be truncated/verbless (i.e., to not contain a clause) when the same is the case for C1. In fact, Hoffmann et al. (2019) found several such cross-clausal associations in English, also with regard to lexical fillers in C1 and C2 (e.g. skôr and lepšie) and filler types (AdjP, AdvP, NP; see above).

That is, despite the manifold choices that speakers have regarding the obligatory and optional slots in the template (10), when C1 is truncated, there is a statistically significant probability of C2 also being truncated, indicating that this cross-clausal association must somehow be stored in speakers’ mental grammars. In Horsch (2021), using a 500-token sample from the SNC, I obtained similar results for Slovak. Again, there was a statistically significant preference for C2 to not contain a verb (i.e., a verbless clause) when the same was the case for C1 (cf. Horsch 2021, p. 214).

This clearly shows that what syntactically happens in C1 has an influence on what happens in C2 and vice versa, something that cannot be captured by maximally abstract templates such as (3). Moreover, this has serious implications for generativist analyses that have claimed that the CC’s “idiosyncrasy[ies] reside wholly in the lexical domain, not in the syntax” (den Dikken 2005, p. 529) and thus presuppose the independent realization of C1 and C2: As noted by Hoffmann et al., cross-clausal associations are “a constraint that cannot be captured by any […] analyses of CCs that treat the two daughter nodes (C1 and C2) as independent slots” (2019, p. 7).

Under a Usage-based CxG approach, cross-clausal association phenomena are easily explained: As patterns that “occur with sufficient frequency” (Goldberg 2006,
they are partly substantive and partly schematic meso-constructions, located at a level somewhere in-between the macro-level and the micro-level of the constructional network underlying the CC. Take the lexical filler pair \( skôr_{C1} \)-lepšie\(_{C2} \), for example, which in my SNC study (Horsch 2021) turned out to appear significantly more often in the data than could be expected by chance (for details on the statistical analysis, see section 2.2 below). Because there is a distinct meaning and a distinct form involved, it is possible to posit the following meso-construction (11) (adapted from Horsch 2021, p. 216):

\[
(11) \quad \text{skôr}_{C1} \text{-lepšie}_{C2} \text{ CC meso-construction}
\]

**Form:**

\[
[tʃi:m] \ [skôr]_{\text{comparative element1}} \ [\ldots]_{\text{optional clause}} \ [C1] \\
[[tʃi:m] \ [lepšie]_{\text{comparative element2}} \ [\ldots]_{\text{optional clause}} \ [C2]
\]

**Meaning:**

‘[As the point in time (with respect to clause 1) decreases] independent variable \[\rightarrow \] {so the degree of quality increases (with respect to clause 2)} dependent variable in a monotonic way’

Note that this is a meso-construction that licenses micro-constructions such as čím skôr začnete, tým lepšie pre vás ‘the sooner you begin, the better for you’ (SNC prim-7.0-public-all HN2012/12) and in turn inherits from the maximally schematic CC macro-construction whose form corresponds to the template above (10).

In Section 4, I will discuss this notion in greater detail using network illustrations. For now, suffice it to say that this constructional network is indeed highly redundant or, in the words of Traugott and Trousdale, “baroque” (2013, p. 53).

Of course, the question remains what exactly “sufficient frequency” (Goldberg 2006, p. 5) should be. I will address this question in section 2.2 below, after a brief overview in section 2.1 of how the corpus data used in the case study was extracted and annotated.

2. **DATA AND METHODOLOGY**

2.1 **Data extraction and annotation**

This corpus study is based on two data sets that were drawn from the following corpora:

- Slovak National Corpus (version: *prim-7.0-public-all*)
- Slovak Web 2011 corpus

The SNC was queried using its web interface at https://korpus.sk/index_en.html. This is a corpus of written Slovak consisting of almost 1 bn words, with 65.1% journalistic, 15.1% fiction, 9.5% professional and 10.3% ‘other’ types of text (Šimková et al. 2017, p. 27). The following CQL query\(^{10}\) was used to extract CCs:

\(^{10}\) The same queries (and resulting data set) were used for previous articles about the Slovak Comparative Correlative Construction (Horsch 2019, 2021).
Using this query, a total of 10,151 tokens was extracted from the SNC, from which a random sample of 500 tokens was drawn. 58 irrelevant tokens were discarded, leaving a data set of 442 C1C2 tokens for statistical analysis.

The Slovak Web 2011 corpus was queried using the SketchEngine web interface at https://www.sketchengine.eu/sktenten-slovak-corpus/. The corpus consists of around 715 m words and contains texts from the internet, thus complementing the SNC (which contains mostly texts from printed sources) well. The following query was used to extract CCs from the corpus:

This yielded a total of 4,998 tokens, of which 503 had been coded by a student assistant at the time of writing. Of these 503 tokens, 343 turned out to be relevant to the study (i.e., C1C2 tokens).

To investigate cross-clausal associations, the variables in Table 2 were coded in spreadsheets using LibreOffice Calc (The Document Foundation 2022):
2.2 Data analysis: Covarying-collexeme analysis

As discussed in section 1.1, one question that needs to be answered is what exactly “sufficient frequency” for the entrenchment of a construction is. An important distinction that needs to be made when answering this question is that between type and token frequency. Whereas token frequency is evidence for specific and substantive constructions, type frequency is evidence for the entrenchment of more abstract constructions. Thus, the utterance *Father is driving a car* (2a) is a token, as are *The uncle is driving the van* (2b) and *The neighbor is driving the truck* (2c). If these three example sentences were to appear in a corpus as tokens, they could then be used to determine a type – in this case, the [Driver] [is Driving] [Vehicle] construction. It is therefore type frequency that plays a crucial role in identifying meso-constructions: If structures “have been encountered with many different lexicalizations [i.e., tokens]”, this will lead “to the entrenchment of […] more abstract constructions [i.e., types]” (Hoffmann 2019, p. 9). In other words, the higher the frequency, the more entrenched a (meso-)construction is. But things are not as simple: Entrenchment is a gradual process and there is no agreement in the literature about a particular frequency threshold that must be crossed to make a structure a construction in its own right.

The question about frequency thresholds therefore has to be slightly rephrased at this point: The method of statistical analysis employed in the present corpus study works with relative frequencies, since it “compares actual frequencies of co-occurrence with expected ones” (Stefanowitsch and Gries 2005, p. 9). That is, a syntagm is compared relative to all other possible instantiations (see below). Thus, the question is not what absolute frequency should be used as sufficient evidence for the existence of a construction, but how the results of statistical analyses should be interpreted.

Following convention in the literature, a $p$-value of 0.05 will therefore be used as a significance threshold. Additionally, two more $p$-value thresholds (0.01 and 0.001) will be used to differentiate between degrees of entrenchment (see below). It will thus be possible to provide evidence for the existence of constructions and their degrees of entrenchment, and visualize how they are interrelated in a network.

Thus, following the methodology of Hoffmann et al. (2019), the annotated SNC and Slovak Web 2011 data was subjected to covarying-collexeme analysis (Stefanowitsch and Gries 2005: 9-11) with the aim of detecting cross-clausal associations between C1 and C2. Known as the “standard test in Construction Grammar for the association of two constructional elements” (Hoffmann et al. 2019: 11), covarying-collexeme analysis belongs to a ‘family’ of statistical analyses devised by Gries and Stefanowitsch in the 2000s (cf. Stefanowitsch – Gries 2003; 2008; Gries – Stefanowitsch 2004; 2006; 2010; Gries 2011). These statistical analyses were developed to provide “empirical support […] to construction-based syntactic theories” (Schmid – Küchenhoff 2013, p. 537). In the
present study, covarying-collexeme analysis was used to test for the association of each of the linguistic variables Lexical Filler (levels e.g. viac, lepšie etc.), Filler Type (levels: AdjP, AdvP, NP, PP), and Verb Presence (levels: No Verb, Full Clause) across C1 and C2, and for interactions between these variables (see below). To this end, the Coll.analysis 3.2a script (Gries 2007) for R for Windows (R Core Team 2020) was employed, which “compares actual frequencies of co-occurrence with expected ones on the basis of a 2-by-2 distribution table” (Stefanowitsch – Gries 2005, p. 9).

As noted by Schmid and Küchenhoff, such two-by-two distribution tables are “familiar to many linguists from applications of the $\chi^2$-test” (2013, p. 534). However, chi-square tests “always require the expected frequencies of all cells to be greater than 5” (Hoffmann 2011, p. 23; referring to Woods et al. 1986, p. 144f.), so that significant portions of data would have to be ignored: In the 442-token SNC data set used in the present study, for example, there were 52 Lexical Filler hapaxes (i.e., words that only appeared once) in C1 and 83 in C2. With a frequency of 1 (and thus $\leq 5$), all of these would have to be excluded from a chi-square test. Covarying-collexeme analysis addresses this issue by employing the so-called Fisher-Yates Exact test, which is “very precise because it handles even small frequencies well” (Horsch 2021, p. 210, referring to Gries 2015, p. 313).

The $p$-values of the Fisher-Yates Exact test are used to “gauge the degree of attraction” (Schmid – Küchenhoff 2013, p. 535) between a construction (e.g. the Comparative Correlative) and constructional elements (e.g. Lexical Fillers, Filler Types and Verb Presence). Note that Stefanowitsch and Gries originally intended covarying-collexeme analysis to be applied to lexemes in constructions, stating that it can be used to investigate “cooccurrence between three signs (lexeme$_1$, lexeme$_2$, and construction)” (Stefanowitsch – Gries 2005, p. 10). As Hoffmann et al. (2019) have demonstrated, however, covarying-collexeme analysis can easily be applied to more abstract constructional slots as well, e.g. Filler Types).

Thus, using observed frequencies as input, covarying-collexeme analysis allows researchers to investigate a “set of choices available in a given position of a syntagmatic structure in relation to the set of choices available in another position” (Stefanowitsch – Gries 2005, p. 10). Applied to the present study, this means that covarying-collexeme analysis could be used to “compare the choices in C1 to the choices in C2 for any variable based on the raw data for the slots in C1 and C2 for these variables” (Horsch 2021, p. 209).

To illustrate this procedure, let us consider the lexical fillers skôr ‘sooner’ in C1 and lepšie ‘better’ in C2 from example (11) and treat them as lexeme$_1$ and lexeme$_2$ of the construction under investigation, the Slovak CC. As Hoffmann has noted, “in order to statistically assess the association” of any two constructional elements using covarying-collexeme analysis, “this syntagm must be compared to all other logically possible lexicalizations” (2011, p. 23). Thus, the Coll.analysis 3.2a script requires as
input the (observed) frequency of tokens with skôr ‘sooner’ in C1 and lepšie ‘better’ in C2, which it will compare to the token frequencies of

- skôr in C1 occurring with all lexical fillers in C2 other than lepšie (skôr + ¬lepšie)
- lepšie in C2 occurring with all lexical fillers in C1 other than skôr (¬skôr + lepšie)
- neither skôr in C1 nor lepšie in C2 occurring (¬skôr + ¬lepšie)

Table 3 shows the ‘input’ of a covarying collexeme-analysis as a two-by-two distribution table, using the frequencies found in the SNC sample (the totals are relevant for calculating expected frequencies).

Table 3: Example of a two-by-two table as used in covarying-collexeme analysis

<table>
<thead>
<tr>
<th>skôr in C1 (word skôr in C1)</th>
<th>lepšie in C2 (word lepšie in C2)</th>
<th>¬lepšie in C2 (all words except lepšie in C2)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(frequency of skôr in C1 and lepšie in C2)</td>
<td>35 (frequency of skôr in C1 and lepšie in C2)</td>
<td>17 (frequency of skôr in C1 and ¬lepšie in C2)</td>
<td>52</td>
</tr>
<tr>
<td>¬skôr in C1 (all words except skôr in C1)</td>
<td>62 (frequency of ¬skôr in C1 and lepšie in C2)</td>
<td>328 (frequency of ¬skôr in C1 and ¬lepšie in C2)</td>
<td>390</td>
</tr>
<tr>
<td>Totals</td>
<td>97</td>
<td>345</td>
<td>442</td>
</tr>
</tbody>
</table>

Note that Table 3 shows only one out of many two-by-two tables that the Coll.analysis 3.2a script will test. In fact, it will test all possible Lexical Filler combinations (including combinations not attested in the data) based on the individual Lexical Fillers that it finds in the data set. Thus, in case of the SNC data set, the Coll.analysis 3.2a script actually applied the Fisher-Yates Exact test to a total of 7,526 two-by-two tables (71 unique Lexical Fillers in C1 × 106 unique Lexical Fillers in C2).

In the output provided by the Coll.analysis 3.2a script, the most important value is the so-called collostructional strength, a negative log-transformed p-value of the Fisher-Yates Exact test (Gries 2007) that shows the strength of attraction between a construction and constructional elements (see above). Put differently, collostructional strength is an indication of how strong the evidence against the null hypothesis is, which translates to “the independence of occurrence” of a construction and the constructional elements under investigation (Schmid – Küchenhoff 2013, p. 535). Thus, “[t]he lower the p-value, the stronger the evidence against the null hypothesis” (Schmid – Küchenhoff 2013, p. 539).

The collostructional strength values provided by the Coll.analysis 3.2a script correspond to the following p-values:
• $p$-value of 0.001 = coll. strength of 3 (***)
• $p$-value of 0.01 = coll. strength of 2 (**) 
• $p$-value of 0.05 = coll. strength of 1.30103 (*) 

Thus, collostructional strengths higher than 1.30103 indicate a $p$-value of less than 0.05, a widely accepted significance threshold in statistics. This also explains why collostructional strength is provided as a log-transformed $p$-value: Significant results (i.e., $p$-values < 0.05), i.e. those that are of greatest interest, are easier to interpret this way (as noted by Stefanowitsch and Gries 2005, p. 7; Gries et al. 2005, p. 671). By way of example, consider a $p$-value of 0.0000000000001, which as a negative log-transformed value amounts to 13.37 (this is the collostructional strength of *skôr* ‘sooner’ in C1 and *lepšie* ‘better’ in C2). This is a value that is much easier to interpret both against the significance threshold of 1.30103 ($p = 0.05$) (as noted by Horsch 2021, p. 210) as well as other significant results.

Beyond just indicating collostructional strength, covarying-collexeme analysis also provides information about attraction and repulsion, based on observed and expected frequencies: If the observed frequency is higher than the expected frequency, the output of the Coll.analysis 3.2a script will indicate ‘attraction’ in a separate column; conversely, if the expected frequency is higher than the observed frequency, the output will indicate ‘repulsion’.

In this context, another advantage of using negative log-transformed $p$-values (also rendered as ‘$p_{\log_{10}}$’) to indicate collostructional strength is that in this notation, “the sign of the resulting $p_{\log_{10}}$ value is rendered as ‘a plus when the observed frequency is higher than the expected one’ (Stefanowitsch – Gries 2005, p. 7)” (Hoffmann 2011, p. 24), indicating attraction. Conversely, “a $p_{\log_{10}}$ value of −2 indicates that a pair of words appears significantly less often in a construction than expected” (Hoffmann 2011, p. 24), indicating repulsion.

That is, covarying-collexeme analysis provides either a positive (‘attraction’) or a negative (‘repulsion’) collostructional strength value, the latter of which indicates that the observed frequency is lower than the expected frequency (Stefanowitsch – Gries 2005, p. 7). Thus, it is also possible to determine which constructional elements particularly ‘dislike’ each other. As an example, a collostructional strength of −2 means that in the data, there was a significantly lower observed frequency of co-occurrence of two constructional elements than could be expected by chance, whereas a collostructional strength of +2 means a significantly higher observed frequency of co-occurrence of two constructional elements than could be expected by chance. In both cases, the $p$-value yielded by a simple statistical analysis would simply be 0.01, a result that does not provide any information about attraction or repulsion.

Expected frequencies are calculated using the formula $(row\ total \times\ column\ total) \div\ grand\ total$. Returning to the example used above (*skôr* ‘sooner’ in C1 and
lepšie ‘better’ in C2, cf. Table 3), this would amount to \((52 \times 97) \div (442) = 11.41\). The expected frequency (11.41) is thus considerably lower than the observed frequency (35), so the Coll.analysis 3.2a script will indicate ‘attraction’ in this case.

The output of the Coll.analysis 3.2a script is a table, “ranked from attraction to repulsion and from the highest to the lowest collostructional strength” (Horsch 2021, p. 212). The table also shows the observed frequency of each level of the first factor by itself and the same for the second factor; and it shows the observed and expected frequencies of each factor level combination. Furthermore, the relation (attraction or repulsion) is shown, as well as a collostructional strength value and, using asterisks, which significance threshold has been crossed (see above).\(^\text{16}\)

Table 4, adapted from a previous study using data from the SNC (Horsch 2021) shows the first row of the results table of a covarying-collexeme analysis carried out on the data set used for the present investigation. We see that skôr in C1 and lepšie in C2 had the strongest collostructional strength, i.e., it was the combination with the largest difference between observed frequency (35) and expected frequency (11.41). The table also shows that skôr appeared 52 times in a C1 subclause and lepšie 97 times in a C2 subclause, and that they are attracted to each other with a highly significant collostructional strength of 13.37, which amounts to a p-value of less than 0.001, hence the three asterisks (**). Any significant results that exhibit attraction are rendered with light gray shading in the table; any significant results with repulsion with dark gray.

**Table 4:** Example of covarying-collexeme analysis results table (adapted from Horsch 2021, p. 212)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>skôr</td>
<td>lepšie</td>
<td>52</td>
<td>97</td>
<td>35</td>
<td>11.41</td>
<td>attraction</td>
<td>13.37</td>
<td>***</td>
</tr>
</tbody>
</table>

Before proceeding with a discussion of how corpus frequencies can be interpreted as evidence for the existence of constructions, it must be noted that covarying-collexeme analysis is not entirely uncontroversial. Perhaps most importantly, it has been criticized that it is “somewhat odd” (Baayen 2011, p. 16) to derive degrees of attraction and repulsion from \(p\)-values, which are indications of likelihood with which a null hypothesis can be rejected (cf. Schmid – Küchenhoff 2013, p. 539-540). In other words, \(p\)-values actually only indicate how strong or weak the evidence against the null hypothesis is, and not really any degree of

---

\(^{16}\) The results table also shows two so-called so-called \(\Delta P\) values that indicate how dependent one constructional element is on the other (cf. Gries 2015). For the present investigation, however, these are not relevant, so they will not be included in the results tables.
attraction/repulsion. Nevertheless, I believe that covarying-collexeme analysis is one of the best tools currently available for describing interdependence between constructional slots, a fact that is also reflected in its widespread use in constructional literature.

3. RESULTS

In the following, the results of the covarying-collexeme analyses are presented, beginning with the individual variables, i.e. **LEXICAL FILLERS** (3.1), **FILLER TYPES** (3.2) and **VERB PRESENCE** (3.3). Note that I have reported the results from the SNC data in a different publication (Horsch 2021). However, in this investigation I will expand on these results by including data from the Slovak Web 2011 corpus and furthermore, testing for interactions between factors. This will allow me to illustrate the constructional network underlying the Slovak CC in considerably greater detail.

Interactions such as **LEXICAL FILLER:FILLER TYPE** were examined in line with Hoffmann et al.’s (2019) methodology. Apart from allowing for a more detailed impression of the constructional network underlying the CC, this was also done to ensure that cross-clausal **FILLER TYPE** associations were really meso-constructions and not just due to “specific **LEXICAL FILLER** interactions that interact with the variable **FILLER TYPE**” (Hoffmann et al. 2019, p. 15). For example, covarying-collexeme analysis might detect an **AdvP C1-AdvP C2** association due to the fact that there was a high frequency of the **LEXICAL FILLER viac** (i.e., concrete instantiations of **AdvPs**) in the data set. Therefore, the variables **LEXICAL FILLER** and **FILLER TYPE** were collapsed and subjected to separate covarying-collexeme analyses.

The following two- and three-way interactions were tested to reveal a more detailed picture of the Slovak CC’s meso-constructional network:
- **LEXICAL FILLER:FILLER TYPE** (3.4.1)
- **LEXICAL FILLER:VERB PRESENCE** (3.4.2)
- **FILLER TYPE:VERB PRESENCE** (3.4.3)
- **LEXICAL FILLER:FILLER TYPE:VERB PRESENCE** (3.4.4)

3.1 LEXICAL FILLER

Table 5 shows the results\(^\text{17}\) of the covarying-collexeme analysis of **LEXICAL FILLERS** in the SNC data; Table 6 shows the results for the Slovak Web 2011 data:

\(^{17}\) Note that in the following, statistically significant results have gray shading; light gray for attraction and dark gray for repulsion. Furthermore, large tables will only show results with an expected frequency of \(\geq 5\), “an assumption often required for goodness-of-fit tests such as the chi-square test” (Hoffmann et al. 2019, p. 15; referring to Gries 2009, p. 152).
Table 5: Covarying-collexeme analysis results of Lexical Filler, SNC (expected frequency ≥5, adapted from Horsch 2021, p. 213)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>skôr</td>
<td>lepšie</td>
<td>52</td>
<td>97</td>
<td>35</td>
<td>11.41</td>
<td>attraction</td>
<td>13.37 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>viac</td>
<td>110</td>
<td>108</td>
<td>47</td>
<td>26.88</td>
<td>attraction</td>
<td>6.21 ***</td>
<td></td>
</tr>
<tr>
<td>viac</td>
<td>viac</td>
<td>88</td>
<td>108</td>
<td>31</td>
<td>21.5</td>
<td>attraction</td>
<td>2.13 **</td>
<td></td>
</tr>
<tr>
<td>viac</td>
<td>menej</td>
<td>88</td>
<td>30</td>
<td>8</td>
<td>5.97</td>
<td>attraction</td>
<td>0.64 n.s.</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>menej</td>
<td>110</td>
<td>30</td>
<td>9</td>
<td>7.47</td>
<td>attraction</td>
<td>0.50 n.s.</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>lepšie</td>
<td>110</td>
<td>97</td>
<td>3</td>
<td>21.14</td>
<td>repulsion</td>
<td>9.70 ***</td>
<td></td>
</tr>
<tr>
<td>skôr</td>
<td>viac</td>
<td>52</td>
<td>108</td>
<td>2</td>
<td>12.71</td>
<td>repulsion</td>
<td>4.47 ***</td>
<td></td>
</tr>
<tr>
<td>viac</td>
<td>lepšie</td>
<td>88</td>
<td>97</td>
<td>10</td>
<td>19.31</td>
<td>repulsion</td>
<td>2.40 **</td>
<td></td>
</tr>
<tr>
<td>dlhšie</td>
<td>lepšie</td>
<td>35</td>
<td>97</td>
<td>2</td>
<td>7.68</td>
<td>repulsion</td>
<td>2.10 **</td>
<td></td>
</tr>
<tr>
<td>dlhšie</td>
<td>viac</td>
<td>35</td>
<td>108</td>
<td>3</td>
<td>8.55</td>
<td>repulsion</td>
<td>1.88 *</td>
<td></td>
</tr>
<tr>
<td>viac</td>
<td>váčšmi</td>
<td>88</td>
<td>41</td>
<td>4</td>
<td>8.16</td>
<td>repulsion</td>
<td>1.23 n.s.</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>váčšmi</td>
<td>110</td>
<td>41</td>
<td>9</td>
<td>10.20</td>
<td>repulsion</td>
<td>0.39 n.s.</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Covarying-collexeme analysis results of Lexical Filler, Slovak Web 2011 (expected frequency ≥5)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ďalej</td>
<td>viac</td>
<td>139</td>
<td>53</td>
<td>38</td>
<td>21.48</td>
<td>attraction</td>
<td>6.21 ***</td>
<td></td>
</tr>
<tr>
<td>skôr</td>
<td>lepšie</td>
<td>41</td>
<td>101</td>
<td>26</td>
<td>12.07</td>
<td>attraction</td>
<td>5.83 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>menej</td>
<td>139</td>
<td>22</td>
<td>18</td>
<td>8.91</td>
<td>attraction</td>
<td>4.26 ***</td>
<td></td>
</tr>
<tr>
<td>menej</td>
<td>lepšie</td>
<td>18</td>
<td>101</td>
<td>11</td>
<td>5.3</td>
<td>attraction</td>
<td>2.38 **</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>horšie</td>
<td>139</td>
<td>13</td>
<td>7</td>
<td>5.27</td>
<td>attraction</td>
<td>0.63 n.s.</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>lepšie</td>
<td>139</td>
<td>101</td>
<td>11</td>
<td>40.93</td>
<td>repulsion</td>
<td>13.57 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej</td>
<td>skôr</td>
<td>139</td>
<td>19</td>
<td>0</td>
<td>7.7</td>
<td>repulsion</td>
<td>4.44 ***</td>
<td></td>
</tr>
<tr>
<td>skôr</td>
<td>viac</td>
<td>41</td>
<td>53</td>
<td>1</td>
<td>6.34</td>
<td>repulsion</td>
<td>2.21 **</td>
<td></td>
</tr>
<tr>
<td>dlhšie</td>
<td>lepšie</td>
<td>17</td>
<td>101</td>
<td>5</td>
<td>5.01</td>
<td>repulsion</td>
<td>0.21 n.s.</td>
<td></td>
</tr>
</tbody>
</table>

71 unique Lexical Fillers in C1 and 106 unique Lexical Fillers in C2 were found in the SNC data set, out of a total of 442 C1C2 tokens. Similarly, in the Slovak Web 2011 data set, there were 72 unique Lexical Fillers in C1 and 81 unique Lexical Fillers in C2, out of 343 C1C2 tokens. These numbers underscore the high degree of variability and productivity of the Lexical Filler slot in the Slovak CC construction, mirroring findings from previous corpus studies that investigated the English CC construction (cf. Hoffmann et al. 2019, p. 15).

Nevertheless, by means of covarying-collexeme analysis, three significantly associated Lexical Filler pairs could be found in the SNC data, and four in the Slovak Web 2011 data:
SNC and Slovak Web 2011: $skór_{c1}$–$lepšie_{c2}$ (12), $d'alej_{c1}$–$viac_{c2}$ (13)
SNC only: $viac_{c1}$–$viac_{c2}$ (14)
Slovak Web 2011 only: $d'alej_{c1}$–$menej_{c2}$ (15), $menej_{c1}$–$lepšie_{c2}$ (16)

(12) $[čím \ skór \ to \ urobiš,]_{c1} [tým \ lepšie \ pre \ teba,]_{c2}$
   ‘The sooner you do this, the better for you’
   <SNC prim-7.0-public-all Source not given>

(13) $A [čím \ d'alej \ ho \ pozerám,]_{c1} [tým \ sa \ mi \ páči \ viac,]_{c2}$
   ‘And the more I look at him, the more I like him.’
   <SNC prim-7.0-public-all BLOG>

(14) $A [čím \ viac \ sa \ Boh \ vzďaloval,]_{c1} [tým \ viac \ si \ proroci \ začínali \ klást' \ otázky]_{c2}$
   ‘And the more God retreated, the more the prophets posed questions’
   <SNC prim-7.0-public-all PYan1>

(15) $Nie \ je \ nás \ veľa, [čím \ d'alej,]_{c1} [tým \ menej,]_{c2}$
   ‘There aren’t many of us, the more, the less.’
   <skTenTen11 token #29079716 doc #63980>

(16) $Najpríjemnejšie \ sa \ cítim \ v \ bavlnených \ veciach \ a [čím \ menej \ ich \ na \ sebe \ mám,]_{c1} [tým \ lepšie,]_{c2}$
   ‘I feel the most comfortable in cotton things and the less of them I have on myself, the better.’
   <skTenTen token #6096469 doc # 12594>

That is, despite the many choices that speakers have to ‘fill’ the **Lexical Filler** slots in C1 and C2 (evident from the many unique **Lexical Fillers**, see above), the above pairings were used significantly more often than could be expected by chance. The pair $skór_{c1}$–$lepšie_{c2}$, for example, is attracted with a collostructional strength of 13.37 in the SNC data, and 6.79 in the Slovak Web 2011 data, which translates to $p$-values of less than 0.0000000000001. In other words, there is overwhelming evidence against the null hypothesis that posits an “independence of occurrence” (Schmid – Küchenhoff 2013, p. 535) of $skór$ in C1 and $lepšie$ in C2. The same applies to $d'alej_{c1}$–$viac_{c2}$ (coll. strength 6.21*** in SNC and 6.75*** in Slovak Web 2011) and the other cross-clausal associations listed above.

Note that some pairs showed attraction, e.g. $viac_{c1}$–$menej_{c2}$ (coll. strength 0.64 n.s.) but did not cross the significance threshold of 1.301, while other pairs, e.g. $d'alej_{c1}$–$lepšie_{c2}$ (coll. strength 9.70*** were significantly repulsed, that is, their observed frequency was lower than their expected frequency.

In Usage-based CxG terms, these results constitute evidence for the entrenchment of at least five so-called meso-constructions, that is, five distinct form-meaning pairings stored in the constructional network. In line with Goldberg’s definition, the
covarying-collexeme analysis provides evidence that these five cross-clausal associations “occur with sufficient frequency” (2006, p. 5) to be considered entrenched.

On a final note, as I have mentioned previously, it seems likely that “čím ďalej, tým viac lit. ‘the more, the more’ might be treated as a separate construction, since it appears to have become a fixed idiomatic expression […] that has lost the original compositional meaning of the CC pattern” (Horsch 2021: 200). This is supported by the covarying-collexeme analysis of the interaction Lexical Filler:Verb Presence (see section 3.4.2), which indicates that, at least in the Slovak Web 2011 data set, the lexical pair *ďalej*<sub>C1</sub>-*viac*<sub>C2</sub> is strongly tied to CCs that feature No Verb in C1 and C2. In other words, *ďalej*<sub>No Verb</sub>-*viac*<sub>No Verb</sub> appears to be a prototype for the more abstract No Verb<sub>C1</sub>-No Verb<sub>C2</sub> meso-construction. Note, however, that in the SNC data the picture is not so clear: While it did reveal a *ďalej*<sub>C1</sub>-*viac*<sub>C2</sub> Lexical Filler meso-construction, in this case testing for interactions showed the pairing to be at least partly tied to the No Verb<sub>C1</sub>-Full Clause<sub>C2</sub> meso-construction.

### 3.2 Filler Type

Table 7 shows the results of the covarying-collexeme analysis of Filler Types in the SNC data:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AdjP</td>
<td>AdjP</td>
<td>13</td>
<td>46</td>
<td>3</td>
<td>1.35</td>
<td>attraction</td>
<td>0.84</td>
<td>n.s.</td>
</tr>
<tr>
<td>AdvP</td>
<td>AdvP</td>
<td>416</td>
<td>374</td>
<td>354</td>
<td>352</td>
<td>attraction</td>
<td>0.71</td>
<td>n.s.</td>
</tr>
<tr>
<td>NP</td>
<td>AdjP</td>
<td>13</td>
<td>46</td>
<td>2</td>
<td>1.35</td>
<td>attraction</td>
<td>0.4</td>
<td>n.s.</td>
</tr>
<tr>
<td>AdvP</td>
<td>NP</td>
<td>13</td>
<td>22</td>
<td>1</td>
<td>0.65</td>
<td>attraction</td>
<td>0.31</td>
<td>n.s.</td>
</tr>
<tr>
<td>NP</td>
<td>AdvP</td>
<td>416</td>
<td>22</td>
<td>21</td>
<td>20.7</td>
<td>attraction</td>
<td>0.21</td>
<td>n.s.</td>
</tr>
<tr>
<td>NP</td>
<td>AdvP</td>
<td>13</td>
<td>374</td>
<td>11</td>
<td>1</td>
<td>chance</td>
<td>0.21</td>
<td>n.s.</td>
</tr>
<tr>
<td>AdvP</td>
<td>AdjP</td>
<td>416</td>
<td>46</td>
<td>41</td>
<td>43.29</td>
<td>repulsion</td>
<td>0.92</td>
<td>n.s.</td>
</tr>
<tr>
<td>AdjP</td>
<td>AdvP</td>
<td>13</td>
<td>374</td>
<td>9</td>
<td>11</td>
<td>repulsion</td>
<td>0.91</td>
<td>n.s.</td>
</tr>
<tr>
<td>NP</td>
<td>NP</td>
<td>13</td>
<td>22</td>
<td>0</td>
<td>0.65</td>
<td>repulsion</td>
<td>0.29</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Interestingly, there were no significant results in the SNC data, suggesting that “there are no meso-constructions with regard to Filler Types” (Horsch 2021, p. 211).

Table 8 shows the results of the covarying-collexeme analysis of Filler Types in the Slovak Web 2011 data:
Table 8: Covarying-collexeme analysis results of Filler Type, Slovak Web 2011 (expected frequency ≥5, only significant results)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AdvP</td>
<td>AdvP</td>
<td>294</td>
<td>277</td>
<td>247</td>
<td>237.43</td>
<td>attraction</td>
<td>3.36 **</td>
<td>***</td>
</tr>
<tr>
<td>AdvP</td>
<td>NP</td>
<td>294</td>
<td>28</td>
<td>19</td>
<td>24</td>
<td>repulsion</td>
<td>2.01 **</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>AdvP</td>
<td>23</td>
<td>277</td>
<td>14</td>
<td>18.57</td>
<td>repulsion</td>
<td>1.75 *</td>
<td></td>
</tr>
<tr>
<td>AdvP</td>
<td>AdjP</td>
<td>294</td>
<td>37</td>
<td>27</td>
<td>31.71</td>
<td>repulsion</td>
<td>1.63 *</td>
<td></td>
</tr>
<tr>
<td>AdjP</td>
<td>AdvP</td>
<td>18</td>
<td>277</td>
<td>11</td>
<td>14.54</td>
<td>repulsion</td>
<td>1.42 *</td>
<td></td>
</tr>
</tbody>
</table>

In the case of the Slovak Web 2011 data, covarying-collexeme analysis turned up evidence for the existence of one meso-construction: AdvP_{C1}-AdvP_{C2} (17).

(17) [Veď čím [neskôr]_{AdvP} pride záchranka,]_{C1} [tým [horšie]_{AdvP} pre zranených.]_{C2}
‘Indeed, the later the ambulance arrives, the worse for the injured.’

3.3 Verb Presence

Table 9 shows the results of the covarying-collexeme analysis of Verb Presence in SNC; Table 10 the results for the Slovak Web 2011 data:

Table 9: Covarying-collexeme analysis results of Verb Presence, SNC (adapted from Horsch 2021, p. 214)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Full clause</td>
<td>Full clause</td>
<td>294</td>
<td>315</td>
<td>252</td>
<td>209.52</td>
<td>attraction</td>
<td>19.84 ***</td>
<td></td>
</tr>
<tr>
<td>No verb</td>
<td>No verb</td>
<td>148</td>
<td>127</td>
<td>85</td>
<td>42.52</td>
<td>attraction</td>
<td>19.84 ***</td>
<td></td>
</tr>
<tr>
<td>Full clause</td>
<td>No verb</td>
<td>294</td>
<td>127</td>
<td>42</td>
<td>84.47</td>
<td>repulsion</td>
<td>19.84 ***</td>
<td></td>
</tr>
<tr>
<td>No verb</td>
<td>Full clause</td>
<td>148</td>
<td>315</td>
<td>63</td>
<td>105.47</td>
<td>repulsion</td>
<td>19.84 ***</td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Covarying-collexeme analysis results of Verb Presence, Slovak Web 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Full clause</td>
<td>Full clause</td>
<td>138</td>
<td>130</td>
<td>107</td>
<td>52.3</td>
<td>attraction</td>
<td>36.44 ***</td>
<td></td>
</tr>
<tr>
<td>No verb</td>
<td>No verb</td>
<td>205</td>
<td>213</td>
<td>182</td>
<td>127.3</td>
<td>attraction</td>
<td>36.44 ***</td>
<td></td>
</tr>
<tr>
<td>Full clause</td>
<td>No verb</td>
<td>138</td>
<td>213</td>
<td>31</td>
<td>85.7</td>
<td>repulsion</td>
<td>36.44 ***</td>
<td></td>
</tr>
<tr>
<td>No verb</td>
<td>Full clause</td>
<td>205</td>
<td>130</td>
<td>23</td>
<td>77.7</td>
<td>repulsion</td>
<td>36.44 ***</td>
<td></td>
</tr>
</tbody>
</table>
Notably, the results are exactly the same in both corpora, indicating the existence of two meso-constructions: Full Clause\textsubscript{C1}-Full Clause\textsubscript{C2} (18) and No Verb\textsubscript{C1}-No Verb\textsubscript{C2} (19), with collostructional strengths of 19.84*** (SNC) and 36.44*** (Slovak Web 2011).

(18) [čím väčším [o tom Tchančchun rozmýšľala]\textsubscript{Full Clause}\textsubscript{C1} [tým väčším [ju to rozčuľovalo]\textsubscript{Full Clause}\textsubscript{C2} ‘The more Tchančchun thought about this, the more it annoyed her.’ <SNC: prim-7.0-public-all CSüe2>

(19) Potrebujeme sa vrátiť domov, [čím skôr]\textsubscript{C1} [tým lepšie.]\textsubscript{C2} ‘We have to return home, the sooner, the better.’ <SNC: prim-7.0-public-all MBal18>

3.4 Interactions

3.4.1 Lexical Filler:Filler Type

Only the Slovak Web 2011 data set revealed a significant Filler Type association; AdvP\textsubscript{C1}-AdvP\textsubscript{C2}. No significant cross-clausal Filler Type associations emerged from the covarying-collexeme analysis of the SNC data (cf. section 3.2).

Thus, testing for interactions to determine whether or not certain Lexical Fillers influenced the outcome of the test for Filler Type associations (as Hoffmann et al. 2019 did for their COCA data) was only necessary for the Slovak Web 2011 data.

Table 11 shows the results of the covarying-collexeme analysis of the Lexical Filler:Filler Type interaction in the Slovak Web 2011 data set.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ďalej_AdVP</td>
<td>viac_AdVP</td>
<td>139</td>
<td>53</td>
<td>38</td>
<td>21.48</td>
<td>attraction</td>
<td>6.21 ***</td>
<td></td>
</tr>
<tr>
<td>skôr_AdVP</td>
<td>lepšie_AdVP</td>
<td>41</td>
<td>91</td>
<td>25</td>
<td>10.88</td>
<td>attraction</td>
<td>6.17 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej_AdVP</td>
<td>menej_AdVP</td>
<td>139</td>
<td>22</td>
<td>18</td>
<td>8.91</td>
<td>attraction</td>
<td>4.26 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej_AdVP</td>
<td>lepšie_AdVP</td>
<td>139</td>
<td>91</td>
<td>10</td>
<td>36.88</td>
<td>repulsion</td>
<td>11.81 ***</td>
<td></td>
</tr>
<tr>
<td>skôr_AdVP</td>
<td>viac_AdVP</td>
<td>139</td>
<td>19</td>
<td>0</td>
<td>7.7</td>
<td>repulsion</td>
<td>4.44 ***</td>
<td></td>
</tr>
</tbody>
</table>

As the results show, there were three statistically significant cross-clausal associations that constitute evidence for the following meso-constructions: ďalej_AdVP\textsubscript{C1}-viac_AdVP\textsubscript{C2} (20), skôr_AdVP\textsubscript{C1}-lepšie_AdVP\textsubscript{C2} (21), and ďalej_AdVP\textsubscript{C1}-menej_AdVP\textsubscript{C2} (22).
A ona sa červenala [čím [d’alej]\_\text{AdvP} –]_{C1} [tým [viac]\_\text{AdvP}]_{C2}
‘And she turned red, the more, the more.’

Platí však: [čím [skôr]\_\text{AdvP}]_{C1} [tým [lepšie!]\_\text{AdvP}]_{C2}
‘However, the following applies: the sooner, the better!’

Nie je nás veľa, [čím [d’alej]\_\text{AdvP}]_{C1} [tým [menej]\_\text{AdvP}]_{C2}
‘There aren’t many of us, the more, the less.’

These results suggest that the Lexical Filler pairs which covarying-collexeme analysis found to be significant in the Slovak Web 2011 data set (see table 6 in section 3.1) are indeed “strongly tied to specific phrasal categories” (Hoffmann et al. 2019, p. 16). For example, as table 11 shows, the most frequent pair is d’alej\_\text{AdvP}-viac\_\text{AdvP} (20) with 38 tokens. This means that the pair “functions as a strong prototype for the abstract AdvP-AdvP meso-construction” (Hoffmann et al. 2019, p. 16) that was determined by covarying-collexeme analysis (see table 8 in section 3.2). In other words, there is an indication that the AdvP\_\text{C1}-AdvP\_\text{C2} meso-construction can in part be explained by lexical effects. However, also note that d’alej\_\text{AdvP}-viac\_\text{AdvP} (38 tokens) and d’alej\_\text{AdvP}-menej\_\text{AdvP} (18 tokens) make up less than 23% of the total amount of AdvP\_\text{C1}-AdvP\_\text{C2} tokens (247), which indicates that this meso-construction cannot be fully accounted for with lexical effects.

3.4.2 Lexical Filler: Verb Presence
Next, the interaction Lexical Filler: Verb Presence was tested using covarying-collexeme analysis. The results are shown in tables 12 and 13.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>skôr_\text{FULL CLAUSE}</td>
<td>lepšie_\text{NO VERB}</td>
<td>38</td>
<td>80</td>
<td>21</td>
<td>6.88</td>
<td>attraction</td>
<td>7.25</td>
<td>***</td>
</tr>
<tr>
<td>viac_\text{FULL CLAUSE}</td>
<td>viac_\text{FULL CLAUSE}</td>
<td>82</td>
<td>90</td>
<td>28</td>
<td>16.7</td>
<td>attraction</td>
<td>3.08</td>
<td>***</td>
</tr>
<tr>
<td>d’alej_\text{NO VERB}</td>
<td>viac_\text{FULL CLAUSE}</td>
<td>97</td>
<td>90</td>
<td>29</td>
<td>19.75</td>
<td>attraction</td>
<td>2.12</td>
<td>**</td>
</tr>
<tr>
<td>d’alej_\text{NO VERB}</td>
<td>lepšie_\text{NO VERB}</td>
<td>97</td>
<td>80</td>
<td>2</td>
<td>17.56</td>
<td>repulsion</td>
<td>6.98</td>
<td>***</td>
</tr>
<tr>
<td>viac_\text{FULL CLAUSE}</td>
<td>lepšie_\text{NO VERB}</td>
<td>82</td>
<td>80</td>
<td>4</td>
<td>14.84</td>
<td>repulsion</td>
<td>3.85</td>
<td>***</td>
</tr>
<tr>
<td>dlhšie_\text{FULL CLAUSE}</td>
<td>lepšie_\text{NO VERB}</td>
<td>34</td>
<td>80</td>
<td>0</td>
<td>6.15</td>
<td>repulsion</td>
<td>3.08</td>
<td>***</td>
</tr>
<tr>
<td>skôr_\text{FULL CLAUSE}</td>
<td>viac_\text{FULL CLAUSE}</td>
<td>38</td>
<td>90</td>
<td>2</td>
<td>7.74</td>
<td>repulsion</td>
<td>2.11</td>
<td>**</td>
</tr>
</tbody>
</table>
Table 13: Covarying-collexeme analysis results of the interaction LEXICAL FILLER:VERB PRESENCE, Slovak Web 2011 (expected frequency ≥5, significant results only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ďalej No Verb</td>
<td>viac No Verb</td>
<td>132</td>
<td>29</td>
<td>28</td>
<td>11.16</td>
<td>attraction</td>
<td>11.17 ***</td>
<td></td>
</tr>
<tr>
<td>skôr No Verb</td>
<td>lepšie No Verb</td>
<td>20</td>
<td>90</td>
<td>17</td>
<td>5.25</td>
<td>attraction</td>
<td>7.66 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej No Verb</td>
<td>menej No Verb</td>
<td>132</td>
<td>15</td>
<td>15</td>
<td>5.77</td>
<td>attraction</td>
<td>6.45 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej No Verb</td>
<td>skôr Full Clause</td>
<td>132</td>
<td>19</td>
<td>0</td>
<td>7.31</td>
<td>repulsion</td>
<td>10.3 ***</td>
<td></td>
</tr>
<tr>
<td>ďalej No Verb</td>
<td>viac Full Clause</td>
<td>132</td>
<td>29</td>
<td>28</td>
<td>11.16</td>
<td>attraction</td>
<td>11.17 ***</td>
<td></td>
</tr>
</tbody>
</table>

Again, several cross-clausal associations emerge as statistically significant: In the SNC dataset, these are skôr Full Clause-C1-lepšie No Verb-C2 (23) (coll. strength 7.25***), viac Full Clause-C1-viac Full Clause-C2 (24) (coll. strength 3.08***), and ďalej No Verb-C1-viac Full Clause-C2 (25) (coll. strength 2.12**).

  ‘And the sooner we change this, the better.’
  <SNC prim-7.0-public-all CCook1>

(24) Každý databázový server miluje pamäť, a
  ‘Every database server loves memory, and the more of it that it has, the more it pays back in terms of its power.’
  <SNC prim-7.0-public-all ITN>

  ‘And the further towards the west, the more this is a completely standard solution.’
  <SNC prim-7.0-public-all HN2006/07>

In the Slovak Web 2011 data, the following cross-clausal associations could be found using covarying-collexeme analysis: dʼalej No Verb-C1-viac No Verb-C2 (26) (coll.strength 11.17***), skôr No Verb-C1-lepšie No Verb-C2 (27) (coll. strength 7.66***), and dʼalej No Verb-C1-menej No Verb-C2 (28) (coll. strength 6.45***).

(26) A ona sa červenala [čím dʼalej –]C1 [tým viac.]C2
  ‘And she turned red, the more, the more.’
  <skTenTen token #7563610 doc #15759>

(27) Platí však: [čím skôr]C1 [tým lepšie!]C2
  ‘However, the following applies: the sooner, the better!’
  <skTenTen token #4500805 doc #9376>

(28) A ovocia je [čím dʼalej]C1 [tým menej.]C2
  ‘And regarding fruits, there are the more, the less.’
  <skTenTen token #17867896 doc #38629>
3.4.3 Filler Type: Verb Presence

The third interaction to be tested was Filler Type: Verb Presence. Table 14 shows the results from SNC; table 15 the results from Slovak Web 2011:

Table 14: Covarying-collexeme analysis results of the interaction Filler Type: Verb Presence, SNC (expected frequency ≥5, significant results only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADVP_NO_VERB</td>
<td>145</td>
<td>118</td>
<td>76</td>
<td>38.71</td>
<td>attraction</td>
<td>16.16***</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>ADVP_FULL_CLAUSE</td>
<td>271</td>
<td>256</td>
<td>185</td>
<td>156.96</td>
<td>attraction</td>
<td>7.6***</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>ADJP_FULL_CLAUSE</td>
<td>271</td>
<td>41</td>
<td>32</td>
<td>25.14</td>
<td>attraction</td>
<td>1.85*</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>ADVP_NO_VERB</td>
<td>271</td>
<td>118</td>
<td>39</td>
<td>72.35</td>
<td>repulsion</td>
<td>12.54***</td>
<td></td>
</tr>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADVP_FULL_CLAUSE</td>
<td>145</td>
<td>256</td>
<td>54</td>
<td>83.98</td>
<td>repulsion</td>
<td>9.14***</td>
<td></td>
</tr>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADJP_FULL_CLAUSE</td>
<td>145</td>
<td>41</td>
<td>4</td>
<td>13.45</td>
<td>repulsion</td>
<td>3.44***</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Covarying-collexeme analysis results of the interaction Filler Type: Verb Presence, Slovak Web 2011 (expected frequency ≥5, significant results only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADVP_NO_VERB</td>
<td>179</td>
<td>168</td>
<td>131</td>
<td>87.67</td>
<td>attraction</td>
<td>20.88***</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>ADVP_FULL_CLAUSE</td>
<td>115</td>
<td>109</td>
<td>75</td>
<td>36.55</td>
<td>attraction</td>
<td>20.02***</td>
<td></td>
</tr>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADVP_FULL_CLAUSE</td>
<td>179</td>
<td>109</td>
<td>17</td>
<td>56.88</td>
<td>repulsion</td>
<td>20.72***</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>ADVP_NO_VERB</td>
<td>115</td>
<td>168</td>
<td>24</td>
<td>56.33</td>
<td>repulsion</td>
<td>13.34***</td>
<td></td>
</tr>
<tr>
<td>ADVP_NO_VERB</td>
<td>ADJP_FULL_CLAUSE</td>
<td>179</td>
<td>12</td>
<td>0</td>
<td>6.26</td>
<td>repulsion</td>
<td>3.94***</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>NO_VERB_NP</td>
<td>115</td>
<td>19</td>
<td>1</td>
<td>6.37</td>
<td>repulsion</td>
<td>2.43**</td>
<td></td>
</tr>
<tr>
<td>ADVP_FULL_CLAUSE</td>
<td>NO_VERB_AdJP</td>
<td>115</td>
<td>25</td>
<td>3</td>
<td>8.38</td>
<td>repulsion</td>
<td>1.94*</td>
<td></td>
</tr>
</tbody>
</table>

Here, three more cross-clausal associations were shown to be statistically significant, providing evidence for the existence of the following meso-constructions: ADVP_NO_VERB<sub>C1</sub>-ADVP_NO_VERB<sub>C2</sub> (29) (coll. strength 16.16*** in SNC, 20.88*** in Slovak Web 2011), ADVP_FULL_CLAUSE<sub>C1</sub>-ADVP_FULL_CLAUSE<sub>C2</sub> (30) (coll. strength 7.6*** in SNC, 20.02*** in Slovak Web 2011), and ADVP_FULL_CLAUSE<sub>C1</sub>-ADJP_FULL_CLAUSE<sub>C2</sub> (31) (coll. strength 1.85*, in SNC only).

(29) Poznáme tú politiku, [čím | horšie]<sub>ADV</sub>[<sub>C1</sub>] [<sub>tým | lepšie.</sub><sub>ADV</sub>[<sub>C2</sub>]
‘We know this (kind of) politics, the worse, the better.’
<SNC prim-7.0-public-all HN2010/05>

(30) Bol som presvedčený, že
[čím | d’alej]<sub>ADV</sub> [v turnaji pôjdeme]<sub>FULL</sub>[<sub>C1</sub>] [<sub>tým | lepšie</sub><sub>ADV</sub> [budeme hrať.]
‘I was sure that the further we will progress in the tournament, the better we will play.’
<SNC prim-7.0-public-all MYTP2011/13>
‘It is said that the more the politicians devour, the hungrier the nation is.’

3.4.4 Lexical Filler:Filler Type:Verb Presence

Finally, the three-way interaction Lexical Filler:Filler Type:Verb Presence was tested using covarying-collexeme analysis. The results are presented in tables 16 and 17:

Table 16: Covarying-collexeme analysis results of the interaction Lexical Filler:Filler Type:Verb Presence, SNC (expected frequency ≥5, significant results only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>skôr _AdvP_Full Clause</td>
<td>lepsie _AdvP_No Verb</td>
<td>38</td>
<td>80</td>
<td>21</td>
<td>6.88</td>
<td>attraction</td>
<td>7.25 ***</td>
<td></td>
</tr>
<tr>
<td>viac _AdvP_Full Clause</td>
<td>viac _AdvP_Full Clause</td>
<td>81</td>
<td>90</td>
<td>28</td>
<td>16.49</td>
<td>attraction</td>
<td>3.19 ***</td>
<td></td>
</tr>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>viac _AdvP_Full Clause</td>
<td>97</td>
<td>90</td>
<td>29</td>
<td>19.75</td>
<td>attraction</td>
<td>2.12 **</td>
<td></td>
</tr>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>lepsie _AdvP_No Verb</td>
<td>97</td>
<td>80</td>
<td>2</td>
<td>17.56</td>
<td>repulsion</td>
<td>6.98 ***</td>
<td></td>
</tr>
<tr>
<td>viac _AdvP_Full Clause</td>
<td>lepsie _AdvP_No Verb</td>
<td>81</td>
<td>80</td>
<td>3</td>
<td>14.66</td>
<td>repulsion</td>
<td>4.51 ***</td>
<td></td>
</tr>
<tr>
<td>dlhšie _AdvP_Full Clause</td>
<td>lepsie _AdvP_No Verb</td>
<td>33</td>
<td>80</td>
<td>0</td>
<td>5.97</td>
<td>repulsion</td>
<td>2.98 **</td>
<td></td>
</tr>
<tr>
<td>skôr _AdvP_Full Clause</td>
<td>viac _AdvP_Full Clause</td>
<td>38</td>
<td>90</td>
<td>2</td>
<td>7.74</td>
<td>repulsion</td>
<td>2.11 **</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Covarying-collexeme analysis results of the interaction Lexical Filler:Filler Type:Verb Presence, Slovak Web 2011 (expected frequency ≥5, significant results only)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>viac _AdvP_No Verb</td>
<td>132</td>
<td>29</td>
<td>28</td>
<td>11.16</td>
<td>attraction</td>
<td>11.17 ***</td>
<td></td>
</tr>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>menej _AdvP_No Verb</td>
<td>132</td>
<td>15</td>
<td>15</td>
<td>5.77</td>
<td>attraction</td>
<td>6.45 ***</td>
<td></td>
</tr>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>lepsie _AdvP_No Verb</td>
<td>132</td>
<td>82</td>
<td>10</td>
<td>31.56</td>
<td>repulsion</td>
<td>8.49 ***</td>
<td></td>
</tr>
<tr>
<td>dalej _AdvP_No Verb</td>
<td>skôr _AdvP_Full Clause</td>
<td>132</td>
<td>19</td>
<td>0</td>
<td>7.31</td>
<td>repulsion</td>
<td>4.15 ***</td>
<td></td>
</tr>
</tbody>
</table>

As tables 16 and 17 show, three significant cross-clausal associations were found in the SNC data set and two in the Slovak Web 2011 data set, revealing further nodes in the Slovak CC constructional network:

- `skôr _AdvP_Full Clause`-`lepsie _AdvP_No Verb` (32) (coll. strength 7.25***)
- `viac _AdvP_Full Clause`-`viac _AdvP_Full Clause` (33) (coll. strength 3.19***)
- `dalej _AdvP_No Verb`-`viac _AdvP_Full Clause` (34) (coll. strength 2.12**)
- `dalej _AdvP_No Verb`-`viac _AdvP_No Verb` (35) (coll. strength 11.17***)
- `dalej _AdvP_No Verb`-`menej _AdvP_No Verb` (36) (coll. strength 6.45***)

728
(32) **Poľská národná banka sa nazdáva, že**

\[ \text{čím [skôr]_{ADV-P} [Poľsko vstúpi do Európskej menovej unie]_{FULL CLAUSE, C2} [ťým [lepšie]_{ADV-P}}_{C2} \]

‘Poland’s national bank reckons that the sooner Poland enters the European monetary union, the better.’

<SNC prim-7.0-public-all HN2004/09>

(33) **[čím [viac]_{ADV-P} [sme sa blížili k Dakaru]_{FULL CLAUSE, C1} [ťým [viac]_{ADV-P} [rástol počet dedín.]_{FULL CLAUSE, C2}**

‘the closer we got to Dakar, the higher the number of villages grew.’

<SNC prim-7.0-public-all DŠ2007/01>

(34) **Mimochodom, [čím [d’alej]_{ADV-P, C1} [ťým [viac]_{ADV-P} [sa mi tá robota nepáči]_{FULL CLAUSE, C2}**

‘Incidentally, the more [I worked], the more I didn’t like this work.

<SNC prim-7.0-public-all BIRDZ2013>

(35) **Míľom psov [čím [d’alej]_{ADV-P, C1} [ťým [viac]_{ADV-P}**

‘I love dogs, the more [of them], the more [I love them].’

<skTenTen token #8703081 doc #18226>

(36) **Lenže [čím [d’alej]_{ADV-P, C1} [ťým [menej]_{ADV-P, C2}**

‘But the more, the more.’

<skTenTen token #5311696 doc #10979>

4. **TOWARDS A VISUALIZATION OF THE SLOVAK CC MESO-CONSTRUCTIONAL NETWORK**

By means of covarying-collexeme analysis, a total of 22 cross-clausal associations and thus, meso-constructions, could be determined in the SNC and Slovak Web 2011 data sets. By testing the data for interactions, it was possible to obtain an even more fine-grained picture of the constructional network underlying the Slovak CC construction than before (cf. Horsch 2021). Before attempting to visualize this network, in the following I will provide an overview of the constructions\(^{18}\) involved. Recall that (meso-)constructions are arbitrary and conventional form-meaning pairings that can be partly substantive and partly schematic.

I will start with the maximally abstract Slovak CC construction (37), whose form side is based on Culicover and Jackendoff’s (1999, p. 567) English CC template (10) and whose meaning side is based on Hoffmann’s (2017, p. 351) English CC paraphrase\(^{19}\):

---

\(^{18}\) Meso-constructions (38), (44) and (45) are taken from Horsch (2021, p. 216-218).

\(^{19}\) Note that sometimes, the meaning encoded by CCs is heavily dependent on context, which cannot be captured entirely by the form-meaning pairings presented here, which are essentially an oversim-
(37) Slovak CC construction

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANING:</td>
<td>‘[As the degree of comparative element 1 increases/decreases with respect to clause 1/ with respect to a contextually retrievable clause] independent variable → [so the degree of comparative element 2 increases/decreases with respect to clause 2/ with respect to a contextually retrievable clause] dependent variable in a monotonic way’</td>
</tr>
</tbody>
</table>

The maximally abstract CC construction in (37) licenses the following less abstract meso-constructions (38-59):

**LEXICAL FILLER meso-constructions**

(38) **skôrC1-lepšieC2** CC meso-construction (SNC, Slovak Web 2011)

<table>
<thead>
<tr>
<th>FORM:</th>
<th>[[tʃi:m] [skôr] comparative element1 (…) optional clause 1]C1 [[tʃi:m] [lepšie] comparative element2 (…) optional clause 2]C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANING:</td>
<td>‘[As the point in time (with respect to clause 1) decreases] independent variable → [so the degree of quality increases (with respect to clause 2)] dependent variable in a monotonic way’</td>
</tr>
</tbody>
</table>

(39) **ďalejC1-viacC2** CC meso-construction (SNC, Slovak Web 2011)

<table>
<thead>
<tr>
<th>FORM:</th>
<th>[[tʃi:m] [ďalej] comparative element1 (…) optional clause 1]C1 [[tʃi:m] [viac] comparative element2 (…) optional clause 2]C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANING:</td>
<td>‘[As the point in time (with respect to clause 1) increases] independent variable → [so the quantity increases (with respect to clause 2)] dependent variable in a monotonic way’</td>
</tr>
</tbody>
</table>

(40) **viacC1-viacC2** CC meso-construction (SNC)

<table>
<thead>
<tr>
<th>FORM:</th>
<th>[[tʃi:m] [viac] comparative element1 (…) optional clause 1]C1 [[tʃi:m] [viac] comparative element2 (…) optional clause 2]C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANING:</td>
<td>‘[As the quantity (with respect to clause 1) increases] independent variable → [so the quantity increases (with respect to clause 2)] dependent variable in a monotonic way’</td>
</tr>
</tbody>
</table>

(41) **ďalejC1-menejC2** CC meso-construction (Slovak Web 2011)

<table>
<thead>
<tr>
<th>FORM:</th>
<th>[[tʃi:m] [ďalej] comparative element1 (…) optional clause 1]C1 [[tʃi:m] [menej] comparative element2 (…) optional clause 2]C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEANING:</td>
<td>‘[As the point in time (with respect to clause 1) increases] independent variable → [so the quantity decreases (with respect to clause 2)] dependent variable in a monotonic way’</td>
</tr>
</tbody>
</table>

plification of a much more complex reality: As remarked by an anonymous reviewer in a previous article, in the case of No Verb constructions, “the degree [of the comparative elements] has to change with respect to something” that is “clear from the context” (i.e. that which is usually expressed by the obligatory clauses in the Full Clause construction). What actually changes is something “with respect to the propositional content of that clause” (anonymous reviewer in Horsch 2021, p. 218).
(42) menej$_{C1}$-lepšie$_{C2}$ CC meso-construction (Slovak Web 2011)

**Form:**

\[
[[tʃi:m][menej]_{\text{comparative element 1}}(...)_\text{optional clause }1]_{C1}
\]

\[
[[tʃi:m][lepšie]_{\text{comparative element 2}}(...)_\text{optional clause }2]_{C2}
\]

**Meaning:**

‘[As the quantity (with respect to clause 1) decreases] independent variable → [so the degree of quality increases (with respect to clause 2)] dependent variable in a monotonic way’

---

**Filler Type meso-constructions**

(43) AdvP$_{C1}$-AdvP$_{C2}$ CC meso-construction (Slovak Web 2011)

**Form:**

\[
[[tʃi:m][...](...)\text{optional clause 1}]_{C1}
\]

\[
[[tʃi:m][...](...)\text{optional clause 2}]_{C2}
\]

**Meaning:**

‘[As the degree of AdvP1 increases/decreases (with respect to clause 1)] independent variable → [so the degree of AdvP2 increases/decreases (with respect to clause 2)] dependent variable in a monotonic way’

---

**Verb Presence meso-constructions**

(44) No Verb$_{C1}$-No Verb$_{C2}$ CC meso-construction (SNC, Slovak Web 2011)

**Form:**

\[
[[tʃi:m][...]]_{\text{comparative element 1}}_{C1}
\]

\[
[[tʃi:m][...]]_{\text{comparative element 2}}_{C2}
\]

**Meaning:**

‘[As the degree of comparative element1 increases/decreases] independent variable → [so the degree of comparative element2 increases/decreases] dependent variable in a monotonic way’

---

(45) Full Clause$_{C1}$-Full Clause$_{C2}$ CC meso-construction (SNC, Slovak Web 2011)

**Form:**

\[
[[tʃi:m][...][...][obligatory clause 1]]_{C1}
\]

\[
[[tʃi:m][...][...][obligatory clause 2]]_{C2}
\]

**Meaning:**

‘[As the degree of comparative element1 increases/decreases with respect to obligatory clause 1] independent variable → [so the degree of comparative element2 increases/decreases with respect to obligatory clause 2] dependent variable in a monotonic way’

---

** Lexical Filler: Verb Presence meso-constructions**

(46) skôr Full Clause$_{C1}$-lepšie No Verb$_{C2}$ CC meso-construction (SNC)

**Form:**

\[
[[tʃi:m][skôr]]_{\text{comparative element 1}}_{C1}
\]

\[
[[tʃi:m][lepšie]]_{\text{comparative element 2}}_{C2}
\]

**Meaning:**

‘[As the point in time with respect to obligatory clause 1 decreases] independent variable → [so the degree of quality increases] dependent variable in a monotonic way’

---

(47) viac Full Clause$_{C1}$-viac Full Clause$_{C2}$ CC meso-construction (SNC)

**Form:**

\[
[[tʃi:m][viac]]_{\text{comparative element 1}}_{C1}
\]

\[
[[tʃi:m][viac]]_{\text{comparative element 2}}_{C2}
\]

**Meaning:**

‘[As the quantity with respect to obligatory clause 1 increases] independent variable → [so the quantity increases with respect to obligatory clause 2] dependent variable in a monotonic way’
(48) **d’alej** _No Verb_C1-**viac** _Full Clause_C2 CC meso-construction (SNC)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] [d’alej]\text{comparative element}_1]_{C1} \\
&[[tʃi:m] [viac]\text{comparative element}_2 […] \text{obligatory clause}_2]_{C2}
\end{align*}
\]

**Meaning:**

‘As the point in time increases\text{independent variable} → so the quantity increases with respect to obligatory clause 2\text{dependent variable in a monotonic way’}

(49) **d’alej** _No Verb_C1-**viac** _No Verb_C2 CC meso-construction (Slovak Web 2011)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] [d’alej]\text{comparative element}_1]_{C1} \\
&[[tʃi:m] [viac]\text{comparative element}_2]_{C2}
\end{align*}
\]

**Meaning:**

‘As the point in time increases\text{independent variable} → so the quantity increases\text{dependent variable in a monotonic way’}

(50) **skôr** _No Verb_C1-**lepšie** _No Verb_C2 CC meso-construction (Slovak Web 2011)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] [skôr]\text{comparative element}_1]_{C1} \\
&[[tʃi:m] [lepšie]\text{comparative element}_2]_{C2}
\end{align*}
\]

**Meaning:**

‘As the point in time decreases\text{independent variable} → so the degree of quality increases\text{dependent variable in a monotonic way’}

(51) **d’alej** _No Verb_C1-**menej** _No Verb_C2 CC meso-construction (Slovak Web 2011)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] [d’alej]\text{comparative element}_1]_{C1} \\
&[[tʃi:m] [menej]\text{comparative element}_2]_{C2}
\end{align*}
\]

**Meaning:**

‘As the point in time decreases\text{independent variable} → so the quantity decreases\text{dependent variable in a monotonic way’

**Filler Type:** **Verb Presence** meso-constructions

(52) **AdvP** _No Verb_C1-**AdvP** _No Verb_C2 CC meso-construction (SNC, Slovak Web 2011)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] […]_{AdvP1}]_{C1} \\
&[[tʃi:m] […]_{AdvP2}]_{C2}
\end{align*}
\]

**Meaning:**

‘As the degree of \text{AdvP1 increases/decreases}\text{independent variable} → so the degree of \text{AdvP2 increases/decreases}\text{dependent variable in a monotonic way’

(53) **AdvP** _Full Clause_C1-**AdvP** _Full Clause_C2 CC meso-construction (SNC, Slovak Web 2011)

**Form:** 
\[
\begin{align*}
&[[tʃi:m] […]_{AdvP1} […] \text{obligatory clause}_1]_{C1} \\
&[[tʃi:m] […]_{AdvP2} […] \text{obligatory clause}_2]_{C2}
\end{align*}
\]

**Meaning:**

‘As the degree of \text{AdvP1 increases/decreases with respect to obligatory clause 1}\text{independent variable} → so the degree of \text{AdvP2 increases/decreases with respect to obligatory clause 2}\text{dependent variable in a monotonic way’

732
(54) **ADVp**

**FORM:**

\[
[[\text{tši:m}] […] \text{ADVp}1 […] \text{obligatory clause 1}]C1
[[\text{tši:m}] […] \text{ADVp2} […] \text{obligatory clause 2}]C2
\]

**MEANING:**

‘[As the degree of \text{ADVp1} increases/decreases with respect to obligatory clause 1]\text{independent variable}

\[\rightarrow [\text{so the degree of} \text{ADVp2 increases/decreases with respect to obligatory clause 2}]\text{dependent variable}

\] in a monotonic way’

**LEXICAL FILLER:**

**FILLER TYPE:**

**VERB PRESENCE**

**meso-constructions**

(55) \text{skôr}_{-}\text{ADVp}_{-}\text{FULL CLAUSE}_{C1}-\text{lepšie}_{-}\text{ADVp}_{-}\text{No VERB}_{C2} **CC** meso-construction (SNC)

**FORM:**

\[
[[\text{tši:m}] \text{skôr} […] \text{ADVp}1 […] \text{obligatory clause 1}]C1
[[\text{tši:m}] \text{lepšie} […] \text{ADVp2}]C2
\]

**MEANING:**

‘[As the point in time decreases with respect to obligatory clause 1]\text{independent variable}

\[\rightarrow [\text{so the degree of quality increases}]\text{dependent variable}

\] in a monotonic way’

(56) \text{viac}_{-}\text{ADVp}_{-}\text{FULL CLAUSE}_{C1}-\text{viac}_{-}\text{ADVp}_{-}\text{FULL CLAUSE}_{C2} **CC** meso-construction (SNC)

**FORM:**

\[
[[\text{tši:m}] \text{viac} […] \text{ADVp}1 […] \text{obligatory clause 1}]C1
[[\text{tši:m}] \text{viac} […] \text{ADVp2} […] \text{obligatory clause 2}]C2
\]

**MEANING:**

‘[As the quantity with respect to obligatory clause 1 increases]\text{independent variable}

\[\rightarrow [\text{so the quantity increases with respect to obligatory clause 2}]\text{dependent variable}

\] in a monotonic way’

(57) \text{ďalej}_{-}\text{ADVp}_{-}\text{No VERB}_{C1}-\text{viac}_{-}\text{ADVp}_{-}\text{FULL CLAUSE}_{C2} **CC** meso-construction (SNC)

**FORM:**

\[
[[\text{tši:m}] \text{ďalej} […] \text{ADVp}1 […] \text{obligatory clause 1}]C1
[[\text{tši:m}] \text{viac} […] \text{ADVp2} […] \text{obligatory clause 2}]C2
\]

**MEANING:**

‘[As the point in time increases with respect to obligatory clause 1]\text{independent variable}

\[\rightarrow [\text{so the quantity increases with respect to obligatory clause 2}]\text{dependent variable}

\] in a monotonic way’

(58) \text{ďalej}_{-}\text{ADVp}_{-}\text{No VERB}_{C1}-\text{viac}_{-}\text{ADVp}_{-}\text{No VERB}_{C2} **CC** meso-construction (Slovak Web 2011)

**FORM:**

\[
[[\text{tši:m}] \text{ďalej} […] \text{ADVp1}]C1
[[\text{tši:m}] \text{viac} […] \text{ADVp2}]C2
\]

**MEANING:**

‘[As the point in time increases]\text{independent variable}

\[\rightarrow [\text{so the quantity increases}]\text{dependent variable}

\] in a monotonic way’
(59) d'alej_AdvP_No VERBc1-menej_AdvP_No VERBc2 CC meso-construction (Slovak Web 2011)

|-------|--------------------------|

Meaning: 

′[As the point in time increases]independent variable

→ [so the quantity decreases]dependent variable

in a monotonic way′

Now, as noted in section 1.1, CxG assumes that the grammar of a language is not simply an inventory of constructions (as might be implied by their listing above). Rather, it features meso-constructions that are interconnected in an elaborate taxonomic network: Every construction inherits from multiple more abstract constructions that represent “generalizations across constructions” (Goldberg 1995: 67).

Take the d'alej_AdvP_No VERBc1-viac_AdvP_No VERBc2 CC meso-construction (58) for example. Apart from the maximally abstract CC macro-construction (37), the results of the covarying-collexeme analyses above suggest that (58) inherits from as many as five more abstract meso-constructions:

- d'alejc1-viacc2 CC meso-construction (39)
- AdvPc1-AdvPc2 CC meso-construction (43)
- No VERBc1-No VERBc2 CC meso-construction (44)
- d'alej_No VERBc1-viac_No VERBc2 CC meso-construction (49)
- AdvP_No VERBc1-AdvP_No VERBc2 CC meso-construction (52)

Of course, since these meso-constructions are abstract to varying degrees, some will inherit from one another as well, in addition to all of them inheriting from the maximally abstract macro-construction (37). This results in the elaborate kind of network that Traugott and Trousdale have described as “rather baroque, involving massive redundancy and vastly rich detail” (2013, p. 53). This fact is better appreciated by means of a visualization20 (Figure 2) analogous to Croft and Cruse’s imperative constructional network (cf. Figure 1):

The visualization shows how the d'alej_AdvP_No VERBc1-viac_AdvP_No VERBc2 CC meso-construction (58) inherits from the less substantive meso-constructions (39), (43), (44), (49), and (52). It also shows how these are in turn interconnected: For example, the AdvP_No VERBc1-AdvP_No VERBc2 CC meso-construction (52) inherits from the No VERBc1-No VERBc2 CC meso-construction (44)

---

20 Note that the meso-constructional level in itself consists of several ‘sub-levels’: As has been noted, “Filler Types are one meso-constructional level above Lexical Fillers” (Horsch 2021: 217) because the latter are more substantive than the former (cf. also table 1 in section 1.1, which shows that constructions can be partially substantive/schematic to varying degrees). This has been accounted for in the illustration, where more substantive meso-constructions are located towards the lower end of the meso-constructional level.
and the $\text{ADV}_1\text{ADV}_2$ CC meso-construction (43). Of course, the same meso-construction can also license several more substantive meso-constructions. Thus, the $\text{NO VERB}_1\text{NO VERB}_2$ CC meso-construction (44) licenses not only the $\text{ADV}_1\text{ADV}_2$ CC meso-construction (52) but also the $\text{dalej NO VERB}_1\text{viac NO VERB}_2$ CC meso-construction (49) and the $\text{dalej ADV}_1\text{NO VERB}_1\text{viac ADV}_2\text{NO VERB}_2$ CC meso-construction (58).

Note that Figure 2 is just a minuscule detail of the overall Slovak CC network: The less substantive meso-constructions without lexically fixed comparative elements (43), (44), and (52) will license many more meso-constructions such as (59). And of course, it should not be forgotten that there is a micro-constructional level on which we find a multitude of concrete utterances such as the examples retrieved from the SNC and Slovak Web 2011 corpus. Behind each of these micro-constructions, there is an elaborate network similar to the one presented in Figure 2. Also note that in the case of constructions involving Full Clauses, even more meso-constructions will be involved to account for e.g. conjugated verbs (recall that morphological phenomena can also be accounted for by means of partly substantive, partly schematic constructions).

5. CONCLUSION

In this paper, I have provided a brief introduction to an innovative and highly successful theory of language that is known as Usage-based Construction Grammar (CxG). The core assumption of this theory is that language knowledge (i.e., grammar)
is constituted by constructions: arbitrary and conventional form-meaning pairings at all levels of linguistic analysis. These constructions are interconnected in an elaborate taxonomic network, which in its entirety is referred to as the construct-i-con. Arguably, the most interesting part of the construct-i-con is the so-called meso-constructional level, on which partly substantive and partly schematic constructions are encountered.

Addressing questions about how to find evidence for the existence of constructions, a corpus study based on a total of almost 800 CC tokens from the SNC (442 tokens) and the Slovak Web 2011 Corpus (343 tokens) was presented. With advanced statistical methods it was possible to detect altogether 22 cross-clausal associations that constitute evidence for the entrenchment of meso-constructions of varying degrees of abstractness.

The existence of such cross-clausal associations has serious implications for Mainstream Generative Grammar-based analyses of language. These strive for “a maximally schematic and economic analysis of CC constructions” (Hoffmann et al. 2019, p. 3) in which “clauses and phrases are […] treated as independent structures whose internal syntax is unaffected by the syntagmatic context in which they occur” (Hoffmann et al. 2019, p. 33). However, the existence of the 22 meso-constructions that were unearthed in the present corpus study suggests that the two clauses that make up the Slovak CC construction (C1 and C2) are anything but independent of each other. Like their English counterparts that were uncovered by Hoffmann et al., they simply “cannot be captured by a single, maximally abstract macro-construction” (2019, p. 32). Rather, the findings of the corpus study presented here suggest that an elaborate network of meso-constructions (cf. Figure 2) underlies the Slovak CC construction: Many of the meso-constructions detected by covarying-collexeme analysis are obviously related to each other, regarding their form (on which the focus of this study has been) and their meaning. This was illustrated with a small detail of the Slovak CC network (Figure 2) involving five of the meso-constructions that were uncovered using covarying-collexeme analysis.

As Croft and Cruse (2004, p. 241) have noted, the existence of meso-constructions implies that speakers possess an extraordinary range of specialized syntactic knowledge that goes beyond general rules of syntax and semantic interpretation on the one hand, and a list of substantive idioms on the other.

Thus, one decisive advantage of Usage-based CxG over generativist approaches becomes obvious when it comes to accounting for phenomena (in this case, cross-clausal associations) that defy analyses using maximally general rules.

Apart from improving our knowledge of the Slovak CC construction, the present study has also demonstrated the importance of corpus data for Usage-based CxG research. With introspection, it would impossible to quantify the entrenchment of constructions such as the cross-clausal associations revealed by covarying-collexeme analysis. Therefore, it is my sincere belief that the venerable and
extremely prolific tradition of corpus linguistics in Czechia and Slovakia holds great potential for the development of Usage-based CxG, in particular by putting this highly successful theory to the test in languages other than English.

References


Resumé

**Od korpusových dát ku konštrukčným sieťam: analýza jazyka v rámci tzv. usage-based konštrukčnej gramatiky**

Konštrukčná gramatika (CxG) predstavuje inovatívny prístup k štúdiu jazyka, ktorý v anglo-saskom svete posledných tridsať rokov postupne získaval na popularite. V CxG sa za základnú jednotku jazykové analýzy považujú konštrukcie: arbitrárne a konvencionalizované kombinácie formy a významu, ktoré pripomínajú Saussurove jazykové znaky, na rozdiel od nich sú však vymedzované na rôznych úrovniach jazykové analýzy aj za hranicami lexikónu. V tejto oblasti vzniklo v poslednom období množstvo prác, ktoré potvrdili viacero hypotéz CxG. Napriek tomu však daný teoretický prístup ostáva pomerne málo známy mimo anglosaského kontextu. V štúdii sa zamierame na jednu špecifickú oblasť CxG, ktorá sa označuje ako „usage-based“. V tomto prístupe sa predpokladá, že konštrukcie sa osvojujú na základe frekvencie príslušného jazykového inputu, teda prostredníctvom opakovaného vystavenia sa lingvistickým štruktúram (preto atribút „usage-based“). Hlavným cieľom prispevku je ukázať, ako možno analizovať konštrukcie a oblast s cieľom doložiť tzv. „ukotvenie“ (z angl. *entrenchment*) lingvistických štruktúr, a teda existenciu konštrukcií. Na potvrdenie týchto predpokladov budeme používať metódu tzv. kovariantnej analýzy, ktorá bude vychádzať z dát Slovenského národného korpusu (SNK) a Slovenského webového korpusu (skTenTen11), odkiaľ bolo extrahovaných 785 tokenov reprezentujúcich tzv. komparatívnu korelatívnu (CC) konštrukciu (napr. Čím viac čítam, tým viac rozumiem.).

**Kľúčové slová:** usage-based konštrukčná gramatika, komparatívna korelatívna konštrukcia, slovenčina, mezokonštrukcia, mentálne siete