DOI: https://doi.org/10.31577/slovrec.2025.2.2

ŠTÚDIE A ČLÁNKY

CONSTRUCTION GRAMMAR-BASED RESEARCH ON SLOVAK, IN SLOVAKIA: A CASE STUDY AND A ROAD MAP

Jakob Horsch – Hans C. Boas

HORSCH, J. – BOAS, H. C. (2025): Construction Grammar-Based Research on Slovak, in Slovakia: A Case Study and a Road Map. In: *Slovenská reč*, 90/2, 11 – 50.

Abstract: This paper presents a pioneering application of Construction Grammar (CxG) to Slovak, a language largely underrepresented in constructionist research. After providing a comprehensive introduction to CxG and its theoretical foundations, we present a case study of the Slovak Comparative Correlative (CC) construction (e.g., *Čím menej rečí tu bude, tým skôr zaspím* 'The less talking there is here, the sooner I will fall asleep'). In the context of a large-scale corpus study based on over 3,500 tokens from the Slovak Web 2011 corpus, we demonstrate how constructionist methodology is applied, employing covarying-collexeme analysis to reveal statistically significant patterns of cross-clausal association across the CC's subclauses. Our findings uncover an intricate network of interconnected meso-constructions, demonstrating both high productivity and a strong tendency toward formal symmetry. The case study thus provides robust empirical support for a usage-based, network-oriented view of Slovak grammar, in line with previous investigations of the English and Spanish CC constructions. Our study not only enriches our understanding of the Slovak CC construction but also argues for the value and feasibility of CxG approaches in 'smaller' Slavic languages. We conclude with a roadmap for future CxG-based research on Slovak, advocating for comprehensive construct-i-con projects that 'map out' the constructional networks of Slovak, and the integration of complementary methodologies in doing so. Our results contribute both theoretically and methodologically to constructionist linguistics and serve as a template for more comparable studies in Slovak and other underexplored languages.

Keywords: Construction Grammar, Slovak, Comparative Correlative, corpus study.

Highlights:

- The study presents a brief introduction to Construction Grammar (CxG), followed by a case study illustrating how a construction – in this case, the comparative correlative (CC) – may be analyzed within the CxG framework on the basis of corpus data.
- The study shows that Slovak lends itself very well to CxG-based studies.
- The detailed corpus study of CC constructions demonstrated the feasibility of a CxG-based analysis of syntactic patterns in Slovak based on corpus data.

- The study promotes future constructionist research on Slovak to contribute insights into the nature of the Slovak construct-i-con as well as into how construction entries are combined to license complete sentences.

1. INTRODUCTION

Construction Grammar (CxG) is a cognitive approach to language that has gained considerable popularity, with a large body of research published over the last three decades (see Section 2 below). While it initially started out in the English-speaking countries, it has since firmly established itself in many other countries such as Japan, Germany, Sweden, Brazil and France. Despite its extraordinary success, it remains largely unknown in Slovakia[1]. Horsch, for example, has noted that CxG "does not appear to be well-known in […] Czechia and Slovakia" (2023a, p. 702). We acknowledge that this observation is based on anecdotal evidence, but the lack of CxG studies on 'smaller' Slavic languages such as Slovak is obvious and striking (although there is a small and growing number, cf. e.g. Horsch 2021, 2023a; Ivanová 2021). As we argue in this paper, however, there is great potential in the study of these languages against a CxG background. Firstly, assuming a cognitive perspective that pays equal attention to what has traditionally been labeled the 'core' and the 'periphery' can greatly improve our understanding of specific languages. This is especially true of phenomena that have proven difficult to describe with other approaches, e.g. island structures in English (Ambridge – Goldberg 2008), so-called ethical datives in Czech (e.g. Fried 2010), or the Comparative Correlative construction in Slovak (Horsch 2021, 2023a)[2]. Secondly, CxG, which has so far focused mainly on English and a few other languages such as French, German, Japanese, Brazilian Portuguese, Spanish, and Swedish, stands to benefit substantially from demonstrating that it is able to accommodate typologically different languages such as Slovak, thereby contributing to its "aspirations toward universal applicability" (Fried 2017, p. 249) and demonstrating how research on other languages can contribute to CxG as a theory in general. And finally, colleagues operating within other theoretical frameworks may also find useful the data and results of constructional studies, which are conducted according to the standards of scientific rigor and following good experimental practice (see Cappelle 2024 and Boas – Leino – Lyngfelt 2024).

In the remainder of this paper, we provide a brief introduction to Construction Grammar (Section 2), followed by a case study illustrating how a construction – in this case, the comparative correlative (briefly introduced in Section 3) – may be

[1] This also applies to other countries, albeit not uniformly across the former Eastern Bloc. While colleagues from Ukraine have pointed out that constructional approaches "are not widely practiced" and "not well-known in Ukraine" (Karamysheva, p.c. 2024), there is a large body of CxG research on Czech (e.g. Fried 2007, 2013, 2017; Fried – Lipská 2020) and for Russian, there is even a construct-i-con (cf. Janda et al. 2018; Janda et al. 2020; Endresen et al. 2020).

[2] We would like to thank an anonymous reviewer for pointing this out.

analyzed within the CxG framework. Specifically, we conduct a corpus study, whose methodology and data we discuss in Section 4. Results are presented in Section 5, followed by a discussion in Section 6. Finally, Section 7 offers a conclusion and a roadmap that suggests further steps for CxG-based research on Slovak.

2. CONSTRUCTION GRAMMAR: A BRIEF INTRODUCTION

Construction Grammar is a framework for linguistic analysis whose beginnings can be traced back to the mid-1980s, when it was developed by Charles Fillmore, Paul Kay, and George Lakoff, among others, at the University of California, Berkeley. Its intellectual roots can be traced back to Fillmore's seminal (1968) paper *The Case for Case*, which was highly influential during the late 1960s and early 1970s, but which was then largely abandoned because of a number of theoretical problems and inconsistencies (see Levin – Rappaport Hovav 2005 and Boas – Dux 2017 for details). During the late 1970s and early 1980s, Fillmore worked on revising and expanding his original ideas from his (1968) paper in a number of different ways.

One major outcome of these efforts was a new framework for dealing with linguistic meaning that he termed Frame Semantics (Fillmore 1982, 1985a), which was refined throughout the 1980s and into the 1990s. In contrast to the then mainstream prevalent theories of meaning that worked with, among other things, semantic features and terms from predicate logic to model meaning, Fillmore proposed a theory of understanding whose explicit goal was to model what a speaker of a language needs to know in order to encode and decode meaning ("semantics of understanding"). One of the central ideas of Fillmore's usage-based approach to meaning is the so-called semantic frame, which should be regarded as primary for the description and analysis of meaning (as well as its syntactic relevance, as we will see below). Fillmore and Atkins (1992, p. 76 – 77) characterize the nature of semantic frames for the understanding of words as follows:

> A word's meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes. Within such an approach, words or word senses are not related to each other directly, word to word, but only by way of their links to common background frames and indications of the manner in which their meanings highlight particular elements of such frames.

One of Fillmore's examples illustrating the central role played by semantic frames concerns the meaning of the verb *to sell*, whose meaning requires an understanding of a number of different concepts. In frame-semantic terms, *to sell* evokes the `Commercial Transaction` frame, which requires knowledge of the concepts of exchange and market economy, and which consists of the frame elements (situation-specific semantic roles) BUYER (the person acquiring the GOODS by giving the seller MONEY), SELLER (the

person wanting to give the GOODS away in exchange for MONEY), MONEY (units of payment), and GOODS (things that can be acquired and given away) (among others). The `Commercial Transaction` frame is not only evoked by different types of verbs such as *to buy, to sell, to pay*, and *to purchase*, but also by different types of nouns such as *buyer, seller, payment, receipt*, and *money* and different types of adjectives such as *expensive, cheap*, and *pricey*. Mentioning of any of these different words evokes, according to Fillmore, the `Commercial Transaction` frame, and each of these words gives a slightly different perspective of the commercial transaction event as is evidenced by their syntactic behaviors. For example, *to buy* requires overt syntactic realization of the BUYER and the GOODS while making realization of the MONEY and SELLER optional. Compare the following examples (1a) – (1d):

(1) a. Sascha bought the table (from Kim) (for $200).
b. *Sascha bought.
c.*Sascha bought for $200.
d.*Sascha bought from Kim.

In contrast, the verb *to sell* requires overt syntactic realization of the SELLER and the GOODS while making realization of the BUYER and the MONEY optional, as the following examples (2a – d) illustrate:

(2) a. Kim sold the table (to Sascha) (for $200).
b. *Kim sold.
c.*Kim sold for $200.
d.*Kim sold to Sascha.

Finally, the adjective *cheap*, which also evokes the `Commercial Transaction` frame, requires obligatory syntactic realization of the GOODS and it does not allow the syntactic realization of the Seller, as the following examples show (*Kim* is supposed to be the SELLER).

(3) a. The table was expensive.
b.*The table was expensive for Kim.

During the 1980s and into the 1990s, Fillmore and his associates developed Frame Semantics further, investigating different types of meaning structures that can be modeled with semantic frames (for an overview, see Petruck 1996, Busse 2012, Matsumoto 2025). One of the major outcomes of these efforts was the start of the FrameNet project at the International Computer Science Institute in Berkeley, California, in 1997, which applied the concept of semantic frame to analyze and structure the English lexicon. Over the past 25+ years, FrameNet has produced a lexicographic database of English that structures the English lexicon with semantic frames together with the words that evoke them (see Fillmore et al. 2003, Fillmore – Baker 2010). The FrameNet database contains, among other things, several

thousand semantic frames organized in a frame hierarchy and more than 10,000 frame-evoking lexical entries that provide detailed information about the semantic frames they evoke, together with specific valence information about how the semantics of frames are realized syntactically (e.g. as in the three examples (1) – (3) above) (for more information on FrameNet, see Boas 2017 and Boas – Ruppenhofer – Baker 2024).

The second major outcome of Fillmore's revisions of his original ideas in *The Case for Case* during the 1980s is the concept that meaning is intricately related to form, as has already been discussed in the context of the three examples above. Fillmore formulates this close relationship between meaning and form as follows:

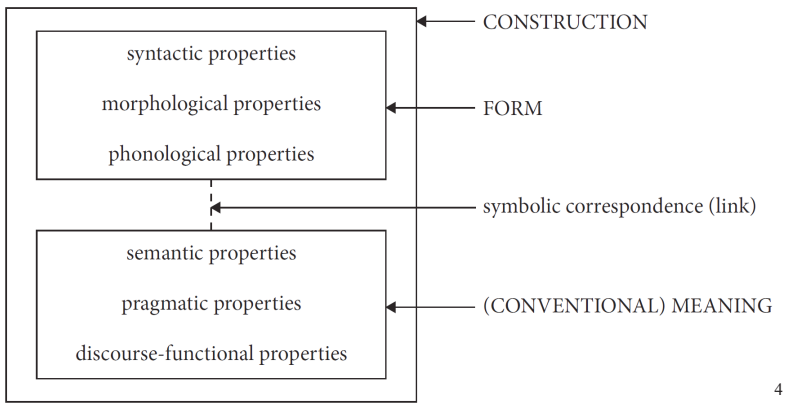
> If new-style lexical entries for content words were to be seen instead of constructions capable of occupying particular higher-phrase positions in sentences and included both the needed semantic role and the needed specifications of structural requirements [...], we could see such structures as providing expansions of their existing categories. (Fillmore 1985b, p. 84)

Fillmore's suggestion to directly tie the meaning and form of words together in a lexical entry in a way to present them so that they can be employed for modeling different types of more abstract constructions can be seen as one of the first specifications of the notion of construction as a pairing of form with meaning. In other words, Fillmore extended the concept of Saussure's linguistic sign – an arbitrary symbolic relationship between form and meaning – to any and all linguistic material, from morphemes to complex syntactic patterns (cf. Goldberg 2003, p. 220). In a series of publications in the mid- and late-1980s (Fillmore 1985, Fillmore 1986, Fillmore 1988, Fillmore 1989, Fillmore et al. 1988, Fillmore 1990), Fillmore and associates laid out some of the basic principles of CxG as a maximalist approach to grammar (see Fried – Östman (2004, p. 24) and Boas (2025a, 2025b)). On this view, CxG aims for a comprehensive (ideally full) coverage of linguistic phenomena within a single theoretical framework, i.e. it aims to account for both peripheral intransparent (i.e., opaque) grammatical phenomena such as partially filled idioms (e.g. *jog <someone's> memory* (Goldberg 2006)), semi-productive constructions such as *What's* X *Doing* Y? (e.g. *What's that fly doing in my soup?*) (Kay – Fillmore 1999)), and fully regular semantic and syntactic structures such as word order constructions (Boas – Ziem 2018), the subject-predicate construction (Fillmore et al. 2012), and passives (e.g. Ackerman – Webelhuth 1998, Lasch 2016) in terms of a non-modular and non-derivational architecture of grammar (for details, see Hoffmann (2020), Boas (2021), and Boas (2025a, 2025b)).

As a usage-based theory (Barlow – Kemmer 2000, Perek 2023), CxG aims to model what a language user needs to know in order to fully understand any linguistic expression. Central to this approach is the concept of "construction," defined as a pairing of form with meaning, according to Goldberg (1995):

> C is a CONSTRUCTION if and only if C is a form-meaning pair $<F_i, S_i>$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions (Goldberg 1995, p. 4).[3]

On this view, the basic units of language are constructions, i.e. conventionalized pairings of form with meaning at varying levels of abstraction and complexity that must be learned based on linguistic input. If an utterance cannot be licensed based on the existing inventory of constructions (or a combination of existing constructions), then one has to posit a new construction. The architecture of a construction in Figure 1 illustrates what different types of form can be associated with what different types of meaning (and function) in a construction. It is important to remember that form and meaning/function cannot be separated from each other since form and meaning/function do not typically exist on their own, e.g. as autonomous (sub-)modules as is often postulated in other syntactic theories. Since form and meaning/function constitute inseparable parts of the linguistic sign as in Figure 1, this means that a difference in form usually also signals a difference in meaning. This is known as the 'Principle of No Synonymy' (Goldberg 1995, p. 67). Note, however, that recently the Principle of No Synonymy has been challenged. Laporte et al., for example, have noted that "the Principle of No Synonymy largely breaks down at low levels of formal description" (2021, p. 230; cf. also Uhrig 2015). These issues have been addressed by the "Principle of No Equivalence" (Leclerq – Morin 2023; Leclercq 2024).



[4]

Figure 1: Types of information in constructions (Croft 2001, p. 18)[5]

[3] For newer alternative definitions, see Goldberg (2006, 2019).

[4] Constructionist principles have also been applied to morphology, see Booij (2010, 2018) and Audring and Jackendoff (2025).

[5] Opinions differ on how to model constructional meaning and whether all constructions do in fact bear meaning. For example, Fillmore (1999) argues that the English Subject Auxiliary Inversion construction does not bear any clearly identifiable meaning, while Goldberg (2006) argues otherwise. See Willich (2020), Boas (2025a), and Boas (2025b) for a discussion of constructional meaning(s).

The question of how many different types of constructions there are in a language is still a matter of on-going research. Consider, for example, a simple English sentence such as *The tacos taste delicious*, which is licensed by a combination of different types of English constructions, including the Intransitive construction, the VP construction, the AdvP construction, the NP construction, the Plural construction, the Verb-inflection construction, and a number of different lexical constructions. In CxG, utterances are licensed by taking the necessary construction entries from the construct-i-con[6] (the inventory of construction entries, see Fillmore et al. 2012, Lyngfelt 2018, Boas et al. 2019) and combining them. Since in CxG there is no strict separation between syntax and the lexicon, all constructions, from the most abstract/schematic (such as Subject-predicate or Intransitive) to the most specific/concrete (such as entries for senses of words as in FrameNet (Fillmore – Baker 2010)) are assumed to have the same data structure, which typically makes them compatible with each other so they can combine to license utterances.[7]

Fillmore et al. (2012) provide a cursory overview of the different types of English constructions they found during a one-year long pilot project exploring the organizational principles of the English construct-i-con that employed the workflow and analytical apparatus of lexical FrameNet to cover linguistic structures beyond the lexicon. The constructions that Fillmore et al. (2012) discuss cover a wide range of different structures, including, among others, lexical idiom constructions (e.g., *in the distance*), constructions with gaps, clause-level constructions, absolute constructions, auxiliary-initial constructions, co-text specification constructions, degree modification constructions, nominal pumping constructions, verb pumping constructions, and measurement expression constructions (for details, see Fillmore et al. 2012, p. 332 – 368).

An important assumption of CxG is that language consists of a network of constructions (or, perhaps multiple networks, this is an open question) in which constructions are organized in terms of different levels of schematicity and specificity (Fillmore – Kay 1993, Goldberg 1995, Boas 2011, Diessel 2019, Schmid 2020). As a maximalist approach (see above), CxG assumes that such networks are "baroque, involving massive redundancy and vastly rich detail" (Traugott – Trousdale 2013, p. 53). On this view, constructional networks are useful for representing and organizing constructions that share certain common properties, though the exact nature of how constructional networks are organized and whether all constructional networks share the same or different organizational principles remains an open

[6] This term is "a morphological blend of construction and lexicon" (Ziem – Flick 2019, p. 203).

[7] For a discussion of different concepts of the nature of the construct-i-con (and how it connects to the lexicon), see Fillmore et al. (2012), Boas (2017), Lyngfelt et al. (2018), Diessel (2023), and Boas (2025b).

question.[8] For example, Goldberg (1995) proposes a core sense and six sense extensions to cover the various realizations of ditransitive in a network of ditransitive argument structure constructions. Other researchers propose differently structured networks for different groups of constructions, including passives (Ackerman – Webelhuth 1998), conatives (Medina 2017), subject-auxiliary constructions (Fillmore 1999, Goldberg 2006), support verb constructions (Zeschel 2008), relative clause constructions (Diessel 2019), and the V-*that* construction (Perek – Patten 2019).

The concept of productivity plays an important role in CxG. While abstract schematic constructions such as the SUBJECT-PREDICATE construction impose relatively few restrictions on what types of elements can fill its open slots, other constructions that are less schematic such as the PASSIVE construction (Ackerman – Webelhuth 1998, Lasch 2016) or the CAUSED-MOTION construction (Goldberg 1995) impose significantly more restrictions of different types on their open slots (for details see Boas 2021). This means that the notion of productivity of constructions exists on a continuum that coincides with the level of abstraction and schematicity of constructions, ranging from fully productive constructions to semi- and non-productive constructions (see also Fillmore 1999, Goldberg 2006, Boas – Ziem 2018, Boas 2021).[9]

Since the mid-1980s, several different so-called "flavors" of CxG have developed out of the "original" Berkeley Construction Grammar (Fillmore et al. 1988, Fillmore – Kay 1993, Kay – Fillmore 1999, Fillmore 2013). Taken together, these are also known as "constructionist approaches" (Goldberg 2019, Ungerer – Hartmann 2023), indicating that they all share a certain **set of common assumptions** about the nature of language, including the following: (1) The **basic building block of language is a construction**, a pairing of form with meaning/function (in Goldberg's words, "it's constructions all the way down" (2003, p. 223)) and (almost) all of language consists of constructions of various degrees of abstraction and schematicity.; (2) Analysis of language should be **usage-based** (Barlow – Kemmer 2000); (3) The architecture of language is **non-modular**, i.e. unlike in most other models of language there are no separate modules for syntax, the lexicon, semantics,

[8] Another open question is the exact nature of how construction entries interact with each other to license specific sentences. As discussed in detail by Boas (2025b), this central aspect of constructional syntax has so far not been addressed in sufficient detail and needs to be investigated more thoroughly by future research.

[9] In contrast to other frameworks, in which frequency is treated as external to the grammar, many constructionist analyses regard frequency as a central concept when it comes to analyzing and explaining the nature and distribution of constructions. For the differences between a construction's type frequency and token frequency, see Clausner and Croft (1997), Goldberg (2006), Bybee (2010), Diessel (2019), Schmid (2020), Gries (2022), and Hilpert (2025).

phonology, pragmatics, etc. This means that mechanisms that connect and relate between different so-called "modules" in other approaches are not necessary in CxG; (4) There are **no different levels of** (underlying and surface) **representation**, i.e. there are no derivations in the classical sense; (5) Language is learned on the basis of **input** and there are **no empty categories** or silent operators, i.e. 'what you see is what you get.' (6) The **meaning side** of (many, if not most) constructions can be modeled with **Frame Semantics** (Fillmore 1982), the sister theory of CxG. These basic assumptions about the nature of language are shared more or less by all constructionist researchers, which means that in theory it should be possible to 'translate' a constructional analysis from one flavor of CxG into other flavors of CxG (see Sag 2012 for details).[10]

Berkeley Construction Grammar started out in the mid-1980s as an alternative model of language seeking to account for all linguistic phenomena and not only those at what was then known as the 'core' (as opposed to the 'periphery') and which occurred with sufficient frequency (for a discussion of what constitutes 'sufficient' frequency, see Section 4.1 below). In this context it is important to note that there are different views[11] on what constitutes 'constructionhood': For example, some approaches posit a form-meaning pairing "such that some aspect of [the form] or some aspect of [the meaning] is **not strictly predictable** from [the construction's] component parts" (Goldberg 1995, p. 4, emphasis by the authors). In other words, under this approach a construction typically has a specific meaning element that is not recoverable from the meanings of its components[12]. Goldberg later developed this concept into a more extensive concept according to which **fully compositional** structures are also analyzed as constructions in their own right, as long as they are frequent enough ("sufficient frequency", cf. (Goldberg 2006, p. 5)). At the center of this alternative approach is a sign-based conception of language that emphasizes constructions as the basic building blocks of language, pairing form with meaning and function. As Fillmore (2013, p. 112) puts it:

> [T]he grammar of a language is the set of its grammatical constructions, the rules that unite formal and semantic information into various kinds of linguistic objects, together with the principles that constrain and connect them . . . Any well-formed linguistic entity can be interpreted as an assembly of the constructions that jointly license it, and an ambiguous expression permits more than one such assembly […] **What you see is what you get**. (Fillmore 2013, p. 112, emphasis by the authors)

[10] Whether the meanings of all constructions can be modeled with semantic frames remains an open question. For a discussion, see Willich (2023) and Boas (2025b).

[11] We would like to thank an anonymous reviewer for pointing this out.

[12] In other words, under this approach frequency is not a necessary condition for 'constructionhood', rather, it is a matter of the internal makeup and meaning of a structure.

In the early 1990s, Cognitive Construction Grammar (Goldberg 1995, Boas 2013) grew out of Berkeley Construction Grammar by focusing on the nature of one particular set of constructions, namely argument structure constructions such as the *Ditransitive*[13] construction, the *Caused-Motion* construction, and the *Resultative* construction (Goldberg 1995, Goldberg 2006), and by highlighting the cognitive aspects of language, i.e. how linguistic information is learned, stored, and processed. While Cognitive Construction Grammar does not put an emphasis on notation and formalization, Berkeley Construction Grammar is keen on providing a more formalized notation of linguistic insights, similar to Sign-based Construction Grammar (Sag 2012, Michaelis 2013), which grew out of Berkeley Construction Grammar and Head-driven Phrase Structure Grammar (Pollard – Sag 1994), focusing on a more detailed unification-based formalism using Attribute-Value Matrices (AVMs) for representing constructions and interactions between constructions (see Sag – Boas – Kay 2012). Other flavors of CxG include Radical Construction Grammar (Croft 2001, Croft 2013), which is based, among other things, on insights from linguistic typology. According to Croft, each language comes with its own categories and constructions that sets it apart from all other languages. More computationally oriented flavors of CxG include Embodied Construction Grammar (Bergen and Chang 2013) and Fluid Construction Grammar (Steels 2013), which are interested in computational simulation and implementation of constructional insights into the nature of language. Finally, there is a variety of applications of CxG to fields such as second language acquisition and language pedagogy, as, for example, in Pedagogical Construction Grammar (Herbst 2016, Boas 2022), which seeks to apply constructional insights to the teaching and learning of languages. For an overview and a comparison of the various 'flavors' of CxG, see Boas (2020) and Boas (2021).

One of the main goals of constructional research is to find, document, and analyze all constructions of a language in order to determine how they interact to license utterances in a language. In the 1980s, CxG started out by primarily focusing on English constructions, as laid out by Fillmore and Kay (1993, p. 4 – 5):

> We will be satisfied with the technical resources at our disposal, and with our use of them, if they allow us to represent, in a perspicuous way, everything that we consider to be part of the conventions of the grammar of the first language we work with. We will be happy if we find that a framework that seemed to work for the first language we examine also performs well in representing grammatical knowledge in other languages.

This quote illustrates the idea that one should first try to determine the constructions of one language before moving on to other languages. This reflects the

[13] Following Boas and Sag (2012), constructions are rendered here with the Courier New font in italics.

assumption that there might not be an innate language faculty with universal principles (a.k.a. Universal Grammar) as proposed by other linguistic theories. Following the pioneering constructional research on English, linguists started exploring other languages, including Brazilian Portuguese, French, German, Japanese, Spanish, Swedish, and many others from a constructional perspective. As already discussed above, Croft (2001) claims that all constructions are language-specific and that it might be very difficult to come up with constructional generalizations across languages (see also Croft 2025). Other constructionist researchers, however, propose that certain types of generalizations across languages are indeed possible, especially from a contrastive approach. For example, the contributions in Boas (2010) compare a variety of different English constructions with their counterparts in other languages, including Finnish, German, Japanese, Russian, Spanish, Swedish, and Thai. Similarly, Kuzar (2012) provides a contrastive analysis of sentence patterns in English and Hebrew, while Dux (2020) shows how generalizations can be made across the different types of constructions instantiated by change and theft verbs in English and German. Other efforts looked at the nature of constructions in particular language families such as Romance (see the contributions in Boas – Gonzálvez García 2014) and the Nordic Languages (see the contributions in Coussé et al. 2023) or particular languages such as French (see the contributions in Bouveret – Legallois 2012), German (see the contributions in Boas and Ziem 2018), and Spanish (see the contributions in Hennecke – Wiesinger 2023).

With this short overview of how CxG has developed over the past 35+ years, in particular with respect to its expansion from English to other languages, we now turn to the main part of our paper. That is, in the following sections we will discuss the nature of a specific construction, namely the Comparative Correlative (CC) construction (e.g. *The more we walk, the more tired we get*), which has been analyzed in English (see e.g. Hoffmann 2019; Hoffmann et al. 2019, 2020; Horsch 2023b) and other languages (e.g. Spanish (Horsch 2024) and Slovak (Horsch 2021, 2023a)). More specifically, we discuss the various semantic, pragmatic, and syntactic similarities and differences between the Slovak CC Construction and its English counterpart to demonstrate how the principles of CxG can be applied to the Slovak language.

3. THE SLOVAK COMPARATIVE CORRELATIVE CONSTRUCTION

Prototypically, the Slovak Comparative Correlative (CC) consists of two subclauses C1 and C2[14], as example (4) illustrates:

[14] As an anonymous reviewer has noted, some analyses of the English CC have posited that it is a hypotactic construction, with C1 being the subordinate and C2 the main/matrix clause (e.g. Culicover – Jackendoff 2005, p. 504 – 507). However, recent CxG-based corpus studies have adduced evidence

(4) [*Čím menej rečí tu bude,*]$_{C1}$ [*tým skôr zaspím.*]$_{C2}$
'The less talking there is here, the sooner I will fall asleep.'
(from Horsch 2019, p. 180)

Regarding its semantics, the Slovak CC – like its counterparts in many other languages – is quite complex. It encodes two distinct meanings: (1) **cause-effect**, which has been described as a "conditional, **asymmetric**, effect-cause relationship between C1 and C2" (Horsch 2021, p. 196) (talking less results in falling asleep sooner) and (2) **parallel change over time** that amounts to a "pair of semantic differentials" between which a "monotonic relationship" exists (Hoffmann 2019, p. 38; cf. also Sag 2010, p. 525 – 526) (over the same time period, there is less talking while the likelihood of falling asleep simultaneously increases). Horsch refers to this meaning as "**symmetric**" (2021, p. 196).

The Slovak CC also has a complex form, which was first extensively described by Horsch (2019, 2020, 2021). The first thing to note in this regard are the lexically/phonologically fixed clause-initial elements *čím*$_{C1}$ and *tým*$_{C2}$ that introduce the subclauses C1 and C2. As Horsch (2019, p. 182) has noted, there appears to be a "general uncertainty concerning the classification" of these clause-initial elements: Betáková assigned them to the category of correlative conjunctions (1955, p. 319), Dvonč et al. (1966) describe them as hypotactic conjunctions, while Oravec (1954) suggests that they have become fossilized as a pair of conjunctions. Similarly, Slovak lexicographic works have analyzed *čím*$_{C1}$ and *tým*$_{C2}$ as analytic subordinating conjunctions (Kačala et al. 2003; Buzássyová – Jarošová 2006)[15]. However, Marsinová (1955, p. 39) suggests that the pair be excluded "a priori" from the category of conjunctions because she considers the word *čím* not to be related to the conjunction *čo*. Finally, Horsch (2021, p. 197) points to an analysis by Sabol, who has described *čím*$_{C1}$ and *tým*$_{C2}$ as particles that modify comparatives, noting that *čím* loses its (syntactic) validity when *tým* is removed from the structure (1982, p. 53). Finally, and in a similar vein, Horsch calls the clause-initial elements *čím*$_{C1}$ and *tým*$_{C2}$ "construction-specific elements" based on the observation that they resemble instrumental-case pronouns but do not function as such in the CC construction (2021, p. 197).

Also note that each subclause of the Slovak CC consists of "obligatory and optional slots" (Horsch 2021, p. 197). The obligatory slot "contains a comparative

---

that suggests it is a paratactic structure, based on the strong formal parallelism between C1 and C2 (Hoffmann et al. 2020, p. 213). The question whether the relationship between C1 and C2 is hypotactic or paratactic is therefore also relevant against the theoretical background of CxG. In fact, this was one of the central research questions of Horsch's study on the English CC, which was based on several thousand tokens of corpus data and a large-scale psycholinguistic experimental investigation with several hundred participants from various English-speaking countries (Horsch 2023b).

[15] We would like to thank an anonymous reviewer for pointing this out.

element" in the form of a noun phrase, an adjective phrase, or an adverb phrase (Horsch 2021, p. 197). This is followed by a slot containing a verb, which is optional (i.e., it does not have to be realized), as demonstrated by example (5):

(5) [*Čím ďalej,*]$_{C1}$ [*tým lepšie*]$_{C2}$
'The further, the better'

Horsch has therefore distinguished between "verbless" and "full" clauses (2021, p. 198); in the literature on the CC in other languages, these are also referred to as "deleted" and "full" clauses (cf. Hoffmann 2019; Hoffmann et al. 2019). In the present study, we will adopt the latter terminology, i.e., 'deleted'[16] to describe C1 and C2 subclauses without realization of the optional clause slot and 'full clause' to describe realized optional clause slots.

Based on the formal features outlined above, Horsch has suggested the following maximally schematic template (6) for the Slovak CC (2019, p. 184; 2021, p. 198, based on Culicover – Jackendoff's (1999, p. 567) template of the English CC):

(6) [[tʃi:m] [...]$_{\text{comparative element}}$ (clause)]$_{C1}$ [[ti:m] [...]$_{\text{comparative element}}$ (clause)]$_{C2}$

Note that from a CxG perspective, this is a construction, because there is a distinct form (cf. template (6)) that encodes a distinct meaning (asymmetric and symmetric, see above)[17] (from Horsch 2023a, p. 730):

FORM:	[[tʃi:m] […]$_{\text{comparative element1}}$ (…)$_{\text{optional clause 1}}$]$_{C1}$ [[ti:m] […]$_{\text{comparative element2}}$ (…)$_{\text{optional clause 2}}$]$_{C2}$
MEANING:	'[As the degree of comparative element 1 increases/decreases with respect to clause1/ with respect to a contextually retrievable clause]$_{\text{independent variable}}$ → [so the degree of comparative element 2 increases/decreases with respect to clause2/ with respect to a contextually retrievable clause]$_{\text{dependent variable}}$ in a monotonic way'

However, this template is not completely satisfactory, for several reasons: Firstly, as noted by Horsch, the order of the optional and obligatory slots is quite flexible in the Slovak CC (2021, p. 198 ff.). For example, a corpus study by Horsch

[16] Note that we use the term "deletion" here to stay with conventions in the literature (e.g. Hoffmann 2019; Hoffmann et al. 2019; Horsch 2021; Horsch 2023b). As two anonymous reviewers have pointed out, this terminology implies an absence of derivation operations, which does not align with the surface-based approach of CxG. Hoffmann et al. have therefore explicitly noted that "the absence of BE [and by extension, whole clauses] can also be modelled constructionally without the assumption of a deleted, covert form" (2019, p. 7).

[17] The meaning side is based on Hoffmann's (2017, p. 351) English CC paraphrase.

has shown that the comparative element does not have to immediately follow the clause-initial element (e.g. *čím som sa* ***viac*** *usiloval* […] 'the more I tried […]'); this is not restricted to clitics (e.g. *Čím bol spisovateľ* ***odvážnejší*** […] 'the braver an author was […]') (2019, p. 185; 2021, p. 199). Secondly, the order of C1 and C2 may also be changed, so that C2 precedes C1 (e.g. [*filozofi boli tým lepší,*]$_{C2}$ [*čím boli starší*]$_{C1}$ 'the philosophers were better, the older they were', from Horsch 2021, p. 200). Thirdly, there are 'stacked' CCs that consist of more than two subclauses (e.g. [*čím viac rastie cena komodít a energie,*]$_{C1}$ [*tým viac je americká ekonomika „spanikárená",*]$_{C2}$ [*tým viac sa prepadá do recesie*]$_{C3}$ 'the more the price of commodities and energy grows, the more the American economy panics, the more it sinks into recession', from Horsch 2019, p. 187)[18].

Most importantly from a CxG perspective, however, previous corpus studies of the Slovak CC (Horsch 2021, 2023a) have uncovered statistically significant cross-clausal associations across C1 and C2[19]. Specifically, in these studies Horsch used so-called covarying-collexeme analyses (see Section 4.1 below) that were first used to investigate the English CC construction by Hoffmann et al. (2019). For example, in a corpus study based on a 500-token sample from the Slovak National Corpus (SNC), Horsch found a statistically significant preference for a 'deleted' (i.e., an unrealized) clause slot in C2 when the same was the case in C1, and vice versa (2021, p. 214). Similarly, in a 503-token data set from the Slovak Web 2011 corpus, Horsch found a statistically significant preference for the obligatory comparative element slot in C2 to contain an AdvP when the same was the case in C1, and vice versa (2023a, p. 723).

That is, despite the many different choices that speakers have when realizing C1 and C2, in the corpus data there are certain patterns that appear significantly more often than could be expected by chance. This suggests that the two subclauses C1 and C2 are formally/syntactically interdependent. Apart from implications for generativist analyses of the CC[20] (for more detailed discussions, cf. Horsch 2021 and in press), from a CxG point of view these cross-clausal associations are of interest

[18] As an anonymous reviewer has remarked, such stacked clauses "generally refer to multiple relative clauses modifying the same head noun [and demonstrate that] CCs are not restricted to the two-part structure, as argued by Culicover and Jackendoff (2005, p. 501) and den Dikken (2005) but there are 'tripartite' structure CCs as well."

[19] These corpus studies were based on an earlier study of the English CC by Hoffmann et al. (2019), which turned up very similar results.

[20] Specifically, generativist approaches assume that C1 and C2 are daughter nodes of a mother node and otherwise independent of each other. This means that the 'granddaughter nodes' C1 and C2 are communicating with each other (principle of locality). As noted by Hoffmann (2019, p. 139), in syntactic trees only immediate constituents are available to a mother node: "the [mother] construction can only access information of its direct [daughters] (C1 and C2 in this case), but not information embedded within its [daughters]" (Hoffmann 2019, p. 139).

because they are evidence for so-called meso-constructions. That is, form-meaning pairings less schematic than the maximally schematic CC 'macro-construction' as shown in template (6) that "occur with sufficient frequency" (Goldberg 2006, p. 5) to warrant calling them constructions in their own right. Inheriting from the maximally schematic macro-construction (see template (6)) and each other, they form the nodes of an elaborate, interconnected constructional network that represents linguistic knowledge.

These meso-constructions are of varying degrees of schematicity: For example, Horsch (2021, 2023a) also detected a preference for the less schematic lexical filler pair *ďalej*$_{C1}$-*viac*$_{C2}$ 'further-more' in his SNC and Slovak Web 2011 data. Of course, as constructions in their own right, corresponding form-meaning pairings can be posited for each of these statistically significant cross-clausal associations. In the case of *ďalej*$_{C1}$-*viac*$_{C2}$, Horsch (2023a, p. 712) has suggested the following form-meaning pair[21]:

FORM:	[[tʃi:m] [*ďalej*]$_{\text{comparative element1}}$ (…)$_{\text{optional clause 1}}$]$_{C1}$ [[ti:m] [*viac*]$_{\text{comparative element2}}$ (…)$_{\text{optional clause 2}}$]$_{C2}$
MEANING:	'[As the point in time (with respect to clause 1) increases]$_{\text{independent variable}}$ → [so the quantity increases (with respect to clause 2)]$_{\text{dependent variable}}$ in a monotonic way'

To illustrate the constructional network underlying the Slovak CC construction, and how CC meso-constructions inherit from each other as well as the CC macro-construction, consider Figure 2 (from Horsch 2023a, p. 735), which is a visualization of the results of Horsch's statistical analyses of data from the SNC and Slovak Web 2011 corpora. As is evident from the arrows in Figure 2, the network is taxonomic, i.e., less schematic constructions are licensed by/inherit from more schematic ones. Figure 2 shows that there are three levels: at the macro-constructional level at the very top, there is a maximally schematic CC macro-construction that is based on template (6). This macro-construction licenses various "semi-productive, partly substantive, partly schematic intermediate meso-constructions" (Hoffmann et al. 2019, p. 26), which in turn license so-called micro-constructions, i.e., "specific, substantive instances of a construction" (Hoffmann et al. 2019, p. 26).

Returning to the meso-constructional level, in Figure 2 we see the *ďalej*$_{C1}$-*viac*$_{C2}$ meso-construction discussed above, which can be found on the very left of the figure. It is licensed by the CC macro-construction, as is evident from the

[21] Note that the clause-initial elements are rendered in IPA here to indicate that they are phonologically fixed/invariable (and therefore construction-specific), unlike actual pronouns, which can change their form according to e.g. case and number.

arrow connecting the CC macro-construction with the *ďalej*$_{C1}$-*viac*$_{C2}$ meso-construction. Following the arrows that protrude from this meso-construction, we see that it in turn licenses an even less schematic *ďalej*_AdvP_No Verb$_{C1}$-*viac*_AdvP_No Verb$_{C2}$ meso-construction[22]. That is, a cross-clausal association in which not only *ďalej* in C1 and *viac* in C2 are significantly associated, but, furthermore, the comparative elements in C1 and C2 are adverb phrases and the optional clause slot is not realized. Note that constructions, including meso-constructions, can inherit from multiple other constructions. Thus, in addition to the CC macro-construction, the *ďalej*_AdvP_No Verb$_{C1}$-*viac*_AdvP_No Verb$_{C2}$ meso-construction inherits from a total of 5 meso-constructions, some of which inherit from one another (cf. Figure 2):

- *ďalej*$_{C1}$-*viac*$_{C2}$ CC meso-construction, e.g. (7),
- AdvP$_{C1}$-AdvP$_{C2}$ CC meso-construction, e.g. (8),
- No Verb$_{C1}$-No Verb$_{C2}$ CC meso-construction, e.g. (9),
- *ďalej*_No Verb$_{C1}$-*viac*_No Verb$_{C2}$ CC meso-construction, e.g. (10),
- AdvP_No Verb$_{C1}$-AdvP_No Verb$_{C2}$ CC meso-construction, e.g. (11):

(7) [*čím* ***ďalej*** *ideme,*]$_{C1}$ [*tým sa počasie* ***viac*** *zlepšuje.*]$_{C2}$
‘The further we go, the more the weather will improve.’
<skTenTen token #41933821 doc#93295>

(8) [*čím* **[*rýchlejšie*]$_{AdvP}$** *vyjdeš na vrchol,*]$_{C1}$ [*tým* **[*kratšie*]$_{AdvP}$** *sa tam udržíš.*]$_{C2}$
‘The faster you go to the top, the shorter you will stay there.’
<skTenTen token #328586 doc#657>

(9) *A jasné, že* [*čím bližšie k turistom,*]$_{C1}$ [*tým drahšie.*]$_{C2}$
‘And of course, the closer to the tourists, the more expensive.’
<skTenTen token #8750626 doc#18336>

(10) *Je veľa dobrých seriálov, v poslednom čase* [*čím ďalej,*]$_{C1}$ [*tým viac.*]$_{C2}$
‘There are many good TV shows, lately the further, the more.’
<skTenTen token #19809765 doc#43183>

(11) *A ovocia je* [*čím ďalej,*]$_{C1}$ [*tým menej.*]$_{C2}$
‘There are many good TV shows, lately the further, the more.’
<skTenTen token #17867896 doc#38629>

[22] Note that Horsch uses “No Verb” here to refer to what others (e.g. Hoffmann et al. 2019) refer to as “Truncated”. While in general, we use the latter term, in this case we use the former to facilitate interpretation of Figure 2 below, which was taken from Horsch (2023a). The two terms, however, can be used interchangeably.

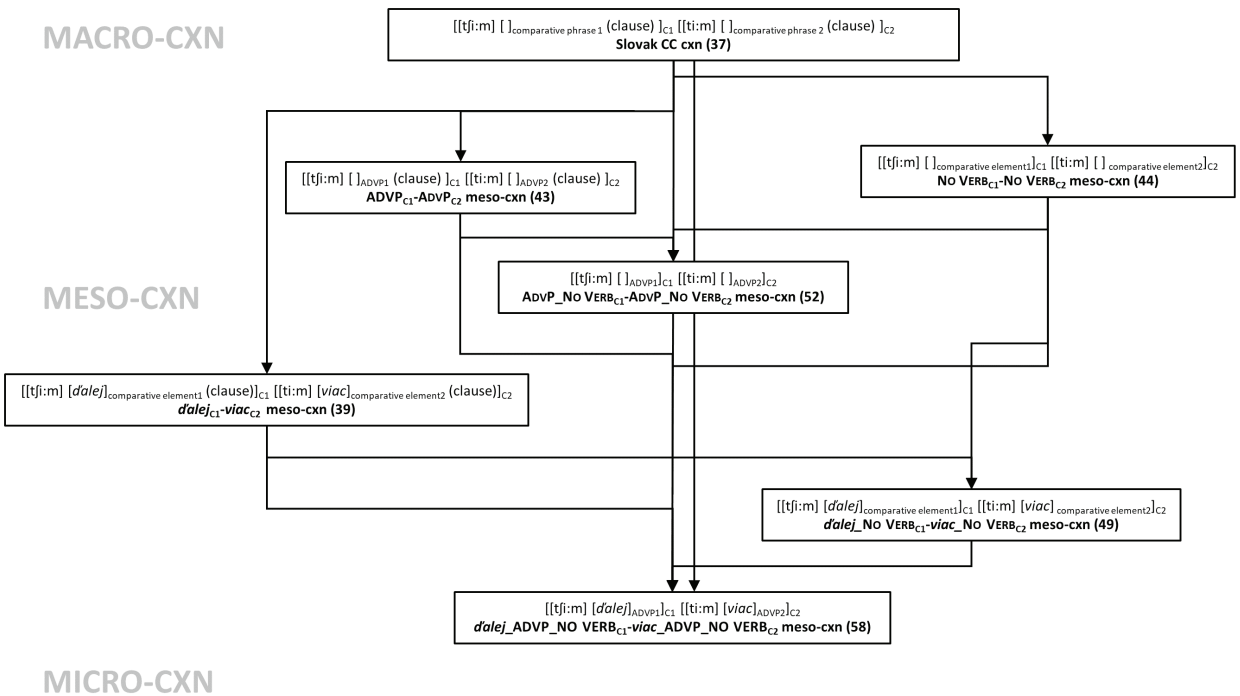


Figure 2: Detail of the taxonomy of the Slovak CC construction, based on SNC and Slovak Web 2011 data (Horsch 2023a, p. 735)

There is empirical evidence for the existence of all of these meso-constructions in the form of statistically significant cross-clausal associations in the corpus data. They can therefore also be rendered as form-meaning pairs (12 – 15) in their own right (taken from Horsch 2023a, p. 731 – 733):

(12) ADVP$_{C1}$-ADVP$_{C2}$ CC meso-construction

FORM:	[[tʃi:m] […]$_{\text{ADVP1}}$ (…)$_{\text{optional clause 1}}$]$_{C1}$ [[ti:m] […]$_{\text{ADVP2}}$ (…)$_{\text{optional clause 2}}$]$_{C2}$
MEANING:	'[As the degree of ADVP1 increases/decreases (with respect to clause 1)]$_{\text{independent variable}}$ → [so the degree of ADVP2 increases/decreases (with respect to clause 2)]$_{\text{dependent variable}}$ in a monotonic way'

(13) NO VERB$_{C1}$-NO VERB$_{C2}$ CC meso-construction (SNC, Slovak Web 2011)

FORM:	[[tʃi:m] […]$_{\text{comparative element1}}$]$_{C1}$ [[ti:m] […]$_{\text{comparative element2}}$]$_{C2}$
MEANING:	'[As the degree of comparative element1 increases/decreases]$_{\text{independent variable}}$ → [so the degree of comparative element2 increases/decreases]$_{\text{dependent variable}}$ in a monotonic way'

(14) *ďalej*_NO VERB$_{C1}$-*viac*_NO VERB$_{C2}$ CC meso-construction (Slovak Web 2011)

FORM:	[[tʃi:m] [*ďalej*]$_{\text{comparative element1}}$]$_{C1}$ [[ti:m] [*viac*]$_{\text{comparative element2}}$]$_{C2}$
MEANING:	'[As the point in time increases]$_{\text{independent variable}}$ → [so the quantity increases]$_{\text{dependent variable}}$ in a monotonic way'

(15) ADVP_NO VERB$_{C1}$-ADVP_NO VERB$_{C2}$ CC meso-construction (SNC, Slovak Web 2011)

FORM:	[[tʃi:m] […]$_{\text{ADVP1}}$]$_{C1}$ [[ti:m] […]$_{\text{ADVP2}}$]$_{C2}$
MEANING:	'[As the degree of ADVP1 increases/decreases]$_{\text{independent variable}}$ → [so the degree of ADVP2 increases/decreases]$_{\text{dependent variable}}$ in a monotonic way'

In the following, we apply Hoffmann et al.'s (2019) methodology to a considerably larger data set that consists of well over 3,500 CC tokens from the Slovak Web 2011 corpus. In doing so, we intend to not only show how many meso-constructions underlie the Slovak CC construction but also provide fellow researchers with step-by-step instructions of how to investigate the relationship between different slots of a construction using corpus data. Our hope is that this will inspire more CxG-based investigations of various Slovak multi-slot constructions.

4. METHODOLOGY AND DATA

In this section, we present the methodology we used to detect the entrenchment of constructions. In CxG, any meaningful linguistic pattern that "occurs with sufficient frequency" is considered entrenched as a construction (Goldberg 2006, p. 5). Entrenchment refers to "the ongoing reorganization and adaptation of communicative knowledge" and has been described as "a lifelong cognitive reorganization process, the course and quality of which are conditioned by *exposure to and use of language*" (Schmid 2017, p. 3, emphasis by the authors). That is, constructions are 'learned' as a result of their repeated use by speakers (cf. Croft 2013, p. 224). Entrenched constructions, as noted by Goldberg, are stored as "a network of constructions" (2003, p. 219; cf. also Traugott – Trousdale 2013, p. 45) such as the one illustrated in Figure 2 above. Because entrenchment is dependent on usage frequency, corpus data is highly suitable for CxG-based investigations. This follows from Schmid's "from corpus-to-cognition principle" (2000, p. 39, as discussed in Hoffmann et al. 2019, p. 10), which posits that "text frequency 'instantiates entrenchment in the cognitive system'" (Hoffmann 2019, p. 18, citing Stefanowitsch – Flach 2017, p. 102; for more comments and suggestions cf. also Gries 2024).

The intention of this section is to provide fellow researchers with step-by-step instructions for the statistical analysis of corpus data in the context of usage-based CxG studies[23]. We hope that this will encourage others to model more multi-slot constructions in Slovak (and, of course, any other languages in which they may be interested).

4.1. COVARYING-COLLEXEME ANALYSIS

To test for cross-clausal associations between C1 and C2 and, thus, the entrenchment of meso-constructions, we employ a statistical test of relative frequencies known as **covarying-collexeme analysis** (Stefanowitsch and Gries

[23] For more details on how we proceeded, we also suggest taking a look at the coding guide that was used to train the student assistants who annotated the corpus data. The link to the OSF repository is provided in Section 4.2 below.

2005, p. 9 – 11; cf. also Gries – Stefanowitsch 2004; Gries 2019, 2022). While covarying-collexeme analysis was originally developed for testing the association of *lexical items* between two slots of a construction, Hoffmann et al. (2019) have shown that it can also be used to "extend collostructional methods by exploring the elements in slots on *a more schematic level*" (Martínez – Gries 2024, p. 164, emphasis by the authors). Thus, covarying-collexeme is really a "corpus-based method for investigating correlations between […] items occurring in two different slots of a grammatical construction" (Stefanowitsch – Gries 2005, p. 2); it has been described as a "standard test in Construction Grammar for the association of two constructional elements" (Hoffmann et al. 2019, p. 11). Covarying-collexeme analysis works with 2-by-2 contingency tables and a Fisher-Yates exact test[24]. The Fisher-Yates exact test is "very precise because it handles even small frequencies well" (Horsch 2021, p. 210, referring to Gries 2015, p. 313), which is particularly useful for investigating the lexical fillers in C1 and C2, many of which appeared with a very low frequency in the corpus data (see the note on hapaxes in section 5.1 below). We used the *Coll. analysis 3.2a* script for *R* (Gries 2007) to investigate the following linguistic variables: Lexical Filler, Filler Type and Deletion in C1 and C2 (i.e., the constructional slots), as well as their respective interactions.

Covarying-collexeme analysis works with relative frequencies to detect the entrenchment of constructions. This is based on the assumption that syntactic patterns become entrenched as conventionalized form-meaning pairings when they "occur with sufficient frequency" (Goldberg 2006, p. 5). While the literature does not agree on a particular frequency threshold that must be crossed for a pattern to be regarded entrenched as a construction, covarying-collexeme analysis uses a *p*-value of 0.05, a common convention in statistics. To distinguish between *degrees* of entrenchment, two more *p*-values are used as thresholds: $p<0.01$ and $p<0.001$. Assuming that the null hypothesis amounts to independence of occurrence and that "[t]he lower the *p*-value, the stronger the evidence against the null hypothesis" (Schmid – Küchenhoff 2013, p. 539), this means that lower *p*-values translate to higher entrenchment. Collostructional strength as output by the *Coll.analysis 3.2a* script thus indicates the "probability of obtaining [a particular] distribution or a more extreme one" (Schmid – Küchenhoff 2013, p. 535).

Note that the *Coll.analysis 3.2a* script does not provide *p*-values, but a value called 'collostructional strength'. This is a value that was originally intended to "gauge the degree of attraction" (Schmid – Küchenhoff 2013, p. 535) between constructions and lexemes but can be used to investigate more schematic 'slots' of a construction as well, as Hoffmann et al. (2019) first demonstrated. It is rendered as

[24] Such a "contingency or two-by-two table [is] familiar to many linguists from applications of the χ^2 test" (Schmid – Küchenhoff 2013, p. 534).

a negative log-transformed *p*-value (cf. Gries 2007), which facilitates the interpretation of significant results (i.e., $p<0.05$) (Stefanowitsch – Gries 2005, p. 7, Gries et al. 2005, p. 671). The *p*-value thresholds correspond to the following collostructional strengths:

- $p=0.001$ coll. str. = 3 (***)
- $p=0.01$ coll. str. = 2 (**)
- $p=0.05$ coll. str. = 1.30103 (*)

Following conventions in the literature, any rows in the results table that contain statistically significant results will be shaded gray (cf. Hoffmann et al. 2019) and "ranked from […] the highest to the lowest collostructional strength" (Horsch 2021, p. 212). Only pairs with an expected frequency of ≥5 will be displayed, "an assumption often required for goodness-of-fit tests such as the chi-square test" (Hoffmann et al. 2019, p. 15). This is done to preclude "the identification of pairs as significant that have an observed frequency of only 1" (Horsch 2023b, p. 46; cf. also Hoffmann et al. 2019, p. 15). Note that the results tables output by the *Coll.analysis 3.2a* script include more columns with information about attraction and repulsion (Gries 2015, p. 314), as well as so-called ΔP values that indicate how dependent one constructional element is on the other (Gries 2015, p. 314). These, however, will be excluded from results tables shown in the following, since they are of no relevance to this study.

4.2. CORPUS DATA

The corpus study is based on a data set consisting of 3,670 Slovak CC tokens, extracted from the Slovak Web 2011 corpus (skTenTen 11) using the Sketch Engine web interface[1] (Kilgariff et al. 2004, 2014). As implied by its name, the 540 million-word Slovak Web 2011 corpus consists of "texts collected from the Internet" (https://www.sketchengine.eu/sktenten-slovak-corpus/).

The Slovak Web 2011 corpus was queried in January 2021 using the regular expression (16), which was composed based on the template (6) for the Slovak CC construction as first suggested by Horsch (2019, p. 184). No filters or subcorpora were used. The regular expression (16) accounts for the lexically fixed clause-initial elements ([word="čím|tým"])[25], the obligatory comparative element slots ([tag="(k2*(d1|d2|d3)|k6*(d1|d2|d3)).*"]), and the flexibility in ordering clause and comparative element slots ([]{,6}).

[25] Note that this part of the regular expression also accounts for CCs with C2C1 order, which, however, are not part of the present investigation.

(16) **Regular expression (CQL) for extracting CCs from skTenTen11**[26]
[word="čím|tým"] []{,6} [tag="(k2*(d1|d2|d3)|k6*(d1|d2|d3)).*"] []*
[word="tým|čím"] []{,6} [tag="(k2*(d1|d2|d3)|k6*(d1|d2|d3)).*"] within <s/>

The query yielded a total of 4,998 tokens, which were manually annotated by two Slovak student assistants[27]. Each student assistant annotated half of the data set; interrater agreement was not measured. However, to ensure maximum consistency in annotation, following the methodology of Hoffmann et al. (2019, p. 11 – 12), the student assistants (both L1 speakers of Slovak) received extensive training based on sample data sets and were provided with a coding guide that contained detailed instructions. Moreover, the data set was reviewed by one of the authors after annotation by the student assistants. After identifying and discarding false positives[28], a data set[29] of 3,002 relevant tokens remained. These were then annotated for the following linguistic variables:

- Lexical Filler C1, Lexical Filler C2 (levels: e.g. *viac* 'more', *skôr* 'sooner', 405 levels in C1 and 616 levels in C2)
- Filler Type C1, Filler Type C2 (levels: AdjP, AdvP, NP, PP)
- Deletion C1, Deletion C2 (levels: Full Clause, Truncated)

Note that there are already corpus studies on the Slovak CC construction employing the same methodology (Horsch 2021, 2023a). However, this study was considerably smaller in scale, with just 442 CC tokens (Horsch 2021, p. 204). Now, with almost seven times as many tokens, it is possible to conduct a much more detailed investigation that will provide a considerably more fine-grained picture of the Slovak CC: As noted by Schmid and Küchenhoff (2013, p. 547), covarying-collexeme analysis yields more results the larger the tested sample is. Moreover, by applying the same methodology to a larger dataset it will be possible to see whether

[26] The regular expression retrieved strings with the clause-initial elements ([word="čím"] and [word="tým"]), each followed by a comparative adverb ((k2*(d1|d2|d3)) or adjective (k6*(d1|d2|d3)). To account for optional clauses, within <s/> was used; to account for variation in word order, the wild card []{,6} was inserted between the clause-initial element and the comparative words. Also note that the regular expression was formulated such that both C1C2 and C2C1 patterns were extracted (by using the pipe command "|"): [word="čím|tým"]).

[27] The student assistants were employed in the context of the research project "Comparing Comparative Correlatives: The more languages, the better" (DFG-HO 3904/5–2; 2021-24). This project was generously funded by the German Research Foundation (DFG).

[28] Including cases of C2 preceding C1, e.g. *Tie sa vyvíjajú* [*tým rýchlejšie,*]$_{C2}$ [*čím lepšie podmienky jej ľudský organizmus nechtiac vytvára*]$_{C1}$ 'These develop (the) faster, the better conditions the human organism creates unintentionally'.

[29] This data set, along with the results of covarying-collexeme analysis and the student assistant coding guide, is available in an OSF repository[2].

earlier studies can be replicated. Finally, it is an important contribution to overcoming the "big/small divide in corpus work" (Mair 2005, p. 24).

After annotation, the linguistic variables were subjected to covarying-collexeme analysis to detect statistically significant cross-clausal associations, following the methodology of Hoffmann et al. (2019). In addition to the individual variables, their interactions were also tested, for two reasons: (1) to test for prototype effects (for example, in Hoffmann et al.'s (2019) corpus study, the significantly associated LEXICAL FILLER association *more*$_{C1}$-*more*$_{C2}$ turned out to function "as a strong prototype" for the cross-clausal association ADVP$_{C1}$-ADVP$_{C2}$ (2019, p. 16)); (2) to obtain a more fine-grained picture of the CC meso-constructional network by testing for less schematic meso-constructions (as demonstrated by Horsch 2025, in press).

5. CORPUS STUDY RESULTS

The following sections (3.2.1 – 3.2.7) present the results of covarying-collexeme analysis, revealing a great number of statistically significant cross-clausal associations and thus, evidence for not only numerous meso-constructions, but also an extensive, elaborate network in which these meso-constructions are interconnected. Where applicable, results will be compared to Horsch's earlier (2021) corpus study. The results tables will show cross-clausal combinations ranked from highest to lowest collostructional strength, i.e., from highest to lowest statistical significance.

5.1. LEXICAL FILLER

The data set contained 1,165 different LEXICAL FILLER pairs, of which 985 (84.5%) were hapaxes, which can be seen as evidence of the high degree of variability and productivity of this constructional slot. It is interesting to note that similar numbers were observed in previous corpus studies on the CC construction in English (855/965=88.6%) (Hoffmann et al. 2019, p. 15) and Spanish (2,053/2,228=92.15%) (Horsch 2024, p. 229).

Testing so many different levels using covarying-collexeme analysis produced a large results table. For ease of exposition, the tables below are reduced in size by showing only pairs exhibiting attraction and a collostructional strength of ≥1.31, i.e. those that are of interest in the context of the present study.

The results Table 1 reveals 12 significant LEXICAL FILLER associations.

Table 1: Covarying-collexeme analysis of LEXICAL FILLER in C1 and C2 (expected frequency ≥5, Coll. strength ≥1.31, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
skôr 'sooner'	*lepšie* 'better'	461	841	295	129.15	69.46	***
skôr 'sooner'	*skôr* 'sooner'	461	141	99	21.65	51.04	***
ďalej 'further, longer'	*viac* 'more'	1133	342	228	129.08	30.03	***
horšie 'worse'	*lepšie* 'better'	48	841	44	13.45	19.92	***
ďalej 'further, longer'	*menej* 'less'	1133	263	167	99.26	18.14	***
ďalej 'further, longer'	*častejšie* 'more often'	1133	84	70	31.7	17.41	***
menej 'less'	*lepšie* 'better'	146	841	73	40.9	8.18	***
ďalej 'further, longer'	*horšie* 'worse'	1133	92	61	34.72	7.78	***
ďalej 'further, longer'	*ťažšie* 'heavier, more difficult'	1133	97	56	36.61	4.41	***
viac 'more'	*viac* 'more'	102	342	25	11.62	3.93	***
menej 'less'	*menej* 'less'	146	263	26	12.79	3.57	***
dlhšie 'longer'	*ťažšie* 'heavier, more difficult'	180	97	14	5.82	2.8	**
vyššie 'higher'	*lepšie* 'better'	72	841	27	20.17	1.31	*

Note that all three LEXICAL FILLER pairs determined by Horsch (*skôr*$_{C1}$-*lepšie*$_{C2}$ 'sooner$_{C1}$-better$_{C2}$', *ďalej*$_{C1}$-*viac*$_{C2}$ 'further/longer$_{C1}$-more$_{C2}$', *viac*$_{C1}$-*viac*$_{C2}$ 'more$_{C1}$-more$_{C2}$') as statistically significant (2021, p. 213) also emerged as significant here. Moreover, mirroring Horsch's (2021) results, *skôr*$_{C1}$-*lepšie*$_{C2}$ 'sooner$_{C1}$-better$_{C2}$' turned out to be the most significantly associated FILLER TYPE pair.

5.2. FILLER TYPE

Table 2 shows the FILLER TYPE pairs that emerged as statistically significant: NP$_{C1}$-NP$_{C2}$ (coll. str. 12.03***), AdvP$_{C1}$-AdvP$_{C2}$ (coll. str. 10.44***) and AdjP$_{C1}$-AdjP$_{C2}$ (coll. str. 8.71***). It is notable that all of them are symmetric. That is, C1 mirrors C2 and vice versa. While in Horsch's (2021) smaller study, no statistically significant FILLER TYPE pairs could be determined, the results of the present study are reminiscent of previous investigations of the CC in English (Hoffmann et al. 2019) and Spanish (Horsch 2024). In both of these studies, exactly the same three cross-clausal associations emerged as statistically very significant (i.e., with a collostructional strength of >3***). This suggests a strong cross-linguistic tendency towards symmetry on more abstract meso-constructional levels that Horsch has previously described as "'under-the-surface' parataxis" (2024, p. 232, referring to Hoffmann et al. 2019).

Table 2: Covarying-collexeme analysis of FILLER TYPE in C1 and C2 (expected frequency ≥5, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
NP	NP	167	194	38	10.79	12.03	***
ADVP	ADVP	2617	2282	2043	1989.34	10.44	***
ADJP	ADJP	210	524	71	36.66	8.71	***

5.3. DELETION

Table 3 shows the results for the variable DELETION. Two pairs emerged as statistically significant: FULL CLAUSE$_{C1}$-FULL CLAUSE$_{C2}$ (coll. str. 251.12***) and TRUNCATED$_{C1}$-TRUNCATED$_{C2}$ (coll. str. 251.12***). As was the case with FILLER TYPES, there were only symmetric pairs. Furthermore, these results reflect Horsch's study[30], where the same pairs also exhibited very significant (collostructional strength >3***) attraction (2021, p. 214).

Table 3: Covarying-collexeme analysis of DELETION in C1 and C2 (attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
FULL_CLAUSE	FULL_CLAUSE	1289	1381	1037	592.97	251.12	***
TRUNCATED	TRUNCATED	1713	1621	1369	924.97	251.12	***

5.4. LEXICAL FILLER:FILLER TYPE INTERACTION

To address Horsch's call for a "closer inspection of the interaction between FILLER TYPES and LEXICAL FILLERS" in the Slovak CC construction (2021, p. 217), we also tested for this interaction. This was done by "collapsing the factors FILLER TYPE C1 and DELETION C1 as well as FILLER TYPE C2 and DELETION C2" (Hoffmann et al. 2020, p. 203) before inputting them into the *R* script.

One reason for testing for LEXICAL FILLER:FILLER TYPE interactions[31] was to ensure that no significant FILLER TYPE associations (see Section 5.2) were due to "specific LEXICAL FILLER interactions that interact with the variable FILLER TYPE" (Hoffmann et al. 2019, p. 15). For example, the AdvP$_{C1}$-AdvP$_{C2}$ association (collostructional strength 10.44, see section 5.2) could be due to a high frequency of the LEXICAL FILLER pair *skôr*$_{C1}$- *lepšie*$_{C2}$ (collostructional strength 69.46***, see Section 5.1) that would simply have been annotated as ADVP in the data (Horsch 2024, p. 233).

Secondly, we wanted to obtain a more fine-grained picture of the meso-constructional network by testing for less schematic constructions. For example, the interaction *skôr*_ADVP$_{C1}$-*lepšie*_ADVP$_{C2}$, which turned out to be very significant (collostructional strength 74.02***, see Table 4 below), can be said to inherit from two more schematic meso-constructions: *skôr*$_{C1}$-*lepšie*$_{C2}$ (collostructional strength 69.46***) and ADVP$_{C1}$-ADVP$_{C2}$ (collostructional strength 10.44***). (We discuss this

[30] Note that Horsch called this variable "VERB PRESENCE" (2021, p. 214). To avoid confusion, we chose to use "DELETION" instead, in line with most other studies (e.g. Hoffmann 2019; Hoffmann et al. 2019, 2020; Horsch 2023a, 2023b, 2024).

[31] As one anonymous reviewer pointed out, it would also be an option to examine lexically partially specific cases, for example, LEXICAL FILLER in C1 and FILLER TYPE in C2. We suggest incorporating such cases in future studies.

network consisting of meso-constructions of varying degrees of schematicity in more detail in section 6).

As Table 4 shows, a total of 12 significant LEXICAL FILLER:FILLER TYPE interactions could be detected. A glance at the observed frequency column indicates that indeed, several LEXICAL FILLER pairs might be suspected to have acted as prototypes for the ADVP_{C1}-ADVP_{C2} pair. However, when comparing their observed frequencies to the frequency of ADVP_{C1}-ADVP_{C2}, it becomes clear that, in the words of Hoffmann et al., the ADVP_{C1}-ADVP_{C2} pair "cannot be fully explained by lexical effects" (2019, p. 16). For example, the most frequent pair, *skôr*_ADVP_{C1}-*lepšie*_ADVP_{C2} (288 tokens), makes up just about 14% of the ADVP_{C1}-ADVP_{C2} pair (2,043 tokens). Note that the other two FILLER TYPE pairs (ADJP_{C1}-ADJP_{C2} and NP_{C1}-NP_{C2}) do not show up at all in Table 4 below. We therefore conclude that lexical effects play a negligible role and that all FILLER TYPE pairs that covarying collexeme analysis detected (see Section 5.2) may be considered meso-constructions in their own right.

Table 4: Covarying-collexeme analysis of LEXICAL FILLER:FILLER TYPE in C1 and C2 (expected frequency ≥5, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
*skôr*_ADVP	*lepšie*_ADVP	461	776	288	119.17	74.02	***
*skôr*_ADVP	*skôr*_ADVP	461	140	99	21.5	51.51	***
*ďalej*_ADVP	*viac*_ADVP	1132	342	228	128.96	30.1	***
*horšie*_ADVP	*lepšie*_ADVP	47	776	44	12.15	22.41	***
*ďalej*_ADVP	*menej*_ADVP	1132	262	167	98.8	18.44	***
*ďalej*_ADVP	*častejšie*_ADVP	1132	83	69	31.3	17.07	***
*ďalej*_ADVP	*ťažšie*_ADJP	1132	36	31	13.57	8.63	***
*ďalej*_ADVP	*horšie*_ADJP	1132	39	32	14.71	7.84	***
*menej*_ADVP	*lepšie*_ADVP	145	776	68	37.48	7.77	***
*viac*_ADVP	*viac*_ADVP	102	342	25	11.62	3.93	***
*menej*_ADVP	*menej*_ADVP	145	262	26	12.65	3.64	***
*ďalej*_ADVP	*horšie*_ADVP	1132	53	29	19.98	2.09	***

5.5. LEXICAL FILLER:DELETION INTERACTION

Next, we tested the LEXICAL FILLER:DELETION interaction. Table 5 presents the results, a total of 20 cross-clausal associations ranging from collostructional strengths of 84.96*** (*skôr*_TRUNCATED_{C1}-*lepšie*_TRUNCATED_{C2}) to 1.8* (*dlhšie*_FULL CLAUSE_{C1}-*lepšie*_FULL CLAUSE_{C2}).

Table 5: Covarying-collexeme analysis of Lexical Filler:Deletion in C1 and C2 (expected frequency ≥5, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
*skôr*_TRUNCATED	*lepšie*_TRUNCATED	179	753	164	44.9	84.96	***
*skôr*_FULL_CLAUSE	*skôr*_FULL_CLAUSE	282	138	96	12.96	71.43	***
*ďalej*_TRUNCATED	*menej*_TRUNCATED	1073	133	129	47.54	53.5	***
*ďalej*_TRUNCATED	*viac*_TRUNCATED	1073	105	101	37.53	40.52	***
*horšie*_TRUNCATED	*lepšie*_TRUNCATED	44	753	42	11.04	22.86	***
*ďalej*_TRUNCATED	*častejšie*_TRUNCATED	1073	48	48	17.16	21.75	***
*menej*_TRUNCATED	*lepšie*_TRUNCATED	40	753	37	10.03	18.87	***
*ďalej*_TRUNCATED	*ťažšie*_TRUNCATED	1073	42	40	15.01	15.51	***
*viac*_TRUNCATED	*lepšie*_TRUNCATED	28	753	25	7.02	11.98	***
*ďalej*_TRUNCATED	*horšie*_TRUNCATED	1073	78	56	27.88	10.25	***
*viac*_FULL_CLAUSE	*viac*_FULL_CLAUSE	74	237	24	5.84	9.26	***
*skôr*_FULL_CLAUSE	*lepšie*_TRUNCATED	282	753	109	70.74	7.08	***
*dlhšie*_TRUNCATED	*lepšie*_TRUNCATED	24	753	18	6.02	6.44	***
*vyššie*_TRUNCATED	*lepšie*_TRUNCATED	34	753	21	8.53	5.24	***
*ďalej*_TRUNCATED	*viac*_FULL_CLAUSE	1073	237	112	84.71	4.01	***
*ďalej*_TRUNCATED	*rýchlejšie*_TRUNCATED	1073	14	12	5	3.76	***
*skôr*_FULL_CLAUSE	*lepšie*_FULL_CLAUSE	282	88	19	8.27	3.44	***
*dlhšie*_FULL_CLAUSE	*viac*_FULL_CLAUSE	156	237	22	12.32	2.34	**
*ďalej*_TRUNCATED	*častejšie*_FULL_CLAUSE	1073	36	20	12.87	1.94	*
*dlhšie*_FULL_CLAUSE	*menej*_FULL_CLAUSE	156	130	13	6.76	1.8	*

5.6. DELETION:FILLER TYPE INTERACTION

The third two-way interaction to be tested using covarying-collexeme analysis was Deletion:Filler Type. As Table 6 below shows, 11 pairs emerged as statistically significant, ranging in collostructional strength from 127.83*** (Full Clause_AdvP$_{C1}$-Full Clause_AdvP$_{C2}$) to 1.83* (Full Clause_NP$_{C1}$-Full Clause_AdjP$_{C2}$).

Table 6: Covarying-collexeme analysis of Deletion:Filler Type in C1 and C2 (expected frequency ≥5, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
FULL_CLAUSE_ADVP	FULL_CLAUSE_ADVP	1079	1108	704	398.24	127.83	***
TRUNCATED_ADVP	TRUNCATED_ADVP	1538	1174	861	601.47	85.72	***
TRUNCATED_ADVP	TRUNCATED_ADJP	1538	344	290	176.24	41.5	***
TRUNCATED_ADJP	TRUNCATED_ADJP	89	344	40	10.2	15.28	***
FULL_CLAUSE_ADJP	FULL_CLAUSE_ADJP	121	180	30	7.26	11.36	***
FULL_CLAUSE_ADVP	FULL_CLAUSE_ADJP	1079	180	107	64.7	10.5	***
TRUNCATED_ADVP	TRUNCATED_NP	1538	101	77	51.74	6.86	***
FULL_CLAUSE_ADJP	FULL_CLAUSE_ADVP	121	1108	63	44.66	3.43	***

FULL_CLAUSE_NP	FULL_CLAUSE_ADVP	88	1108	47	32.48	3	***
FULL_CLAUSE_ADVP	FULL_CLAUSE_NP	1079	93	48	33.43	2.92	**
FULL_CLAUSE_NP	FULL_CLAUSE_ADJP	88	180	11	5.28	1.83	*

5.7. DELETION:FILLER TYPE:LEXICAL FILLER INTERACTION

Finally, the three-way interaction DELETION:FILLER TYPE:LEXICAL FILLER was tested. As Table 7 shows, 20 pairs emerge as statistically significant. These range in collostructional strength from 90.36*** (TRUNCATED_ADVP_*skôr*$_{C1}$-TRUNCATED_ADVP_*lepšie*$_{C2}$) to 1.57* (FULL_CLAUSE_ADVP_*dlhšie*$_{C1}$-FULL_CLAUSE_ADVP_*menej*$_{C2}$).

Table 7: Covarying-collexeme analysis of DELETION:FILLER TYPE:LEXICAL FILLER in C1 and C2 (expected frequency ≥5, attraction only)

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Coll. str.	Sign.
TRUNCATED_ADVP_*skôr*	TRUNCATED_ADVP_*lepšie*	179	703	164	41.92	90.36	***
FULL_CLAUSE_ADVP_*skôr*	FULL_CLAUSE_ADVP_*skôr*	282	138	96	12.96	71.43	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADVP_*menej*	1072	133	129	47.49	53.56	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADVP_*viac*	1072	105	101	37.49	40.56	***
TRUNCATED_ADVP_*horšie*	TRUNCATED_ADVP_*lepšie*	43	703	42	10.07	25.37	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADVP_*častejšie*	1072	47	47	16.78	21.31	***
TRUNCATED_ADVP_*menej*	TRUNCATED_ADVP_*lepšie*	40	703	36	9.37	18.47	***
TRUNCATED_ADVP_*viac*	TRUNCATED_ADVP_*lepšie*	28	703	25	6.56	12.71	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADJP_*ťažšie*	1072	28	27	10	10.9	***
FULL_CLAUSE_ADVP_*viac*	FULL_CLAUSE_ADVP_*viac*	74	237	24	5.84	9.26	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADJP_*horšie*	1072	35	30	12.5	8.92	***
FULL_CLAUSE_ADVP_*skôr*	TRUNCATED_ADVP_*lepšie*	282	703	108	66.04	8.57	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADVP_*ťažšie*	1072	14	13	5	4.86	***
TRUNCATED_ADVP_*ďalej*	FULL_CLAUSE_ADVP_*viac*	1072	237	112	84.63	4.03	***
TRUNCATED_ADVP_*ďalej*	TRUNCATED_ADVP_*horšie*	1072	43	26	15.36	3.12	***
TRUNCATED_ADVP_*vyššie*	TRUNCATED_ADVP_*lepšie*	27	703	14	6.32	2.93	**
FULL_CLAUSE_ADVP_*dlhšie*	FULL_CLAUSE_ADVP_*viac*	151	237	21	11.92	2.18	**
TRUNCATED_ADVP_*ďalej*	FULL_CLAUSE_ADVP_*častejšie*	1072	36	20	12.86	1.95	*
FULL_CLAUSE_ADVP_*skôr*	FULL_CLAUSE_ADVP_*lepšie*	282	73	13	6.86	1.78	*
FULL_CLAUSE_ADVP_*dlhšie*	FULL_CLAUSE_ADVP_*menej*	151	129	12	6.49	1.57	*

6. RESULTS AND DISCUSSION

Using covarying-collexeme analysis, our corpus study of the Slovak CC construction has revealed a total of 80 statistically significant associations across C1 and C2. Based on this, we argue that these cross-clausal associations can be considered to "occur with sufficient frequency" (Goldberg 2006, p. 5) to be regarded as "conventionalized form-meaning pairings" (Hoffmann 2019, p. 6). Following the tenets of CxG, we assume these meso-constructions of varying degrees of

schematicity to form interconnected nodes of an elaborate network of linguistic knowledge (see Section 3) – in the words of Langacker, a "structured inventory of conventional linguistic units" (2008, p. 222). For example, less schematic meso-constructions, e.g. TRUNCATED_ADVP$_{C1}$-TRUNCATED_ADVP$_{C2}$ (see Section 5.6) **inherit** from more schematic ones (in this case, ADVP$_{C1}$-ADVP$_{C2}$ (Section 5.2) and TRUNCATED$_{C1}$-TRUNCATED$_{C2}$ (Section 5.3)). In turn, they **license** even less schematic meso-constructions, e.g. TRUNCATED_ADVP_*ďalej*$_{C1}$-TRUNCATED_ADVP_*viac*$_{C1}$ (Section 5.7). This is illustrated in Figure 3 below:

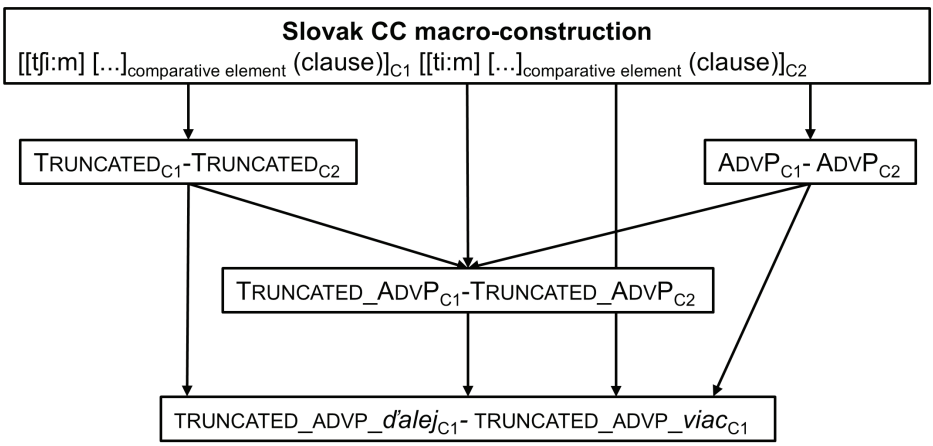


Figure 3: Excerpt of the Slovak CC network based on corpus study results (cf. Sections 5.1 – 5.7)

Note that as Figure 3 shows, any meso-construction can inherit from *and* simultaneously license multiple other meso-constructions. Thus, the least schematic meso-construction TRUNCATED_ADVP_*ďalej*$_{C1}$-TRUNCATED_ADVP_*viac*$_{C1}$ at the bottom of Figure 3 inherits from all other meso-constructions as well as the macro-construction.

On a final note, most of the meso-constructions (cf. Sections 5.2 – 5.7) are symmetric in nature, i.e., C1 formally mirrors C2, and vice versa. For example, the only FILLER TYPE associations that emerged as statistically significant were NP$_{C1}$-NP$_{C2}$, AdvP$_{C1}$-AdvP$_{C2}$ and AdjP$_{C1}$-AdjP$_{C2}$ (section 5.2); the only DELETION associations to do so were FULL CLAUSE$_{C1}$-FULL CLAUSE$_{C2}$ and TRUNCATED$_{C1}$-TRUNCATED$_{C2}$ (Section 5.3). This also concerns the interactions that were tested. For example, of the five most significant DELETION:FILLER TYPE interactions, only one (TRUNCATED_ADVP$_{C1}$-TRUNCATED_ADJP$_{C2}$) was not symmetric (Section 5.6).

These results reflect findings of previous corpus studies on Slovak (Horsch 2021, 2023a) and other languages (Hoffmann et al. 2019; Hoffmann et al. 2020;

Horsch 2023b (English) and Horsch 2024 (Spanish)). The strong preference for symmetric form regarding cross-clausal associations has previously been attributed to the cognitive principle of iconicity (Bybee 2012), which posits that "the meaning of constructions is encoded in their form" (Horsch 2021, p. 219). That is, the **symmetric**, parallel change-over-time semantics[32] (see Section 3) appear to influence speakers' choice of Filler Types and Deletion patterns.

Of course, in this context it must be noted that the CC also encodes **asymmetric** semantics (as discussed in Section 3). That is, we can expect a competition between formal symmetry and asymmetry. This is evident, as Horsch has previously argued, from the form of the phonologically/lexically fixed clause-initial elements *čím... tým...* that, unlike in English, clearly "differ from each other in form" (Horsch 2021, p. 197). For a more detailed discussion of the implications of these results – that is, cross-clausal associations, meso-constructions and the formal encoding of symmetric and asymmetric semantics – we would like to refer the reader to two recent corpus investigations of the Slovak CC (Horsch 2021; 2023a).

7. CONCLUSION AND ROADMAP

As our corpus study has demonstrated, the mental constructional network underlying the Slovak CC construction is, to return to Traugott and Trousdale's quote, "baroque, involving massive redundancy and vastly rich detail" (2013, p. 53). That is, many empirically attested cross-clausal associations that, according to the basic tenets of Construction Grammar, are conventionalized form-meaning pairings – meso-constructions that are formally related to each other. Each meso-construction may license and inherit from a multitude of other meso-constructions, of which the corpus study has identified 80 in total. This provides us with a small glimpse of the Slovak construct-i-con, i.e., "[t]he totality of our knowledge of language" (Goldberg 2003, p. 223).

In the following, we intend to lay out a **roadmap** for further CxG-based research on Slovak. The ultimate goal is a Slovak construct-i-con, modeled after similar projects for other languages. To achieve this, we suggest the following short-, medium- and long-term objectives.

In the short term, we would like to encourage fellow researchers to start exploring more individual Slovak constructions. We welcome any studies applying our methodology for investigating entrenchment in constructional slots, i.e., by

[32] Recall that these symmetric semantics, as discussed in section 3, are monotonic in nature, that is, "the meaning of each individual clause denotes an increase or decrease over time" (Hoffmann et al. 2019, p. 26). That is, each clause separately encodes a monotonic (but not necessarily linear) change over the same time period, which is then reflected in symmetric structures (principle of iconicity). This aspect of the CC semantics does not cover the conceptual relationship between the two events, which is causal in nature and is covered by the asymmetric cause-effect meaning.

using covarying-collexeme analysis. In this regard, we specifically call for studies of all kinds of syntactic patterns. We have shown how this may be done by investigating the linguistic variables Filler Type and Deletion (see Sections 4 and 5). Any construction, no matter whether fully substantive (e.g. collocations), partly substantive and partly schematic (e.g. the Comparative Correlative) or fully schematic (e.g. the ditransitive construction), is suitable for this kind of investigation. The same holds for what in traditional terms is called the 'core' and the 'periphery'; both are worthy of our attention. This is fully in line with the CxG approach; as noted by Goldberg, "the constructionist approach does not assume that language should be divided up into 'core' grammar and the to-be-ignored 'periphery'" (2003, p. 223). Rather, CxG distinguishes itself by its "holistic framework: no one level of grammar is autonomous, or 'core'" (Traugott – Trousdale 2013, p. 3).

The more Slovak constructions that are 'mapped out' in this way, the closer we will come to the ultimate goal of a Slovak construct-i-con (see below). In this regard, there is much work to be done: The Swedish construct-i-con, for example, has 441 entries[3] and the Russian construct-i-con comprises over 3600 constructions[4]. However, we believe there is great potential, especially because the data on which such investigations are based – that is, corpus data – is readily available in Slovakia. The Slovak National Corpus department at the Slovak Academy of Sciences provides access to a wealth of corpora that cover many different registers, periods and even dialects of Slovak[5]. This will also address a more general issue in CxG-based research: As Fried has noted, "syntactic patterning […] has been generally left just about untouched in Slavic linguistics" (2017, p. 241).

Of course, "[c]orpus data can [...] be said to offer one, but obviously not the only window onto the entrenchment of constructions" (Hoffmann et al. 2019, p. 10). In fact, as Horsch has noted, "even the largest corpora have an inherent flaw: As language *samples* they are by definition *finite*. However, [language] is by definition *infinite*" (2023c, p. 60, emphasis Horsch). This flaw, referred to as "negative data problem" (Hoffmann 2011, p. 1), cannot be resolved with ever larger corpora, since they will always be, by definition, finite samples of language (Horsch 2023c, p. 61). There is also the "positive data problem", which Hoffmann summarizes as follows: "just because a construction appears in a corpus it does not automatically follow that it is grammatical" (2011, p. 1). Put more succinctly, absence of evidence is not evidence of absence, and vice versa.

This is why we recommend using a corroborative approach that employs other types of data to complement corpus studies. The importance of such "complementary" approaches (McEnery – Wilson 2001, p. 16) has been acknowledged before. Szmrecsányi et al., for example, specifically suggest "spot-check[ing] the cognitive robustness of […] corpus-derived probabilities via rating experiments" (2016,

p. 134). We believe that such rating experiments, if carried out following good scientific practice, are highly suitable for complementing corpus data. As noted by Hoffmann, "several studies have [...] shown that the results from experimental and corpus-based studies often converge" (2019, p. 18). We therefore explicitly recommend psycholinguistic experimental studies to corroborate corpus-based claims about the entrenchment of constructions (as suggested by Hoffmann et al. 2019, p. 10; cf. also Horsch (2023c)).

Given that CxG-based investigations of Slovak are still relatively uncommon, we encourage a contrastive approach that uses English as a basis, because it is "the most widely analyzed language" in CxG (Boas 2010, p. 3). This will also address the "striking absence of cross-linguistic generalizations" in CxG (Boas 2010, p. 2) and contribute to CxG's "aspirations toward universal applicability" (Fried 2017, p. 249).[33] We therefore encourage researchers to present their studies at domestic and international conferences alike.

In the medium term, we recommend identifying and describing interactions between the individual constructions that will have been explored at this point. Again, this will be an important step towards the ultimate goal of mapping the Slovak construct-i-con (see below) in its entirety. We also call for a full CxG-based syntactic analysis of an entire Slovak sentence to demonstrate the feasibility of this approach, following Ziem et al.'s example (2014), who did the same for a German sentence. In this context, we would like to encourage colleagues from other theoretical backgrounds to contribute by analyzing the same sentence against their specific theoretical background – again, following an example from German (cf. Hagemann – Staffeldt 2014).

At this stage, researchers should also visualize their results to illustrate meso-constructional networks, following Hoffmann et al. (2019) and Horsch (2023a, 2023b). For example, an expanded version of Figure 2 in Section 6 that incorporates *all* 80 meso-constructions that we have identified in Section 5.

Finally, **in the long term**, we call for creating a construct-i-con, following the methodology that has been employed for other languages, such as English (Fillmore et al. 2012), Swedish (Lyngfeldt 2018), Russian (Janda et al. 2018), and other languages. Creating a construct-i-con is a large and complicated task, both methodologically and financially. Diessel (2023) discusses some of the key challenges that researchers face when compiling a construct-i-con, while Boas (2025b) compares different empirical methodologies for creating construction

[33] Of course, any other language can be taken as a basis for such contrastive investigations, too. The reason why we are suggesting to start with English is because at this point, constructional research on English is the most prominent and there are the greatest number of publications couched in constructional terms available.

entries. At the heart of the matter is the open question of how many constructions there are in a given language, how many different types of constructions there are in a language, and how construction entries are organized in a construct-i-con, both internally (what types of information should be stored in a construction entry?) and externally (how are construction entries organized vis-a-vis each other?). When compiling a construct-i-con for Slovak, it is also not clear how many of the organizational principles of construct-i-cons of other languages can be applied to Slovak or not. Once there is a significant number of construction entries stored in the construct-i-con, we then expect it to be possible to conduct systematic investigations of how different types of construction entries interact in order to license full sentences. This interaction of different constructions, i.e. combining them, is then what constructional researchers would call the 'grammar' part of CxG.

In summary, what we have shown in this case study is that Slovak lends itself very well to CxG-based studies. Our detailed corpus study has demonstrated the feasibility of a CxG-based analysis of syntactic patterns in Slovak based on corpus data. It has also shown that CxG is an empirically falsifiable theory. While corpus data play a central role in this regard, as we have argued above, it is desirable to incorporate other types of data due to the shortcomings of corpora. We expect future constructional research on Slovak to contribute insight into the nature of the Slovak construct-i-con as well as into how construction entries are combined to license complete sentences (the 'grammar part'). We would also like to suggest that the data and results obtained by such constructional research in the future will not only be of interest to constructional researchers, but also to the broader discipline of linguistics, including other theoretical frameworks that might benefit from constructional insights. Finally, given how constructional insights have been applied to the teaching and learning of foreign languages (see, e.g. Herbst 2016 and Boas 2022), we also fully expect for constructional insights to be applied to the teaching and learning of Slovak.

Acknowledgements

The paper has been written in the context of the research project "Weight Effects in English and Slovak" (2024-26), generously funded by the EU´s *NextGenerationEU* program through the Recovery and Resilience Plan for Slovakia (project no. 09I03-03-V04-00713).

We are grateful to the German Research Foundation (DFG), which has generously supported us financially in the context of the research project "Comparing Comparative Correlatives: The more languages, the better" (2021-24). Funding covered a part-time research assistant position, student assistant positions and access to the Slovak Web 2011 corpus on Sketch Engine.

We would like to thank our two student assistants, Michaela and Helena, who diligently coded thousands of corpus tokens.

Finally, we express our gratitude to two anonymous reviewers, whose comments have greatly improved this paper.

References

AMBRIDGE, B. – GOLDBERG, A. E. (2008): The island status of clausal complements: Evidence in favor of an information structure explanation. In: *Cognitive Linguistics*, 19/3, 357 – 389.

ACKERMAN, F. – WEBELHUTH, G. (1998): *A Theory of Predicates*. Stanford: CSLI Publications.

AUDRING, J. – JACKENDOFF, R. (2025): Construction Morphology and Relational Morphology. In: Fried, M. – Nikiforidou, K. (eds.): *The Cambridge Handbook of Construction Grammar*. Cambridge: Cambridge University Press, 101 – 128.

BARLOW, M. – KEMMER, S. (eds.) (2000): *Usage-based Approaches*. Stanford: CSLI Publications.

BERGEN, B. – CHANG, N. (2013): Embodied Construction Grammar. In: Hoffmann, Th. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar.* Oxford: Oxford UP, 191 – 210.

BETÁKOVÁ, V. (1955). Poznámky k Učebnici jazyka slovenského pre štvorročné odborné školy. In: *Slovenská reč*, 20/5, 313 – 323.

BOAS, H. C. (2010): Comparing constructions across languages. In: Boas, H. C. (ed.): *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins, 1 – 20.

BOAS, H. C. (2011): Zum Abstraktionsgrad von Resultativkonstruktionen. In: Engelberg, S. – Proost, K. – Holler, A. (eds.): *Sprachliches Wissen zwischen Lexikon und Grammatik.* Berlin/New York: Mouton de Gruyter, 37 – 69.

BOAS, H. C. (2013): Cognitive Construction Grammar. In: Hoffmann, Th. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar.* Oxford: Oxford UP, 233 – 254.

BOAS, H. C. (2017): Computational Resources: FrameNet and Constructicon. In: Dancygier, B. (ed.): *The Cambridge Handbook of Cognitive Linguistics.* Cambridge: Cambridge UP, 549 – 573.

BOAS, H. C. (2020): Constructions in English Grammar. In: Aarts, B. – McMahon, A. – Hinrichs, L. (eds.): *The Handbook of English Linguistics.* Oxford: Wiley, 277 – 297.

BOAS, H. C. (2021): Construction Grammar and Frame Semantics. In: Wen, X. – Taylor, R. J. (eds.): *The Routledge Handbook of Cognitive Linguistics.* New York/London: Routledge, 43 – 77.

BOAS, H. C. (ed.) (2022): *Directions for Pedagogical Construction Grammar*. Berlin/Boston: De Gruyter.

BOAS, H. C. (2025a): What happened to Frame Semantics? In: *Journal of English Linguistics*, 41, 1 – 55.

BOAS, H. C. (2025b): Constructional Syntax. In: Fried, M. – Nikiforidou, K. (eds.): *Cambridge Handbook of Construction Grammar*. Cambridge: Cambridge UP, 44 – 70.

BOAS, H. C. – DUX, R. (2017): From the Past into the Present: From Case Frames to Semantic Frames. In: *Linguistics Vanguard*, 2017, 1 – 14.

BOAS, H. C. – GONZÁLVEZ-GARCÍA, F. (eds.) (2014): *Romance Perspectives on Construction Grammar*. Amsterdam/Philadelphia: John Benjamins.

BOAS, H. C. – LEINO, J. – LYNGFELT, B. (2024): Constructionist Views on Construction Grammar. In: *Constructions and Frames*, 16/2, 1 – 22.

BOAS, H. C. – LYNGFELT, B. – TORRENT, T. T. (2019): Framing Constructicography. In: *Lexicographica*, 35/1, 41 – 95.

BOAS, H. C. – RUPPENHOFER, J. – BAKER, C. (2024): FrameNet at 25. In: *International Journal of Lexicography*, 37/3, 263 – 284.

BOAS, H. C. – ZIEM, A. (eds.) (2018a): *Constructional Approaches to Syntactic Structures in German*. Berlin/Boston: De Gruyter Mouton.

BOAS, H. C. – ZIEM, A. (2018b): Approaching German syntax from a constructionist perspective. In: Boas, H. C. – Ziem, A. (eds.): *Constructional Approaches to Syntactic Structures in German.* Berlin/Boston: De Gruyter Mouton, 1 – 46.
BOUVERET, M. – LEGALLOIS, F. (eds.) (2012): *Constructions in French*. Amsterdam/Philadelphia: John Benjamins.
BUSSE, D. (2012): *Frame-Semantik: Ein Kompendium*. Berlin/New York: De Gruyter.
BUZÁSSYOVÁ, K. – JAROŠOVÁ, A. (eds.) (2006): *Slovník súčasného slovenského jazyka*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
BYBEE, J. (2010): *Language, usage and cognition*. Cambridge UP.
BYBEE, J. (2012): Domain-general processes as the basis for grammar. In: Tallerman, M. – Gibson, K. R. (eds.): *The Oxford Handbook of Language Evolution.* Oxford: Oxford UP, 528 – 536.
CAPPELLE, B. (2024): *Can Construction Grammar Be Proven Wrong?* Oxford: Oxford UP.
CLAUSNER, T.C. – CROFT, W. (1997): Productivity and schematicity in metaphors. In: *Cognitive Science*, 21/3, 247 – 282.
COUSSÉ, E. – HÖDER, S. – LYNGFELT, B. – PRENTICE, J. (2023): *Constructional Approaches to Nordic Languages*. Amsterdam/Philadelphia: John Benjamins.
CROFT, W. (2001): *Radical Construction Grammar*. Oxford: Oxford UP.
CROFT, W. (2013): Radical Construction Grammar. In: Hoffmann, Th. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar.* Oxford: Oxford UP, 439 – 468.
CROFT, W. (2025): Constructions in typological and cross-linguistic contexts. In: Fried, M. – Nikiforidou, K. (eds.): *Cambridge Handbook of Construction Grammar*. Cambridge: Cambridge UP, 44 – 70.
CULICOVER, P. W. – JACKENDOFF, R. (1999): The View from the Periphery: The English Comparative Correlative. In: *Linguistic Inquiry*, 30/4, 543 – 571.
CULICOVER, P. W. – JACKENDOFF, R. (2005): *Simpler syntax*. Oxford: Oxford UP.
DEN DIKKEN, M. (2005): Comparative Correlatives Comparatively. In: *Linguistic Inquiry*, 36/4, 497 – 532.
DIESSEL, H. (2019): *The Grammar Network*. Cambridge: Cambridge UP.
DIESSEL, H. (2023): *The Constructicon*. Oxford: Oxford UP.
DUX, R. (2020): *Frame-constructional Verb Classes: Change and Theft Verbs in English and German*. Amsterdam/Philadelphia: John Benjamins.
ENDRESEN, A. – ZHUKOVA, V. – MORDASHOVA, D. – RAKHILINA, E. LYASHEVSKAYA, O. (2020): Русский конструктикон: Новый лингвистический ресурс, его устройство и специфика [The Russian Constructicon: A new linguistic resource, its design and key characteristics]. In: *Computational linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue-2020"*, 19, 226 – 241.
FILLMORE, C. J. (1968): The case for case. In: Bach, E. – Harms, R. T. (eds.): *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston, 1 – 88.
FILLMORE, C. J. (1982): Frame Semantics. In: Linguistic Society of Korea (ed.): *Linguistics in the Morning Calm*. Seoul: Hanshin, 111 – 138.
FILLMORE, C. J. (1985a): Syntactic intrusions and the notion of grammatical construction. In: *Berkeley Linguistic Society*, 11, 73 – 86.
FILLMORE, C. J. (1985b): Frames and the Semantics of Understanding. In: *Quadernie di Semantica*, 6/2, 222 – 254.
FILLMORE, C. J. (1986): Pragmatically controlled zero anaphora. In: *Berkeley Linguistics Society*, 12, 95 – 107.
FILLMORE, C. J. (1988): The Mechanisms of "Construction Grammar. In: *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35 – 55.
FILLMORE, C. J. (1989): Grammatical construction theory and the familiar dichotomies. In: *North-Holland Linguistic Series: Linguistic Variations*, 54, 17 – 38.

FILLMORE, C. J. (1990): Epistemic stance and grammatical form in English conditional sentences. In: *Chicago Linguistic Society*, 26, 137 – 162.
FILLMORE, C. J. (1999): Inversion and constructional inheritance. In: Webelhuth, G. – Koenig, J.-P. – Kathol, A. (eds.): *Lexical and Constructional Aspects of Linguistic Explanation*. Stanford: CSLI Publications, 113 – 128.
FILLMORE, C. J. (2013): Berkeley Construction Grammar. In: Hoffmann, Th. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar*. Oxford: Oxford UP, 111 – 132.
FILLMORE, C. J. – ATKINS, B. T. S. (1992): Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. In: Lehrer, A. – Kittay, E. (eds.): *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale: Erlbaum, 75 – 102.
FILLMORE, C. J. – BAKER, C. (2010): A frames approach to semantic analysis. In: Heine, B. – Narrog, H. (eds.): *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford UP, 13 – 340.
FILLMORE, C. J. – BAKER, C. – CRONIN, B. (2003): The structure of the FrameNet database. In: *International Journal of Lexicography*, 16/3, 281 – 296.
FILLMORE, C. J. – KAY, P. (1993): *Construction Grammar Course Book*. UC Berkeley: Department of Linguistics.
FILLMORE, C. J. – KAY, P. – O'CONNOR, M.-C. (1988): Regularity and idiomaticity in grammatical constructions: The case of 'let alone.' In: *Language*, 64, 501 – 538.
FILLMORE, C. J. – LEE-GOLDMAN, R. – RHOMIEUX, R. (2012): The FrameNet Constructicon. In: Boas, H. C. – Sag, I. A. (eds.): *Sign-based Construction Grammar*. Stanford: CSLI Publications, 309 – 372.
FRIED, M. (2007): Constructing grammatical meaning: isomorphism and polysemy in Czech reflexivization. In: *Studies in Language*, 31/4, 721 – 764.
FRIED, M. (2013): Pojem konstrukce v konstrukční gramatice. In: *Časopis pro moderní filologii*, 95/1, 9 – 27.
FRIED, M. (2017): Construction Grammar in the Service of Slavic Linguistics, and Vice Versa. In: *Journal of Slavic Linguistics*, 25/2, 241 – 276.
FRIED, M. (2010): Between verb semantics and interpersonal meanings in participant realization: a constructional account of 'ethical' datives. Keynote Lecture at the *International workshop on Variation and Change in Argument Realization*. Naples, Italy.
FRIED, M. – ÖSTMANN, J.-O. (2004): Construction Grammar: a thumbnail sketch. In: Fried, M. – Östman, J.-O. (eds.): *Construction Grammar in a cross-language perspective.* Amsterdam: John Benjamins, 11 – 86.
FRIED, M. – LIPSKÁ, K. (2020): Relativizační strategie v češtině: konstrukčně-gramatický přístup. In: *Naše řeč*, 103/1 – 2, 37 – 54.
GOLDBERG, A. E. (1995): *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: The University of Chicago Press.
GOLDBERG, A. E. (2003): Constructions: A new theoretical approach to language. In: *TRENDS in Cognitive Sciences*, 7/5, 219 – 224.
GOLDBERG, A. E. (2006): *Constructions at Work: The nature of generalization in language*. Oxford: Oxford UP.
GOLDBERG, A. E. (2019): *Explain me this*. Princeton: Princeton UP.
GRIES, S. Th. (2007): *Coll.analysis 3.2a*. A program for R for Windows 2.x.
GRIES, S. Th. (2015): Quantitative designs and statistical techniques. In: Biber, D. – Reppen, R. (eds.): *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge UP, 50 – 71.
GRIES, S. Th. (2019): 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistic measures). In: *International Journal of Corpus Linguistics*, 24/3, 386 – 412.
GRIES, S. Th. (2022): Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more. In: *Research Methods in Applied Linguistics*, 1/1, 1 – 19.

GRIES. S. Th. (2024): *Frequency, Dispersion, Association, and Keyness: Revising and tupleizing corpus-linguistic measures*. Amsterdam: John Benjamins.
GRIES, S. Th. – HAMPE, B. – SCHÖNEFELD, D. (2005): Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. In: *Cognitive Linguistics*, 16/4, 635 – 676.
GRIES, S. Th. – STEFANOWITSCH, A. (2004): Extending collostructional analysis. In: *International Journal of Corpus Linguistics*, 9/1, 97 – 129.
HAGEMANN, J. – STAFFELDT, S. (eds.) (2014): *Syntaxtheorien. Analysen im Vergleich*. Tübingen: Stauffenburg.
HENNECKE, I. – WIESINGER, E. (eds.) (2023): *Constructions in Spanish*. Amsterdam/Philadelphia: John Benjamins.
HERBST, Th. (2016): Foreign language learning is construction learning – what else? Moving towards Pedagogical Construction Grammar. In: de Knop, S. – Gilquin, G. (eds.): *Applied Construction Grammar*. Berlin/Boston: de Gruyter, 21 – 52.
HILPERT, M. (2025): Frequency: Psychological and Methodological Considerations. In: Fried, M. – Nikiforidou, K. (eds.): *The Cambridge Handbook of Construction Grammar*. Cambridge: Cambridge University Press, 149 – 170.
HOFFMANN, Th. (2011): *Preposition Placement in English: A Usage-based Approach*. Cambridge: Cambridge UP.
HOFFMANN, Th. (2017): Construction Grammar as Cognitive Structuralism: The interaction of constructional networks and processing in the diachronic evolution of English comparative correlatives. In: *English Language and Linguistics*, 21/2, 349 – 373.
HOFFMANN, Th. (2019): *English Comparative Correlatives: Diachronic and Synchronic Variation at the Lexicon-Syntax Interface*. Cambridge: Cambridge UP.
HOFFMANN, Th. (2020): *Construction Grammar*. Oxford: Oxford UP.
HOFFMANN, Th. – HORSCH, J. – BRUNNER, Th. (2019): The More Data, The Better: A Usage-based Account of the English Comparative Correlative Construction. In: *Cognitive Linguistics*, 30/1, 1 – 36.
HOFFMANN, Th. – BRUNNER, Th. – HORSCH, J. (2020): English Comparative Correlative Constructions: A Usage-based account. In: *Open Linguistics*, 6/1, 196 – 215.
HORSCH, J. (2019): Slovak Comparative Correlatives: New Insights. In: *Jazykovedný časopis*, 70/2, 180 – 190.
HORSCH, J. (2021): Slovak Comparative Correlatives: A Usage-based Construction Grammar Account. In: *Constructions and Frames*, 13/2, 193 – 229.
HORSCH, J. (2023a): From corpus data to constructional networks: Analyzing language with the Usage-based Construction Grammar framework. In: *Jazykovedný časopis*, 74/3, 701 – 740.
HORSCH, J. (2023b): *The comparative correlative construction in World Englishes: a usage-based construction grammar approach*. PhD Thesis. Eichstätt (Germany): University of Eichstätt-Ingolstadt.
HORSCH, J. (2023c): Corroborating corpus data with elicited introspection data: A case study. In: *Jazykovedný časopis*, 74/2, 60 – 69.
HORSCH, J. (2024): Cuanto(s) más datos, (tanto) mejor: A corpus-based study of the Spanish comparative correlative construction. In: *Review of Cognitive Linguistics*, 11/1, 204 – 257.
HORSCH, J. (in press): Nodes, networks and redundancies: A constructional account of the Spanish Comparative Correlative construction. In: *Edited Volume inspired by the 2023 Hispanic Linguistics Symposium* (Issues in Hispanic and Lusophone Linguistics).
IVANOVÁ, M. (2021): Láska-neláska, idú zlé časy. Koncesívne štruktúry vo svetle konštrukčnej gramatiky. In: *Slovenská reč*, 86/2, 143 – 161.
JANDA, L. A. – ENDRESEN, A. – ZHUKOVA, V. – MORDASHOVA, D. – RAKHILINA, E. (2020): How to build a constructicon in five years: The Russian Example. In: Brisard, F. – Colleman, T. –

de Wit, A. – Enghels, R. – Koutsoukos, N. – Mortelmans, T. – Sol Sansiñena, M. (eds.): *The Wealth and Breadth of Construction-Based Research [a thematic issue of Belgian Journal of Linguistics 34]*, 162 – 175.
JANDA, L. A. – LYASHEVSKAYA O. – NESSET, T. – RAKHILINA, E. – TYERS, F. M. (2018): Chapter 6. A Constructicon for Russian: Filling in the Gaps. In: Lyngfelt, B. – Borin, L. – Ohara, K. – Torrent, T. T. (eds.): *Constructicography: Constructicon development across languages*. Amsterdam: John Benjamins, 165–181.
KAČALA, J. – PISÁRČIKOVÁ, M. – POVAŽAJ, M. (eds.) (2003): *Krátky slovník slovenského jazyka*. Bratislava: Veda, Vydavateľstvo Slovenskej akadémie vied.
KAY, P. – FILLMORE, C. J. (1999): Grammatical constructions and linguistic generalizations: The 'What's X doing Y?' Construction. In: *Language*, 75, 1 – 33.
KILGARRIFF, A. – BAISA, V. – BUŠTA, J. – JAKUBÍČEK, M. – KOVÁŘ, V. – MICHELFEIT, J. – RYCHLÝ, P. – SUCHOMEL, V. (2014): The Sketch Engine: ten years on. In: *Lexicography ASIALEX*, 1, 7 – 36.
KILGARRIFF, A. – SMRŽ, P. – TUGWELL, D. (2004): The Sketch Engine. In: Williams, G. – Vessier, S. (eds.): *EURALEX 2004: Proceedings of the eleventh EURALEX international congress,* 105 – 116.
KUZAR, R. (2010): *Sentence patterns in English and Hebrew*. Amsterdam/Philadelphia: John Benjamins.
LANGACKER, R. W. (2008): *Cognitive Grammar: A Basic Introduction*. New York: Oxford UP.
LAPORTE, S. – LARSSON, T. – GOULART, L. (2021): Testing the Principle of No Synonymy across levels of abstraction: A constructional account of subject extraposition. In: *Constructions and Frames*, 13/2, 230 – 262.
LASCH, A. (2017): *Nonagentive Konstruktionen des Deutschen*. Berlin/Boston: De Gruyter.
LECLERCQ, B. (2024): *Linguistic Knowledge and Language Use*. Oxford: Oxford UP.
LECLERCQ, B. – MORIN, C. (2023): No equivalence: A new principle of no synonymy. In: *Constructions*, 15/1, 1 – 16.
LECLERCQ, B. – MORIN, C. – PIJPOPS, D. (2024): Modelling the Principle of No Equivalence. In: *Book of Abstracts: The 13th International Conference on Construction Grammar (ICCG),* 90 – 91.
LEVIN, B. – RAPPAPORT HOVAV, M. (2005): *Argument Realization*. Cambridge: Cambridge UP.
LYNGFELT, B. (2018): Introduction: Constructions and constructicography. In: Lyngfelt, B. – Borin, L. – Ohara, K. – Torrent, T. T. (eds.): *Constructicography. Constructicon development across languages.* Amsterdam/Philadelphia: John Benjamins, 1 – 18.
LYNGFELT, B. – BORIN, L. – OHARA, K. – TORRENT, T. T. (2018): *Constructicography: Constructicon development across languages*. Amsterdam/Philadelphia: John Benjamins.
MAIR, Ch. (2005): Tracking Ongoing Grammatical Change and Recent Diversification in Present-day Standard English: The Complementary Role of Small and Large Corpora. In: Schwitalla, J. – Wegstein, W. (eds.): *Korpuslinguistik deutsch: synchron - diachron - kontrastiv: Würzburger Kolloquium 2003*. Tübingen: Max Niemeyer Verlag, 9 – 26.
MARSINOVÁ, M. (1955): Spracovanie gramatických kategórií v normatívnom slovníku. In: *Slovenská reč*, 20/1, 29 – 39.
MARTÍNEZ, J. O. – GRIES, S. Th. (2024): If Not For-If It Weren't/Wasn't For Counterfactual Constructions: A Multivariate Extension of Collostructional Analysis. In: *Cognitive Semantics*, 10, 158 – 189.
MATSUMOTO, Y. (2025): Frame Semantics. In: *Cambridge Handbook of Construction Grammar*. Cambridge: Cambridge UP, 22 – 43.
MCENERY, T. – WILSON, A. (2001): *Corpus Linguistics*. Edinburgh: Edinburgh UP.
MEDINA, P. G. (2017): The English conative as a family of constructions. Towards a usage-based approach. In: de Mendoza Ibanez, F. J. R. – Oyon, A. L. – Sobrino, P. P. (eds.): *Constructing Families of Constructions*. Amsterdam/Philadelphia: John Benjamins, 277 – 300.

MICHAELIS, L. (2013): Sign-Based Construction Grammar. In: Hoffman, T. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar* Oxford: Oxford UP, 133 – 152.
PEREK, F. (2023): Construction Grammar and Usage-Based Theory. In: Díaz-Campos, M. – Balasch, S. (eds.): *The Handbook of Usage-Based Linguistics*. Chichester: John Wiley & Sons, 215 – 231. https://doi.org/10.1002/9781119839859.ch12
PEREK, F. – PATTEN, A. (2019): Towards an English Constructicon using patterns and frames. In: *International Journal of Corpus Linguistics*, 24/3, 354 – 384.
PETRUCK, M. R. L. (1996): Frame Semantics. In: Verschueren, J. – Östman, J.-O. – Blommaert, J. – Bulcaen, C. (eds.): *Handbook of Pragmatics*. Amsterdam: John Benjamins, 1 – 13.
POLLARD, C. – SAG, I. (1994): *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
SABOL, F. (1982): Slovo *čím* v platnosti zámena, častice a spojky. In: *Slovenská reč*, 47/1, 51 – 54.
SAG, I. A. (2010): English Filler-Gap Constructions. In: *Language*, 86/3, 486 – 545.
SAG, I. A. (2012): Sign-based Construction Grammar: An informal synopsis. In: Boas, H. C. – Sag, I. A. (eds.): *Sign-based Construction Grammar*. Stanford: CSLI Publications, 69 – 202.
SAG, I. A. – BOAS, H. C. – KAY, P. (2012): Introducing Sign-based Construction Grammar. In: Boas, H. C. – Sag, I. A. (eds.): *Sign-based Construction Grammar*. Stanford: CSLI Publications, 1 – 29.
SCHMID, H.-J. (2000): *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Mouton de Gruyter.
SCHMID, H.-J. (2017): Introduction. In: Schmid, H.-J. (ed.): *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin: De Gruyter, 1 – 3.
SCHMID, H.-J. (2020): *The Dynamics of the Linguistic System*. Oxford: Oxford UP.
SCHMID, H.-J. – KÜCHENHOFF, H. (2013): Collostructional analysis and other ways of measuring lexicogrammatical attraction: theoretical premises, practical problems and cognitive underpinnings. In: *Cognitive Linguistics*, 24/3, 531 – 577.
STEELS, L. (2013): Fluid Construction Grammar. In: Hoffmann, Th. – Trousdale, G. (eds.): *The Oxford Handbook of Construction Grammar*. Oxford: Oxford UP, 153 – 167.
STEFANOWITSCH, A. – FLACH, S. (2017): The corpus-based perspective on entrenchment. In: Schmid, H.-J. (ed.): *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin: De Gruyter, 101 – 127.
STEFANOWITSCH, A. – GRIES, S. Th. (2005): Covarying Collexemes. In: *Corpus Linguistics and Linguistic Theory*, 1/1, 1 – 43.
SZMRECSÁNYI, B. – GRAFMILLER, J. – HELLER, B. – RÖTHLISBERGER, M. (2016): Around the world in three alternations: Modelling syntactic variation in varieties of English. In: *English World-Wide*, 37, 109 – 137.
TRAUGOTT, E. – TROUSDALE, G. (2013): *Constructionalization and Constructional Change*. Oxford: Oxford UP.
UHRIG, P. (2015): Why the principle of no synonymy is overrated. In: *Zeitschrift für Anglistik und Amerikanistik*, 63/4, 323 – 337.
UNGERER, T. – HARTMANN, S. (2023): *Constructionist Approaches*. Oxford: Oxford UP.
WILLICH, A. (2023): *Konstruktionssemantik*. Berlin/Boston: De Gruyter.
ZESCHEL, A. (2008): Funktionsverbgefüge als Idiomverbände. In: Stefanowitsch, A. – Fischer, K. (eds.): *Konstruktionsgrammatik II. Von der Konstruktion zur Grammatik*. Tübingen: Stauffenburg, 263 – 278.
ZIEM, A. – BOAS, H. C. – RUPPENHOFER, J. (2014): Grammatische Konstruktionen und semantische Frames für die Textanalyse. In: Hagemann, J. – Staffeldt, S. (eds.): *Syntaxtheorien. Analysen im Vergleich*. Tübingen: Stauffenburg, 297 – 333.
ZIEM, A. – FLICK, J. (2019): Constructicography at work: implementation and application of the German Constructicon. In: *Yearbook of the German Cognitive Linguistics Association*, 7/1, 201 – 214.

Internet sources
[1] https://app.sketchengine.eu/ (accessed 12/6/2024).
[2] https://osf.io/j9p5y/ (accessed 12/6/2024).
[3] https://spraakbanken.gu.se/en/resources/konstruktikon (accessed 12/6/2024).
[4] https://constructicon.github.io/russian/about/ (accessed 12/6/2024).
[5] https://korpus.sk/o-korpusoch/korpusy/uvod/ (accessed 12/6/2024).

Konštrukčnogramatické výskumy slovenčiny, na Slovensku: prípadová štúdia a cestovná mapa

Resumé

Štúdia sa venuje možnostiam priekopníckeho uplatnenia konštrukčnej gramatiky (CxG) na slovenčinu, jazyk, ktorý je vo výskume konštrukcionizmu vo veľkej miere nedostatočne zastúpený. Po komplexnom úvode do CxG a jej teoretických základov predstavujeme prípadovú štúdiu slovenskej komparatívnej korelatívnej (CC) konštrukcie (napr. *Čím menej rečí tu bude, tým skôr zaspím*). V kontexte rozsiahlej korpusovej štúdie založenej na viac ako 3500 tokenoch zo slovenského korpusu Web 2011 demonštrujeme, ako možno pri výskume aplikovať konštrukcionistickú metodológiu, pričom využívame analýzu kovariantných kolexém na odhalenie štatisticky významných vzorcov jestvujúcich v rámci CC. Naše zistenia odhaľujú zložitú sieť prepojených mezokonštrukcií, ktoré vykazujú vysokú produktivitu a silnú tendenciu k formálnej symetrii. Prípadová štúdia tak poskytuje robustnú empirickú podporu pre používateľsky založený, sieťovo orientovaný pohľad na slovenskú gramatiku, ktorý je v súlade s predchádzajúcimi výskumami anglických a španielskych CC konštrukcií. Naša štúdia nielen obohacuje chápanie slovenskej CC konštrukcie, ale je významným argumentačne podloženým príspevkom ilustrujúcim hodnotu a realizovateľnosť CxG prístupov v „menších“ slovanských jazykoch. V závere štúdie bližšie opisujeme možnosti budúceho výskumu slovenského jazyka založeného na CxG prístupoch. Jeho súčasťou by mohli byť práce komplexného charakteru zamerané na budovanie tzv. construct-i-conu, v rámci ktorého sa „zmapujú“ konštrukčné siete slovenského jazyka, pričom zároveň môže dôjsť k integrácii komplementárnych metodológií uplatnených pri ich realizácii. Výsledky tejto štúdie sú teoreticky aj metodologicky významným príspevkom k rozvíjaniu konštrukčnogramatickej lingvistiky a môžu poslúžiť ako vzor pre ďalšie analogicky založené štúdie v slovenčine a v iných málo preskúmaných jazykoch.

Jakob Horsch
Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences
Panská 230/26, 811 01 Bratislava, Slovakia
E-mail: jakob.horsch@korpus.sk
ORCID: 0000-0001-7399-9079

Hans C. Boas
University of Texas at Austin, Department of Germanic Studies
1 University Station, C3300 Austin, TX 78712-0304, U.S.A
E-mail: hcb@mail.utexas.edu