Turning a list into a network via family-based expansion of the Russian Constructicon

Anna Endresen, Valentina Zhukova, Laura A. Janda, Daria Mordashova, Ekaterina Rakhilina

**Аbstract**

We expanded the database of the Russian Constructicon from 1,087 to over 2,200 constructions in seven months and worked out a methodology that can be used by other linguists working on comparable resources. We explain how we collected an inventory of constructions from various sources and significantly expanded it in a systematic way by modelling the relationships of constructions in terms of families. Each family was analyzed from the perspective of what members might be missing in the resource and could be added to optimize that family’s representation in the overall inventory. We provide five practical strategies for family-based expansion, each illustrated with an empirical case study of the families of Comparative, Addressee-encoding, Evaluative, Mirative, and Minimizing constructions.

**Keywords:** constructicon, expansion, family, construction, Russian

**1. Introduction**

We follow the conventions of Construction Grammar (Goldberg, 2006; Langacker, 2008; Diessel, 2015) that define “construction” as a learned form-meaning pairing at any level of linguistic complexity. Within this definition we focus on a more specific set of Russian grammatical constructions that are entrenched through repeated use (Langacker, 1987 and 2008; Schmid 2020) and function as multi-word expressions comprised of two elements: an “anchor” and one or more “slots”. The anchor is a fixed part of an expression that may consist of one or more words and/or grammatical patterns. A slot is a part of a construction that can be filled by various lexemes. Example (1) illustrates a grammatical construction of this type in Russian.

(1) ID 831: **NP/VP v ugod-u NP-Dat**

NP/VP in favor-acc.sg NP-dat

*On skaza-l èt-o v ugod-u načal*’*stv-u.*

he.nom say-pst.m this-acc.sg in favor-acc.sg boss.coll-dat.sg

‘He said that in order to please the bosses.’

The anchor in (1) contains two lexemes – the preposition *v* ‘in’ and the word *ugodu* ‘favor’ – and a grammatical pattern for the slots, namely that there is an NP or VP to represent some object or event, and an NP in the Dative case to represent the beneficiary of the object or event. In example (1) the first slot is filled with a VP, namely *skazal èto* ‘(he) said that’, and the second slot is filled with *načal*’*stvu* ‘the bosses’. Example (1) shows the identification number (ID) of this construction in the Russian Constructicon (described below). The first pair of lines shows the name of the construction (a morphosyntactic formula), and below that is the illustration (a short example).[[1]](#footnote-1)

One of the tenets of Construction Grammar is the idea that constructions in a language form not an unordered list but rather a structured inventory, termed a constructicon, or “repertoire” (Fillmore, 1988, p. 37), a global “network” of constructions (Goldberg, 2013, Goldberg & Herbst, 2021, p. 286), or a “set of networks” (Langacker, 2008, p. 237). However, the structural relationships that hold among constructions remain largely unexplored. While many thorough studies focus on alternations within small groups of constructions or a taxonomy of possible horizontal and vertical links among individual constructions, only a large-scale constructicon project can bring this discussion to a qualitatively different level by testing the general idea of a structured inventory against sufficient data. Moreover, we suggest that a large-scale constructicon resource is beneficial not only for testing theoretical claims but also for systematization of the collected constructions and harmonization of the ways in which individual constructions are represented.

Exploring the relationships among constructions and growing the inventory of constructions has been one of the major directions of our work in building a constructicon resource for Russian (<https://constructicon.github.io/russian/>). The Russian Constructicon is relatively “young” compared to other constructicons: the project has been administered over the past six years (2016-2022), and the user interface has been launched very recently, in 2021. Тhat said, the Russian Constructicon is the largest[[2]](#footnote-2) among the existing comparable resources, as it contains over 2,200 partially schematic multi-word constructions, most of which are fully described (and work is ongoing).

In the course of our work on the Russian Constructicon, we have identified various levels of connectedness among constructions, namely families, clusters, and networks. Only the smallest groupings at the family level are relevant for this article. We define a family as a relatively small and homogeneous group of constructions (usually 2 to 9) that exhibit family resemblance and share semantic and often also syntactic or other structural properties (including reduplication, inversion, double negation, etc.) (see Janda et al. 2020; **Endresen** &Janda, 2020). Family resemblance means that the constructions in a family share not necessarily all properties but various subsets of these properties (cf. Wittgenstein, 1953; see also Goldberg & Jackendoff, 2004 and Goldberg, 2006 for discussion of family resemblance both in syntax and semantics; Ruiz de Mendoza Ibáñez et al. 2017; Diessel, 2019: Chapter 10).

When identifying a family of constructions, we prioritize their semantic similarity first and take into account their other linguistic properties (including structural similarity, shared syntactic properties, similar or identical anchor words) secondarily. This approach comports well with the Cognitive Linguistics framework where meaning is understood as the major driving force of linguistic phenomena (Janda, 2015). Prioritizing semantics over syntax is motivated by our observations of constructions with very different syntactic properties that have very similar semantics, justifying their analysis as members of a single family. For example, (1) belongs to the Beneficiary family that encodes an Addressee who benefits from an action. Beneficiary is related to nearby families such as Maleficiary (encoding a participant who is negatively affected by an event), and Audience (encoding a group of people who receive a message) in a cluster of constructions that encode Addressee (a hyperrole that includes not only the Addressee of a speech event, but also the Addressee of any non-verbal action).[[3]](#footnote-3) Further examples are the family of Prohibitive constructions discussed in Section 4 and the family of Mirative constructions presented in Section 5.4.

We used a variety of methods to expand the inventory of constructions and witnessed several stages that differ in terms of efficiency and duration. As a result, we elaborated a methodology that can be useful for other linguists who intend to build comparable constructicon resources. We term this methodology *family-based expansion* because its key idea is to identify constructional families and add those family members that are missing in the resource. The proposed methodology served two purposes: it enabled us to 1) significantly expand our inventory of constructions in a relatively short time (more than doubling the number of constructions over a period of seven months), and 2) turn a list of collected items into a structured network, a system of interrelated constructions. Thus, family-basedexpansion and systematization (or structuralization) of the inventory go hand in hand. The purpose of this article is to detail this methodology and illustrate it with specific examples of constructional families that became better represented due to our commitment to family-based expansion.

The remainder of this article is structured as follows. In section 2, we provide general information on the Russian Constructicon and outline the three stages of expanding the inventory of constructions in the development of this project. Section 3 details the procedure of identifying families of constructions. Section 4 explains why horizontal relationships between constructions are more important for our purposes than vertical ones. Section 5 breaks the proposed methodology of family-based expansion down into five specific strategies, each illustrated with an empirical case study of a family that underwent significant expansion. Our Conclusions summarize the methodological, empirical, and theoretical implications of the proposed methodology for the field of constructicography and building constructicon resources for additional languages.

**2. Ways of expanding the inventory of constructions at different stages of the project**

In this section we present the Russian Constructicon and outline the major stages of our work on the expansion of its inventory. We will show that each stage was necessary and served as the basis for the subsequent stage and describe the methodology that proved most efficient.

**2.1. The resource**

The Russian Constructicon is a free open access electronic resource that offers a searchable database of Russian constructions accompanied with descriptions of their properties and illustrated with corpus examples (<https://constructicon.github.io/russian/>). The collected constructions also carry a set of meta-linguistic tags specifying their semantic, syntactic, and stylistic properties. This resource is designed for users of various kinds, including not only professional linguists working on Russian, but also students and teachers of Russian as a foreign language as well as non-Slavists, typologists, and specialists in natural language processing (Janda et al. 2018 and 2020).

We prioritize collecting multi-word and partially schematic grammatical constructions (Ehrlemark et al., 2016) that are underrepresented in dictionaries, reference grammars, and textbooks for language learners. One of the missions of the Russian Constructicon is to be useful for learners of Russian and therefore we focus on those highly frequent, entrenched but non-transparent constructions that are strategic for second language acquisition.

The representation of each construction contains the following fields, here illustrated by (1) above:[[4]](#footnote-4)

* unique ID number, here 831
* name: a schematic version of the construction with its anchor words, and with slots represented by linguistic annotations (most of these directly derived from the Leipzig glossing rules; Comrie et al. 2008), here ***NP/VP v ugodu NP-Dat****,* see glossing in (1)
* illustration: a brief example, here *On skazal èto v ugodu načal*’*stvu.* ‘He said that in order to please the bosses’
* definition: this explains the meaning of the construction and is annotated for semantic roles given in square brackets, here “The construction has a concrete and an abstract meaning. In the concrete meaning the construction indicates that a participant engages in an [action]Action that is motivated by a desire to do something pleasant for the other [participant]Beneficiary, while at the same time neglecting their own needs and best interests. In the abstract meaning the construction indicates that a participant undertakes an [action]Action in harmony with the [ideals, rules, policies]Theme of a certain group of people.
* 5 corpus-based examples from the Russian National Corpus ([https://ruscorpora.ru](https://ruscorpora.ru/)) that are annotated for the same semantic roles as the definition
* CEFR level: the Common European Framework of Reference for Languages scale of proficiency levels, here B2
* common fillers: a list of lexemes commonly found in the slots
* semantic types: here Addressee with subtype Beneficiary, also Cause and Purpose
* syntactic type of construction: here Head and Modifier Construction
* anchor: syntactic function (here: Modifier) and syntactic structure (here: Nominal Pattern) and part of speech (here: Preposition and Noun)
* dependency structure for both the name and illustration of the construction (according to Universal Dependencies standards)
* additional information such as communicative type, usage label, and references

A separate page of the Russian Constructicon interface is designed for advanced search of constructions according to various linguistic parameters (including Morphological categories, Syntactic type of construction, Syntactic function and structure of the anchor, Part of speech of the anchor, Semantic type, Semantic role, and CEFR level), and it is possible to combine several filters in a single query. It is not a trivial task to make a constructicon a multifunctional resource that would meet the interests of a wide variety of target users. In particular, we must constantly seek the right balance between purely linguistic and practical considerations. Yet, this challenging endeavour is worth undertaking, as it keeps the resource dynamic and alive, although not without various struggles.

**2.2. Stages and means of expansion**

The overall process of collecting constructions for the Russian Constructicon went through three major stages: 1) initial stage, 2) corpus-based expansion, and 3) system-based expansion (Janda et al. 2020; Endresen & Janda, 2020, pp. 6-8). Below we outline the key ideas crucial for each stage and then focus on the system-based expansion, and more specifically, *family-based expansion* and share our methodology of expanding the constructicon via identification and analysis of constructional families.

Figure 1 presents a line chart that tracks the progress in the inventory expansion over time. The horizontal axis is a timeline that starts in January 2016, when we started collecting constructions for the constructicon, and ends in May 2020, when we completed our work on family-based expansion of the inventory and temporarily stopped adding more entries to the database. Since then, our focus shifted from inventory expansion to harmonization of semantic and syntactic annotation as well as detailed description and illustration of collected constructions. Each year is indicated on the horizontal axis by the last two numbers, for example 2016 is 16, 2017 is 17, etc. The vertical axis of the chart indicates the number of constructions present in the resource at each stage. The outcome of each stage is the total number of collected constructions shown in a square box.

A picture containing histogram

Description automatically generated

Figure 1: Stages of inventory expansion of the Russian Constructicon.

At the initial stage we created 660 entries, then reached a milestone of 1087 entries (by adding 427 entries) via corpus-based expansion, with a steep incline to the total of 2267 entries by the end of system-based expansion. The graph shows that the three stages of inventory expansion were uneven in terms of time. Most importantly, the last stage was the shortest (it lasted only seven months) but the steepest, more than doubling the inventory.

We initially collected Russian constructions from scholarly works and textbooks for learners of Russian, authored primarily by the members of the project team and gradually examined more possible sources. The first 660 constructions were collected from a textbook on Russian case based on data collected from newspaper articles (Janda & Clancy, 2002), an anthology dedicated to discussing various groups of constructions from the Construction Grammar perspective (Rakhilina, 2010), and the pioneer monograph on syntactic patterns of spoken Russian written in the 1950s (Švedova 1960). In addition, we used a shared digital spreadsheet for crowdsourcing, where all team members could add the constructions encountered in their teaching and daily life.

Stage 2 employed various techniques applied to corpus data, and thus we term it *corpus-based expansion*. We culled a list of 1000 most frequent collocations (bigrams and trigrams) from the Russian National Corpus ([https://ruscorpora.ru](https://ruscorpora.ru/), over 330 billion words) and analyzed them manually in terms of whether these collocations represent constructions (multi-word expressions comprised of an anchor and one or more slots). A lot of entries were also added through manual examination of a set of running texts of certain genres that feature colloquial discourse (primarily dramaturgy and children’s literature).

Stage 3 was both qualitatively and quantitatively different from the previous two. Its driving force was system-based expansion (more precisely, *family-based expansion*) built on the multi-level semantic and syntactic annotation of constructions that we elaborated during the first two stages. We started classifying constructions early in the project (at the end of Stage 1), gradually working out a system of multi-level annotation of their syntax and semantics. As we added more data during Stage 2, the system of annotation underwent multiple revisions. Often, we had to go back to already analyzed and annotated constructions and reconsider our annotation in the light of newly added entries. Thus, adding new constructions to the inventory forced us to challenge, verify, and often update already established groups. When the inventory reached the size of 1087 constructions, a relatively stable system had emerged and we started working on family-based expansion of the database. By this time, the semantic classification of constructions was complete and no longer required major revision. Newly added constructions fit nicely into already analyzed semantic types (this term is explained in Section 3), and the overall system of semantic annotation since then undewent only minor changes. We found that the size of about one thousand constructions was a critical mass necessary for the semantic classification in our resource to emerge and become robust.

During stages 2 and 3 we elaborated a multi-level semantic and syntactic classification of constructions that subdivided our inventory into meaningful classes and smaller groups and thus facilitated the identification of constructional families. Once we completed the overall semantic and syntactic classification of constructions, we started work on identification of constructional families, which tend to be smaller groupings than those semantic types that we established at Stages 1 and 2. Identification of constructional families, which is the key factor in family-based expansion, resulted from using both semantic and syntactic annotation of constructions, but benefited primarily from the semantic one. We explain the principles and the outcomes of our semantic annotation in Section 3.

Stage 3 of the inventory development is termed *family-based expansion* because it focused on the analysis of constructional families from the perspective of what possible members of each family were still missing in the database. At this stage, the new constructions were added directly to the identified families (previously established meaningful groupings) of constructions that feature family resemblance in terms of their semantic and often also syntactic properties. Often this process involved reanalysis of the identified families. We detail and examplify *family-based expansion* by breaking it down into five practical strategies presented in Section 5.

Our data shows that family-based expansion served as a very effective methodology. Stage 3 was exceptionally successful for the progress of the project and was built on the outcomes of the previous two stages, both quantitatively, in terms of the collected inventory, and qualitatively, in terms of the elaborated system of semantic and syntactic annotation. Our understanding of the data and the notion of a family of constructions evolved and crystallized gradually as we progressed in systematization of the collected data. Therefore, in our view, the expansion and the systematization (or structuralization) of the inventory go hand in hand, they are two sides of the same coin. This also means that expansion of the inventory of a constructicon resource is a positive and useful endeavour because it challenges and verifies the model (or theory) employed in the resource for structuring the data. We argue that a large-scale expansion in the long run reinforces the process of turning a long list of constructions into an organized network that represents the real complexity of a constructicon.

**3. From semantic types to families of constructions**

In this section we explain how our semantic annotation of constructions, and particularly general semantic types of constructions, facilitated identification of coherent families of constructions. Linguistic meaning is conveyed in a variety of ways, via lexemes, grammatical morphemes, and grammatical constructions. Linguistic typology has focused more on the semantics of lexemes and grammatical morphemes, and our semantic classification is informed by this tradition, including, for example, Jespersen’s (1924) list of “conceptual categories”, Nida’s (1949: 166-169) lists of the most common meanings expressed by morphemes, Jakobson’s (1957) classification of grammatical meanings, and semantic types proposed by contemporary scholars such as Jackendoff (1983, 2002), Talmy (1985, 2000), and Mel’čuk (1998). All semantic types are defined in detail on the Russian Constructicon website, complete with scholarly citations.

**3.1. A bottom-up approach informed by typologically attested semantic types**

In elaborating the semantic annotation of constructions, we adopted a bottom-up approach. Our objective was (and remains) to analyze Russian constructions on their own terms, allowing patterns to emerge from the data rather than imposing other models. The annotation of constructions was carried out by a panel of three native speakers (and sometimes more team members were involved) who worked together as a team over a long period of time. They worked towards reaching a consensus about annotation of every single construction.

When we annotated constructions, we considered corresponding categories, distinctions, and terminology already known from typological studies of grammatical meanings and thus verified our results against typological studies of grammatical categories. We adopted and adjusted this terminology for annotation of constructional meanings. In doing so, we tried to understand which types of meanings that are encoded grammatically in other languages can be expressed in Russian by means of syntactic constructions. For example, the construction in (2) illustrates a semantic type that we first termed *Unexpected outcome* and later renamed as *Mirative* after the category that is expressed grammatically in some languages (Aikhenvald, 2012). Mirative semantics entails that the speaker is surprised by new, unexpected information. Even though this semantics is expressed in (2) not by a single grammatical marker, but by a whole construction, as a complex interplay of elements from different “levels” (grammatical markers, specific lexemes, discourse units, word order, prosody), we adopt the term *Mirative*.

(2) ID 61: **NP-Nom voz**’**m-i i VP-Pfv.Imp.Sg!**

NP-nom take.pfv-imp.sg and VP-pfv.imp.sg

*Ivan voz’m-i i kup-i nov-uju mašin-u!*

Ivan.nom.sg take.pfv-imp.sg and buy.pfv-imp.sg new-acc.sg.f car-acc.sg

‘Ivan suddenly bought a new car!’

Our practice of consulting typological studies to inform our semantic annotation is motivated by the idea of the *universal grammatical inventory* defined as “a coherent semantic space such that each language extracts a certain part of it and distributes it between the available grammatical means” (Plungian, 2011, p. 65; 1999, pp. 311-312). Further details about the typological implications of this work go beyond the scope of the present study; see Janda et al. forthc. Many of our semantic types comport with typologically attested grammatical categories. At the same time, we found that our system goes beyond typologically attested and typically grammatical meanings and additionally covers other semantic types that are suggested by our data, for instance such semantic types as Non-existence, Attitude, Assessment, Salient property, and Reaction to the previous discourse.[[5]](#footnote-5)

**3.2 Final outcome: A multi-level semantic classification of constructions**

Our work on semantic annotation of constructions resulted in the overall classification of over 2,200 constructions that subdivides our inventory into meaningful classes and smaller groups and facilitates the identification of constructional families. Table 1 provides an overview of this classification built as a multi-level hierarchy of classes, subclasses, semantic types, and subtypes. The table lists 55 semantic types (marked with bullet points) and shows how they are grouped into superordinate numbered classes (Qualia, Modality, Subjectivity, Discourse, and Parameters) and their subclasses. Most of the semantic types have an additional level of granularity represented by their 182 subtypes (here indicated with a «+» symbol). For example, the type Addressee has subtypes, whereas Instrument has no subtypes.

Graphical user interface, table

Description automatically generated

Table 1: Overview of semantic classification of constructions in the Russian Constructicon.

A detailed account of this classification lies beyond the scope of this article. It can be found on the website of the Russian Constructicon[[6]](#footnote-6) and is the topic of a separate article (Janda et al., forthc.). What is important for our purposes is to explain how we get from these (often rather general) semantic types to specific families of constructions. This is the purpose of the next subsection.

**3.3 Identification of families**

The exact method of how we identified families of constructions (with a certain degree of simplification) can be described as follows. We organized our data in a google doc spreadsheet (comparable with an Excel spreadheet), where all constructions were listed in column A, their semantic types in column B, and their syntactic functions were in column C. We then examined the intersections of semantic and syntactic types by filtering the data first according to the semantic types and then according to their syntactic types (the opposite order turned out to be less successful). We found many examples of such combinations of semantic and syntactic types: e.g., Condition & Biclausal, Assessment & Praedicative, Apprehension & Clause, Distribution & Preposition, Salient property & Nominal pattern, among others.

In many cases this method gave us rather homogeneous small groups of constructions like the family of Minimizing constructions presented in Section 5.5. However, more often we arrived at large groups of constructions with rather diverse semantics that had to be narrowed down in order to achieve smaller groupings. Looking back, we can now see that this outcome was often due to the imprecision or inconsistency of our preliminary syntactic annotation. The complexity, diversity, and multidimensionality of the constructions that we had to annotate presented real challenges, partly overcome by recognizing that many constructions are multiply motivated and thus deserve to be classified as more than one general semantic type. In the case of large groupings, more specific semantics was worked out manually and often over group discussions. We examined one semantic type after another and analyzed smaller groupings with the focus on both semantics and syntax of constructions. In this process we crystallized our understanding of what kinds of constructional grouping can be called a family. We found that in identification of families of constructions the semantics plays a stronger role than syntax, since it is not uncommon that members of the same family exhibit very different syntax (more about this in Section 4.4).

The reader might wonder how exactly the identified families of constructions correspond to the semantic types presented in Table 1. On the one hand, the semantic classification is a hierarchy of superordinate classes, subclasses, semantic types, and their subtypes that are related to one another conceptually and also overlap on the level of individual constructions (when a construction belongs to two or more semantic types at the same time, as is the case for about 40% of constructions in our database, e.g., the Mirative constructions in Section 5.4). On the other hand, we group constructions in families, clusters, and networks. Several families form clusters, and several clusters comprise a network (we present this model in detail for Russian prohibitive constructions in Janda et al., 2020 and for Assessment and Attitude constructions in Endresen & Janda, 2020).

Depending on how many constructions belong to a certain semantic type and how homogeneous or diverse they are in terms of their semantic and syntactic properties, a semantic type can correspond to a whole network, a cluster, or a family of constructions. There are some semantic types that are very heavily populated in our inventory, and they normally constitute an entire network. For example, the constructions that belong to the semantic type Assessment, defined as expressing evaluation of an item external to the speaker,[[7]](#footnote-7) constitute a large group of 225 entries, and they comprise a network that consists of four distinct clusters that are formed by a total of 25 families (described in detail in Endresen & Janda, 2020). The constructions that belong to the semantic type Prohibitive, defined as expressing strict prohibition to perform an action in the future,[[8]](#footnote-8) yield 57 entries that constitute a network comprised of two clusters that are formed by 12 families (cf. Janda et al., 2020). On the other end of the scale are some semantic types that are very sparsely populated, partly because their semantics is very specific and narrow. For instance, the semantic type Inclusive, indicating that an element is included in a set,[[9]](#footnote-9) contains only eight constructions that form a single family.

**4. Family status: vertical and horizontal relationships**

In this section we first establish two types of relationships (vertical and horizontal) among constructions and discuss how they define two crucial types of constructional families (Section 4.1). Then we address the question of whether vertical or horizontal relationships between constructions are more decisive for family status (Sections 4.2 and 4.3) and explain how this motivates our approach to representation of individual constructions in terms of granularity (Section 4.4).

**4.1 Nuclear and extended families**

As evidenced by presentations at the International Conference on Construction Grammar 2021, many recent studies that adopt the constructionist approach are concerned with the two major types of relationships that link constructions to each other: the vertical (hierarchical) relationships and the horizontal (also termed lateral, paradigmatic or alternation/allostructional, cf. Goldberg 2002, Cappelle 2006) relationships between constructions. Scholars often use a family metaphor for both types of relationships.

One type of approach focuses on vertical relationships. A family of constructions is understood as an inheritance structure, where one or more specific constructions instantiate a more abstract construction (or schema) and inherit its properties, fully or partially (see Lyngfelt, 2018, p. 7 for discussion). Here, the family metaphor is manifested by the terms parent and child construction that belong to different levels of abstraction: the parent construction is more abstract, and the child construction is more specific. Since the focus is on parent(s) vs. child(ren), we term this structure a *nuclear* *family*. Depending on whether there is a single parent or more and therefore one or multiple inheritance links (for example, with two parents), a nuclear family can be visualized as shown in Figure 2a and b respectively.

Diagram

Description automatically generated

Figure 2: Types of relationships between constructions.

In the Russian constructicon, we sometimes have both parent and child constructions as distinct entries, for instance when one specific lexical instantiation of a more general schema becomes very entrenched and highly frequent, we represent it as a separate entry, and normally such parent - child counterparts would be found within a single family (as an example consider the family of Minimizing constructions described in Section 5.5).

While vertical relationships reflect certain structural similarity, horizontal relationships are arguably more concerned with semantic kinship. Horizontal relationships between constructions are visualized in Figure 2c. They can exist not only between constructions that have a shared parent schema and similar syntactic structure, but also between constructions that have a more distant syntactic resemblance (see Section 5.1.) and even between constructions that are syntactically unrelated but semantically similar.

When we identify and analyze constructional families in the Russian Constructicon, we focus primarily on the horizontal relationships (while taking into account the vertical relationships as well). Semantic similarity at approximately the same level of abstraction defines horizontal relationships between constructions. Relationships can entail syntactic or structural similarity too, though not necessarily motivated by a shared parent schema. Because a family can extend beyond a nuclear family to non-immediate relatives, we term this structure an *extended family*.

What dimension of relationships between constructions, vertical or horizontal, has more weight in suggesting a family interpretation? In what follows we argue that strong horizontal relationships, even when the connection with the shared parent is problematic, suggest that the constructions in question belong to a single family (Section 4.2). We also show that in the reverse situation, that is when the connection with a shared parent is very strong, but the horizontal relationships are problematic, the constructions can hardly be regarded as belonging to a single family (Section 4.3). Our data suggest that semantic similarity of constructional counterparts (manifested horizontally) plays a more crucial role than syntactic similarity (often motivated vertically by having a shared parent).

**4.2 Strong horizontal relationships and semantic similarity guarantee family status**

Consider a group of three constructions that encode very similar semantics but are syntactically quite distinct. The three constructions in (3)-(5) carry very specific Prohibitive semantics: they encode not the standard prohibition spoken in a strict peremptory tone, but a milder friendly tone and thus can be interpreted as attenuated prohibition. Each of these constructions expresses that the speaker views the activity of the interlocutor as excessive, or too long, and suggests stopping this activity and moving on to something else. This semantics is motivated by having quantifying predicates that “measure” the activity expressed by the VP and denote ‘enough’.

Morphologically, we see very different predicates: *xoroš* in ID 1247 is a short form of an adjective meaning ‘good’, *dovol’no* in ID 114 is an adverb, and *xvatit* in ID 344 is a verb. However, synchronically, they are frozen forms that suggest their morphological status only in terms of the suffixes they contain.

(3) ID 1247: **xoroš VP-Ipfv.Inf!**

good.sg.m.short VP-ipfv.inf

‘Quit X-ing! Cut it out!’

*Èj, naverxu! Xoroš pryg-a-t’!*

hey upstairs good.sg.m.short jump-ipfv-inf

‘Hey, up there! Stop jumping!’

(4) ID 114: **dovol’n-o (PronPers-2.Dat) VP-Ipfv.Inf!**

enough-adv you-dat VP-ipfv.inf

‘Enough X-ing (for you)!’

*Dovol’n-o žalov-a-t’-sja!*

enough-adv complain-ipfv-inf-med

‘Enough complaining!’

(5) ID 344: **Xvat-it (PronPers-2.Dat) VP-Ipfv.Inf!**

be\_enough.pfv-fut.3sg you-dat VP-ipfv.inf

‘Enough X-ing (for you)! That will do!’

*Xvat-it rug-a-t’-sja!*

be\_enough.pfv-fut.3sg fight-ipfv-inf-med

‘Enough fighting! It will do!’

As shown in Figure 3, the three constructions can be analyzed as instantiations of a single more abstract parent schema **[quantifier + VP-Ipfv.Inf]**, where the quantifier has a predicative function, and the VP specifies the activity that should be stopped. In the constructions ID 114 and ID 344, the interlocutor (here: Experiencer) can be optionally overtly expressed, and therefore they can be connected to a subschema that includes this optional open slot.

Diagram

Description automatically generated

Figure 3: A family of prohibitive constructions with predicates quantifying an activity.

The predicates in these constructions are grammaticalized to different degrees, blurring the structural similarities. It is problematic to try to decompose these “frozen” structures and associate them with any more abstract schematic patterns. We suggest that in cases like this, it is more fruitful to focus on horizontal relationships between constructions. These constructions are close synonyms, and we can differentiate between them semantically by looking at the restrictions on the lexical fillers of their open slots. As close synonyms, these constructions are motivated as semantic rivals and interpreted as belonging to a single family.

**4.3 Shared parent schema and syntactic similarity do not guarantee sisterhood**

Let us now consider the reverse situation: strong and well pronounced affiliation with the same parent schema in combination with weak and synchronically irrelevant horizontal links between constructions. Our solution is that such constructions clearly belong to different families.

As an example, consider three head + modifier constructions in (6)-(8) that contain the same anchor preposition *iz* ‘out of’ and are visualized in Figure 4.

(6) ID 919: **VP NP-Acc iz NP-Gen**

VP NP-acc out\_of NP-gen

*Oni vygna-l-i kot-a iz dom-a.*

they.nom banish-pst-pl cat-acc.sg out\_of house-gen.sg

‘They banished the cat from the house.’

(7) ID 133: **VP iz NP-Gen (k NP-Dat)**

VP out\_of NP-gen to NP-dat

*ulybnu-t’-sja iz vežlivost-i*

smile-inf-med out\_of politeness-gen.sg

‘smile out of courtesy’

(8) ID 1979: **VP iz NP-Gen**

VP out\_of NP-gen

*strelja-t’ iz ruž’j-a*

shoot-inf out\_of gun-gen.sg

‘shoot a rifle’

Figure 4 shows that the three constructions under scrutiny are directly related to the more abstract schematic pattern, the parent construction **[VP iz NP-Gen]**. The reading of the preposition *iz* in all three constructions is built on its prototypical image schema ‘out of a container’ (see Endresen, 2019), yet the container in these three cases is interpreted differently. In ID 919, the container is a concrete physical space. In ID 133, we deal with a figurative use: the container is an emotional state (in the illustration: the feeling of politeness), and this motivates the semantics of Cause of the prepositional phrase. In ID 1979, the container is a physical object that is long, hollow, and cylindrical (a gun, spray hose, bottle, shower pipe, etc.), which motivates the instrumental semantics of this construction (the trajector moves out of this container, e.g., a bullet inside a gun or water inside a shower pipe).

Diagram

Description automatically generated

Figure 4: Relationship of related constructions not belonging to a single family due to different semantics.

In Figure 4, the dotted horizontal lines between the boxes indicate that these three constructions, despite their close relations both syntactically and on a deeper semantic level, cannot be analyzed as belonging to a single family because their semantic properties are different. Although they have the same parent construction, a very similar syntactic structure, and the same anchor word, these constructions nevertheless do not belong to a single family. On the contrary, we argue that these three constructions, due to their synchronically different semantics, belong to three different families and three different semantic types indicated in Figure 4: ID 919 belongs to the type Spatial expressions, subtype Path: Source; ID 133 belongs to the semantic type Cause; and ID 1979 belongs to the semantic type Instrument. To conclude, weak horizontal relationships in our view do not suggest membership in the same family.

**4.4 Level of granularity in construction representation**

Building an inventory of constructions inevitably entails making decisions. One such decision concerns determining the level of granularity in representation of constructions that is most optimal for the purposes of the resource. The Russian constructicon is designed to serve the needs not only of linguists but also of second language learners of Russian. This requires finding the right balance between adequate linguistic representation and usefulness for practical purposes.

In this regard, highly abstract representations are not practical for at least two reasons. First, highly abstract representations are less informative and present more difficulties in terms of description that has to take into account too many details of use in a single entry. While abstract schematic patterns like head + modifier, coordination, subordination, and the like are already described in grammars, we observe that it is at the level of specific lexical anchors with non-transparent or otherwise non-trivial argument structure that the idiosyncratic properties of constructions manifest themselves (recall the prohibitive constructions in Figure 3).

Therefore, the vast majority of constructions included in our resource contain a fixed lexical element (the anchor) and one or more open slots that usually represent the arguments or crucial components of a given construction.

Second, our major priority is to represent semantic distinctions. For example, in the case of the three constructions with the same preposition in Figure 4, their semantic distinctions motivate us to represent them as three distinct entries rather than a single construction **VP iz NP-Gen**.

Summing up, the degree of granularity that we choose for representation of an expression as a construction unit of the database is motivated by the semantic distinctions that we observe empirically and want to highlight, and by practical purposes, including language learning and pedagogical considerations.

**5. Strategies of family-based expansion illustrated by case studies**

In this section we demonstrate the strategies that we employed while working with families of constructions in the process of expanding the inventory of the Russian constructicon. We focus on both vertical and horizontal relationships among constructions that facilitate identification of missing family members. Sections 5.1 to 5.5 highlight five major strategies illustrated with example of families of constructions that underwent significant expansion. In Section 5.1, we show how one can add constructions with synonymous lexical anchors and similar syntax and exemplify this technique with a family of Comparative constructions. Section 5.2 is devoted to adding constructions with antonymous semantics and similar syntactic properties, as illustrated with the constructions that encode the Addressee, the Beneficiary, and the Maleficiary participants. Section 5.3 is a discussion of how to expand a family by adding constructions with alternative argument structure patterns related to the same anchor, as detailed in a case study of a family of Attitude constructions expressing Support. Section 5.4 shows how we can “grow” an entire family out of a single construction, in the case of the family of Mirative constructions that signify Disappearance of a key participant in a situation. Section 5.5 makes the case for adding highly frequent and prominent child constructions as distinct entries, as we did with the family of Minimizing constructions.

**5.1. Adding constructions with synonymous anchors and similar syntactic properties**

The most productive strategy for expanding a family of constructions involves exploration of their semantic properties and revealing their synonymous (or antonymous) counterparts. We illustrate this strategy with a case study of a family of Comparative constructions that we expanded from two initial members by adding eight closely related peers.

Overall, the Russian Constructicon contains 164 constructions that express Comparison and represent five semantic subtypes termed Similarity, Equality, Inequality (according to major typological accounts of Comparative constructions, cf. Treis, 2018), as well as Contrast and Imitation. A cluster of constructions that encode Similarity comprises 34 entries that form 7 families. We now focus on one of these families formed by constructions where the indication of similarity between an object and a standard of comparison serves as a means of specifying the former. Therefore, the object of comparison is usually a more generic item, as opposed to the standard of comparison, which is a prototype or a well-known entity. In (9) and (10) we provide the two constructions already present in the database prior to family-based expansion.

(9) ID 171: **kto-to / čto-to vrode NP-Gen[[10]](#footnote-10)**

who-indef what-indef kind\_of NP-gen

‘someone/something like X’

*kto-to vrode tebja*

who.nom-indef like you.acc

‘someone like you’

(10) ID 757: **NP/VP v dux-e NP-Gen**

NP/VP in spirit-loc NP-gen

‘an X/to X in the spirit of Y’

*rasskaz*.nom.sg *v dux-e Čexov-a*

story in spirit-loc Chekhov-gen.sg

‘a short story à la Chekhov’

Both constructions denote an approximate similarity between an object or a participant of a situation (the object of comparison[[11]](#footnote-11)) and another object or participant (the standard of comparison) that normally constitutes the prototypical or a well-known exemplar. The object of comparison is usually named in a more generic way and is specified through comparison to a concrete, well-known or prototypical entity that serves in this case as the standard of comparison: someone like you in (9) and a type of short story that resembles the stories written by Chekhov in (10). Note that in (10) we also deal with metonymy, because the standard of comparison is not the writer himself but his writings and his style.

Syntactically, both constructions have a head + modifier structure, where the head is an NP or VP or a pronoun and the modifier is expressed by a prepositional phrase. In each of these constructions a lexical anchor encodes the idea of comparison, resemblance, similarity and at the same time a certain degree of approximation. Both anchors can be roughly translated by the English suffixoid -*like* (as in *Chekhov-like stories*).

We started our search for additional members of this family based on the analysis of the lexical anchors and their morphological structure, which in this case can be understood best in terms of grammaticalization. The anchor of construction ID 171 is a complex denominal preposition *vrode* ‘kind of’ that is fully grammaticalized. Today, *vrode* is a single word that can be traced back to the prepositional phrase *v* + *rod-e* ‘in kindred-loc.sg’ consisting of the preposition *v* ‘in’ that governs the noun *rod* ‘kindred, ancestry, type’ in the locative case.

Likewise, the anchor of construction ID 757 contains a similar prepositional phrase: *v* + *dux-e* ‘in spirit-loc.sg’, featuring the same preposition *v* ‘in’ and the noun *dux* ‘spirit, soul, mood’ in the locative. As opposed to *vrode*, the anchor *v* *duxe* is arguably not fully grammatizalized into a single lexeme and allows for two syntactic interpretations: it can synchronically be analyzed as a new complex preposition that governs the noun phrase NP-Gen (filled with the proper noun Chekhov in (10)) or, in a more conservative style, the anchor can be analyzed as a prepositional phrase with the noun *dux* (that in its turn governs the noun *Chekhov*). Neither of these analyses is unproblematic because the expression *v* *duxe* is in the intermediate stage of grammaticalization, on its way to turn into a single preposition.

Crucial for our purposes, neither of the two Comparative constructions is one-of-a-kind. Both the fully grammaticalized preposition *vrode* and the partially-grammaticalized PP *v duxe* have counterparts with comparable semantics and structure that arguably belong to the same family. We investigated these rival Comparative constructions and added them as distinct entries, thus expanding the family by eight additional members. Figure 5 visualizes the structure of this family using the following conventions: boxes represent distinct constructions, shading highlights those constructions that already were present in the database, white boxes are their added peers. Vertical lines indicate inheritance links and the parent vs. child relationship.

Diagram

Description automatically generated

Figure 5: Family of Comparative constructions encoding similarity through the reference to a well-known exemplar or prototype.

In Figure 5, we provide approximate literal English glosses for the lexical anchors. Note that the dotted border of the box representing the construction ID 171 means that this construction no longer has the name given in (9). Instead, the name of ID 171 was reformulated in a more general way, to be consistent with other constructions in Figure 6. The initial pronominal anchor of this construction *kto-to/čto-to*‘someone/somewhat’ is a specific instantiation of the more general pattern **NP/VP vrode NP-Gen**, as in (11) and is provided as a filler of the NP slot in one of the illustrative examples for this construction.

(11) ID 171: **NP/VP vrode NP-Gen**

NP/VP kind\_of NP-gen

*nebol*’*š-oe zdani-e vrode časovn-i*

modest-sized-nom.sg.n building-nom.sg kind\_of chapel-gen.sg

‘а modest-sized building somewhat like a chapel’

The fate of construction ID 171 thus shows how family-based expansion of the inventory often entails systematization and harmonization of the existing entries.

The constructions represent different degrees of grammaticalization (compare fully grammaticalized *napodobie* in ID 1202 vs. less grammaticalized *na maner* in ID 1203), but the exploration of exact differences between them in this regard is beyond the scope of this article. For the purposes of our argument, it suffices to say that the anchors spelled as a single word like *vrode* (ID 171) can be considered full-fledged derived denominal prepositions: *tipa* (ID 1174), *srodni* (ID 1183), *vrode* (ID 171), *napodobie* (ID 1202):

(12) ID 1174: **NP tipa NP-Gen**

NP like NP-gen

*dom tipa xižin-y*

house.nom.sg like cabin-gen.sg

‘a cabin-like house’

(13) ID 1183: **NP srodni NP-Dat**

NP akin NP-dat

*Kollektiv tele-kompani-i by-l srodni teatral*’*n-oj*

staff.nom.sg TV-company-gen.sg be-pst.m akin theater-dat.sg.f

*trupp-e.*

troupe-dat.sg

‘The staff of the TV company was akin to a theater troupe.’

(14) ID 1202: **NP/VP napodobie NP-Gen**

NP/VP like NP-gen

*pesn-ja napodobie gimn-a*

song-nom.sg like hymn-gen.sg

‘a hymn-like song’

The remaning anchors resemble *v duxe* ‘in the spirit of’ in ID 757 in being instances of incipient grammaticalization from a free prepositional phrase to a derived preposition: *v stile* ‘in the style of’(ID 1214), *na maner* ‘on manner of’ (ID 1203), and *po tipu* ‘of the type of’ (ID 1204).

(15) ID 1214: **NP/VP v stil-e NP-Gen**

NP/VP in style-loc.sg NP-gen

*istori-i v stil-e Andersen-a*

story-nom.pl in style-loc.sg Andersen-gen.sg

‘stories in the style of Andersen’

(16) ID 1203: **NP/VP na maner NP-Gen**

NP/VP on manner.acc.sg NP-gen

*rubašk-a na maner uniform-y*

shirt-nom.sg on manner.acc.sg uniform-gen.sg

‘a uniform-style shirt’

(17) ID 1204: **NP/VP po tip-u NP-Gen**

NP/VP along type-dat.sg NP-gen

*golovn-aja bol*’ *po tip-u migren-i*

head-nom.sg.f ache.nom.sg along type-dat.sg migrane-gen.sg

‘a migrane-type headache’

The remaining two constructions ID 1232 and ID 1233 contain anchors that almost exclusively modify a VP but are starting to drift towards the NP/VP possibility, similar to their counterparts of the *v duxe* type.

(18) ID 1232: **VP po primer-u NP-Gen**

VP along example-dat.sg NP-gen

*Po primer-u starš-ej sestr-y on sta-l*

along example-dat.sg elder-gen.sg.f sister-gen.sg he.nom become-pst.m

*rabota-t’ v muze-e.*

work-inf in museum-loc.sg

‘Just like his big sister he got a job in a museum.’

(19) ID 1233: **VP po obrazc-u NP-Gen**

VP along model-dat.sg NP-gen

*Kofejn-ja sozda-n-a po obrazc-u*

coffee\_house-nom.sg create-ptcp.pass-f along model-dat.sg

*parižsk-ix uličn-yx kafe.*

parisian-gen.pl outdoor-gen.pl cafe.gen.pl

‘The coffee house is modelled after Parisian outdoor cafes.’

Each of these constructions has its own preferences for lexical fillers of open slots, which supports the approach of representing them as distinct constructions with idiosyncratic combinatorial properties.

**5.2. Adding constructions with antonymous semantics and similar syntax**

It is possible to add not only individual constructions that are synonymous or antonymous to already existing entries, but also to add whole families of constructions that demonstrate semantic asymmetries. We illustrate this with the story of how we started with three constructions encoding a Core Addressee (exemplified in (20) and (21)) and the Benefactive participant (22) (= (1), reproduced here for the benefit of the reader), and added antonymous constructions that encode a Maleficiary (23)-(27).

The construction with the anchor *v adres* ‘to’ (lit. ‘in address’) (ID 752) encodes the Addressee of a speech act, *v otnošenii* ‘in regard’ (lit. ‘in relation’) (ID 789) refers to the Addressee of a non-verbal action, and *v ugodu*‘in order to please’ lit. ‘in favor’) (ID 831) denotes the Addressee of an action who benefits from that action.

(20) ID 752: **NP v adres NP-Gen**

NP in address.acc.sg NP-gen

*V adres Van-i poslyša-l-i-s*’ *oskorbleni-ja.*

in address.acc.sg Vanja-gen.sg be\_heard-pst-pl-med insult- nom.pl

‘Insults were aimed at Vanja.’

(21) ID 789: **VP/NP v otnošeni-i NP-Gen**

VP/NP in regard-loc.sg NP-gen

*Nedavno vve-l-i nov-ye štraf-y v otnošeni-i*

recently introduce-pst-pl new-acc.pl fine-acc.pl in regard-loc.sg

*avtomobilist-ov.*

motorist-gen.pl

‘New fines for motorists have been introduced recently.’

(22) ID 831: **NP/VP v ugod-u NP-Dat**

NP/VP in favor-acc.sg NP-dat

*On skaza-l èt-o v ugod-u načal*’*stv-u.*

he.nom say-pst.m this-acc.sg in favor-acc.sg boss.coll-dat.sg

‘He said that in order to please the bosses.’

Some constructions were added to Core Addessee and Beneficiary (see Figure 6). The entire new family of constructions encoding Maleficiary came into being as the antipode of Beneficiary with the opposite polarity value. This family included five constructions:

(23) ID 904: **VP nazlo NP-Dat**

VP out\_of\_spite NP-dat

*Nazlo muž-u ona obreza-l-a svo-i dlinn-ye volos-y.*

to\_spite husband-dat.sg she.nomcut-pst-f her-acc.pl long-acc.pl hair-acc.pl

‘In order to irritate her husband, she cut off her long hair.’

(24) ID 1786: **VP v pik-u NP-Dat**

VP in pique-acc.sg NP-dat

*Otkry-l-i magazin zdorov-oj ed-y v pik-u*

open-pst-pl store.acc.sg healthy-gen.sg.f food-gen in pique-acc.sg

*fastfud-u.*

fast\_food-dat.sg

‘They opened a health food store in order to undermine the fast food place.’

(25) ID 1787: **VP naperekor NP-Dat**

VP in\_spite\_of NP-dat

*Naperekor mnog-im kolleg-am on ver-it v*

in\_spite\_of numerous-dat.pl colleague-dat.pl he.nom believe-prs.3sg in

*telepati-ju.*

telepathy-acc.sg

‘As opposed to many of his colleagues, he believes in telepathy.’

(26) ID 1788: **VP v ukor NP-Dat**

VP in reproach.acc.sg NP-dat

*Gazet-a napečata-l-a rasskaz v ukor*

newspaper-nom.sg publish-pst-f story.acc.sg in reproach.acc.sg

*Xeminguè-ju.*

Hemmingway-dat.sg

‘The newspaper published a story reproaching Hemmingway.’

(27) ID 771: **VP v nazidani-e NP-Dat**

VP in edification-acc.sg NP-dat

*My pokaza-l-i èt-ot fil’m v nazidani-e*

we.nom show-pst-pl this-acc.sg.m film.acc.sg in edification-acc.sg

*potomk-am.*

descendant-dat.pl

‘We showed the film in order to enlighten the younger generation.’

As a result, we ended up with three families of constructions: Core Addressee, Beneficiary, and Maleficiary constructions visualized in Figure 6.

Diagram

Description automatically generated

Figure 6: Three families of constructions encoding Addressee, Maleficiary, and Beneficiary.

By contrast to Beneficiary, Maleficiary constructions encode the addressee of an action who is negatively affected by the action. In this regard, Beneficiary and Maleficiary constructions are two families with antonymous semantic relationships. The Beneficiary and Maleficiary semantics, in a way, specify the Core Addressee by assessing the impact (positive or negative) of the action on the participant of a situation.

The constructions of these two families employ the same syntactic structures. The first syntactic structure is **[VP v NP-Acc NP-Dat]**, where *v* ‘in’ is a preposition, the verb phrase VP encodes an action, the noun phrase NP-Dat encodes the Addressee of an action, and the slot NP-Acc is filled with a specific noun that determines whether the construction belongs either to Beneficiary or Maleficiary. The second syntactic structure features the preposition *na* ‘on’ instead: **[VP na NP-Acc NP-Dat]**. In the Maleficiary constructions ID 904 and ID 1787, the initial prepositional phrase *na NP-Acc* is grammaticalized into a single word (*nazlo* ‘to spite’ and *naperekor* ‘in spite of’ respectively), whereas in the remaining constructions of this family the anchor combination of *v* ‘in’ andNP-Acc allows two interpretations: it can be analyzed as a single partially grammaticalized complex preposition or as a free prepositional phrase (comparable to the constructions in Section 5.1).

To conclude, this case study serves to illustrate two crucial points. First, we can expand the inventory not only by building on semantic relationships among individual constructions (as in Section 5.1), but also semantic links across whole families (e.g., antonymous relationships of Beneficiary vs. Maleficiary). Second, we observe that semantic horizontal relationships serve as a stronger basis for establishing a single family than syntactic similarities (e.g., constructions that feature the same syntactic pattern **[VP v NP-Acc NP-Dat]** belong to two separate but related families motivated by Beneficiary vs. Maleficiary semantics), as proposed in Section 4.2.

**5.3. Adding constructions with alternative argument structure patterns related to the same lexical anchor**

While previous sections focused on semantic relationships like synonymy and antonymy, this section is devoted to adding constructions on the basis of their syntactic properties. In particular, we will show that it is fruitful to analyze a group of synonymous anchor predicates in terms of their argument structure and add constructions that represent alternative argument structure patterns related to the same lexical anchor. This strategy can be illustrated with a case study of a family of constructions that encode one’s support (or lack of support) for an idea, a participant, or a situation.

Figure 7 shows that all members of the Support family are copular constructions that can be viewed as instantiations of a single abstract schema **[NP Cop Pred Y]**. In this schema, Pred stands for a predicative expression that encodes the semantics of support (or lack of support) and has a valency Y that can be filled with complements of different types: nominal, infinitival, or sentential. The schemas at a lower level of abstraction specify the predicate: *protiv* ‘against’, *za* ‘for’ and *na storone* ‘on the side of’. The boxes represent individual constructions that differ in terms of 1) what type of complement (nominal, infinitival, or sentential) the predicate takes and 2) whether negation is an obligatory, optional, or impossible element of the anchor. The two constructions that feature sentential complements (ID 1803 and ID 43) can be viewed as specific instances (or subtypes) of the constructions with nominal complements (ID 1802 and ID 2347 & ID 1804 respectively) and can thus be analyzed as their child nodes. We represent these vertical relations by locating the constructions with sentential complements at the lower level of abstractness.

Diagram

Description automatically generated

Figure 7: Family of constructions encoding Support via predicatives with comparable argument structures.

Each of the three abstract schemas **[NP Cop protiv Y]**, [**NP Cop za Y]**, and **[NP Cop na storone Y]** has a negated sub-schema (**[NP Cop NEG protiv Y]**, [**NP Cop NEG za Y]**, and **[NP Cop NEG na storone Y]** respectively) that is realized differently depending on the syntactic properties of these predicates. Interestingly, in construction ID 455 with the infinitival complement of the predicate *protiv* ‘against’, the marker of negation *ne* ‘not’ is an obligatory element of the anchor, whereas in the constructions ID 1802 and ID 1803, where the same predicate takes a nominal and sentential complement accordingly, the negation is syntactically optional. Similarly, the construction ID 1136 with a nominal complement of the predicate *na storone* ‘on the side of’ can exist with or without a negation marker. Finally, the construction ID 1804 does not allow the possibility of negating the predicate *za* ‘for’. The negated version *ne za* ‘not for’ (encoding lack of Support) is only used when contrasted with the opposite attitude (presence of Support) to another participant, explicitly stated, as in the construction ID 2347.

Having considered the overall organization of this family of constructions, let us now explain how we arrived at this outcome. Initially, this family consisted of two constructions, ID 455 and ID 43, visualized in Figure 7 as shaded boxes, and five constructions were added, as shown by the white boxes. The construction ID 455 in (28) contains a negated matrix predicate *ne protiv* ‘not against’ that takes an infinitival complement VP-Inf without a complementizer.

(28) ID 455: **NP-Nom Cop ne protiv VP-Inf**

NP-nom cop neg against VP-inf

*Ja ne protiv poexa-t’ v derevnj-u.*

I.nom neg against go-inf in countryside-acc.sg

‘I am not against going to the countryside.’

The synonymous construction ID 43 in (29) contains the matrix predicate *za* ‘for’ that has a valency for a sentential complement, which is introduced by the complementizer *čtoby* ‘in order to’. The sentential complement can be represented by a finite or non-finite (infinitival) subordinate clause, as illustrated in (29-a) and (29-b).

(29) ID 43: **NP-Nom Cop (ne) za t-o, čtoby Cl/VP-Inf**

NP-nom cop neg for that-acc.sg.n in\_order\_to cl/VP-inf

**a.** *Ja za t-o, čtoby ty u nas ži-l.*

I.nom for that-acc.sg.n in\_order\_to you.nom by we.gen live-pst.m

‘I support the idea of you living with us.’

**b.** *Ja za t-o, čtoby poexa-t*’ *v otpusk.*

I.nom for that-acc.sg.n in\_order\_to travel-inf in vacation.acc.sg

‘I support the idea of going on vacation.’

Having these two constructions as a starting point, we aimed at representing other argument structure options available for these two predicates *za* ‘for’ and *protiv* ‘against’ (which are morphologically prepositions). In particular, each of these predicates can govern both sentential and non-sentential complements. Therefore, we added two constructions with the predicate *protiv* ‘against’, where it takes a nominal complement NP-Gen in ID 1802 (30) and a sentential complement Cl/VP-Inf in ID 1803 (31).

(30) ID 1802: **NP-Nom Cop (ne) protiv NP-Gen**

NP-nom cop neg against NP-gen

*Ja protiv škol’n-oj form-y.*

I.nom against school-gen.sg.f uniform-gen.sg

‘I am against school uniforms!’

(31) ID 1803: **NP-Nom Cop (ne) protiv (t-ogo), čtoby Cl/VP-Inf**

NP-nom cop neg against that-gen.sg.n in\_order\_to cl/VP-inf

*Ja ne protiv, čtoby ty na nej ženi-l-sja.*

I.nom neg against in\_order\_to you.nom on she.loc marry-pst.m-med

‘I am not against you marrying her.’

In the construction ID 1803 (31), the pronoun *to* ‘that’ in the genitive case fills the slot NP-Gen of ID 1802. Therefore, ID 1802 can be analyzed as a parent of ID 1803. The pronoun *to* ‘that’ helps to connect the matrix clause to its complement in ID 1803, but it can also be omitted.

Following the same logic, we added the construction ID 1804 (32), where the predicate *za* ‘for’ takes not a sentential (as in the already present construction ID 43), but a nominal complement NP-Acc.

(32) ID 1804: **NP-Nom Cop za NP-Acc**

NP-nom cop for NP-acc

*Ja za revoljuci-ju.*

I.nom for revolution-acc.sg

‘I support the idea of revolution.’

Because ID 1804 does not allow negation (\**ne za NP-Acc*), as opposed to ID 1802 (*ne protiv NP-Gen*), we explored broader contexts, where *ne za* ‘not for’ is attested. We found that the vast majority of attestations contrast the negated complement with a non-negated one, as in (34). We represented this use as a separate entry: ID 2347.

(34) ID 2347: **NP-Nom Cop ne za NP-Acc, a/no za NP-Acc**

NP-nom cop neg for NP-acc and/but for NP-acc

*Oni by-l-i ne za kapitalizm, a za*

they.nom be-pst-pl neg for capitalism.acc.sg but for *demokratij-u.*

democracy-acc.sg

‘They supported not capitalism, but democracy.’

Another added copular construction, ID 1136, with similar semantics of Support, features the predicate *na storone*, which is a prepositional phrase lit. ‘on side’ (35). This construction has two variants: the participant who receives support from the subject can be expressed either by a possessive pronoun or by a nominal phrase in the genitive. Because this participant is normally animate, the predicate *na storone* lit. ‘on side’ takes only a nominal complement, whereas the infinitival and sentential complements available for *protiv* ‘against’ and *za* ‘for’ are not an option.

(35) ID 1136: **NP-Nom Cop (ne) na PronPoss-Loc storon-e /na storon-e NP-Gen**

NP-nom cop (neg) on PronPoss-loc side-loc.sg/on side-loc.sg NP-gen

*V èt-om spor-e ja na vaš-ej storon-e.*

in this-loc.sg.m debate-loc.sg I.nom on your-loc.sg.f side-loc.sg

‘In this debate I am on your side.’

Summing up, the family of constructions encoding the attitude of Support features three predicates that have somewhat similar but not identical argument structure patterns: e.g., the infinitival complement is allowed only by the predicate *ne protiv* ‘not against’, but not by the other two predicates; *na storone* ‘on the side of’ can only take a nominal complement, whereas *za* ‘for’ can take both nominal and sentential complements, but has peculiar idiosyncratic restrictions on negation. Overall, taking into account a range of argument structure patterns and negation options available for the three predicates in question made it possible to achieve a more consistent representation of individual constructions and significantly expand this family by identifying additional members.

**5.4. Growing an entire family out of a single construction**

So far, we have shown how we added missing members of constructional families by searching for synonymous or antonymous anchors (or constructions) or similar syntactic structures. In this subsection we show that sometimes an entire family can be established on the basis of a single construction with productive or cognitively relevant semantics. We illustrate this with a family of Mirative constructions that came together thanks to a single construction in (36) that encodes Disappearance of the key participant of a situation that takes place unexpectedly for the speaker:

(36) ID 619: **Cl – i bud’ zdorov**

cl and be.imp.sg healthy.nom.sg.m.short

‘Cl **–** and keep well’

*Miša zabra-l knig-u – i bud’ zdorov.*

Michael take-pst.m book-acc.sg and be.imp.sg healthy.nom.sg.m.short

‘Michael took the book and unexpectedly disappeared.’

The lexical anchor of construction ID 619 is the expression *bud’ zdorov* that literally encodes wishing health to the interlocutor (used when the interlocutor sneezes). In this construction the anchor is employed in another frequent function, that of a farewell politeness formula meaning ‘Keep well! Good luck! Good-bye!’ that is arguably semantically bleached.

This construction simultaneously belongs to the semantic type Mirative (in the Subjectivity class) and the semantic type Non-Existence, subtype Disappear (in the Qualia class). The combination of these two semantic components, the disappearance of a key participant of a situation (a human or an object) and unexpectedness of this event to the speaker is cognitively relevant, as it can be expressed in Russian by the six constructions that we added with similar semantics, but quite different idiosyncratic syntactic structures illustrated in (37)-(42).

(37) ID 1971: **Cl, i privet!**

cl, and hello

*On uže davno vzja-l u menja èt-u*

he.nom already long\_ago take-pst.m from I.gen this-acc.sg.f

*knig-u, i privet! Uexa-l v otpusk.*

book-acc.sg and hello leave-pst.m in vacation.acc.sg

‘He borrowed my book a long time ago and suddenly disappeared [lit. and hello]. He left for vacation.’

(38) ID 1972: **Cl – i tju-tju / i tjutju![[12]](#footnote-12)**

cl **–** and intj and intj

*On zabra-l mašin-u – i tju-tju! Bol’še my ne*

he.nom take-pst.m car-acc.sg and intj more we.nom neg

*vstreča-l-i-s’.*

meet-pst-pl-med

‘He took the car and suddenly disappeared. I never saw him again.’

(39) ID 2063: **Cl, (a) NP-Gen i sled prosty-l**

cl but NP-gen ptcl trace get\_cold-pst.m

‘Cl, and someone*’*s track went cold.’

*Kogda vs-e prosnu-l-i-s’, ego i*

when everyone-nom.pl wake\_up-pst-pl-med he.gen ptcl

*sled* *prosty-l.*

track.nom.sg get\_cold-pst.m

‘When everyone woke up, he was already long gone.’

(40) ID 2064: **Cl, tol’ko PronPers-Acc i vide-l-i**

cl only PronPers-acc ptcl see-pst-pl

‘Cl, he was gone in a flash. / That was the last they saw of him.’

*On vybeža-l iz kabinet-a, tol’ko ego i vide-l-i.*

he.nom run\_out-pst.m from office-gen.sg only he.acc ptcl see-pst-pl

‘He ran out of the office and that was the last they saw of him.’

(41) ID 1302: **NP-Gen kak ne byva-l-o**

NP-gen as neg exist-pst-n

*Prostud-y kak ne byva-l-o.*

cold-gen.sg as neg exist-pst-n

‘The cold unexpectedly disappeared (as if it was never there).’

(42) ID 1970: **plakat’-Pst PronPoss-Nom NP-Nom!**

cry-pst PronPoss-nom NP-nom

*Plaka-l moj čemodan!*

cry-pst.m my.nom.sg.m suitcase.nom.sg

‘My suitcase is gone for good!’

As demonstrated in Figure 8, this family includes biclausal and monoclausal constructions. In the biclausal constructions the first clause is an open slot that describes the situation prior to the disappearance. Biclausal constructions come in two versions. The subschema **[Cl i X]** (where X is a fixed expression as an anchor) accounts for the constructions ID 619, ID 1971, and ID 1972. The other biclausal schema **[Cl (Conj) X]** (where X is a fixed clause containing an open slot) accounts for the constructions ID 2063 and ID 2064. Two remaining constructions are monoclausal: ID 1302 and ID 1970.

Diagram

Description automatically generated

Figure 8: Family of Mirative constructions that signify disappearance of a key participant.

The constructions in this family cannot be united under a single abstract syntactic pattern. Still, their syntactic variety does not undermine their semantic similarity.

**5.5. Adding highly frequent child constructions as distinct entries**

Frequency is an important factor that contributes to the status of a construction. High frequency is a symptom and at the same time the reason why certain constructions become highly entrenched in language use and in speakers’ minds. Following this line of thinking, individual instantiations of a more general pattern, when highly frequent and entrenched, can qualify as distinct constructions. This idea motivated us to look at some of our general patterns from a different angle and represent certain child constructions as separate entries that deserve their own description. Often this involved a construction with two (or more) slots that were mutually dependent in terms of fillers. In other words, a certain lexical filler in one slot can only combine with a very restricted set of fillers in the other slot. In this section we propose adding highly frequent (and therefore entrenched) child constructions and representing them as distinct entries of the database as another strategy for expanding the constructicon inventory.

We illustrate this idea with a case study of a family of Minimizing constructions that are characterized with two open slots (NP and VP) and a double negation pattern manifested in the combination of two negation markers *ni* and *ne,* both meaning ‘not’. The VP slot is filled with verbs denoting various kinds of actions. The NP slot is filled with head nouns that signify the smallest bit of a substance (physical or metaphorical) that can be affected by the activity or involved in a situation. The overall semantics of the Minimizing construction **[ni NP ne VP]** is an intensified denial that an action takes place. This semantics is motivated by the idea that denying the smallest affected portion or bit of a substance equals denying the activity. In other words, the activity does not affect even the smallest bit of a substance and therefore cannot be viewed as happening.

We started expanding this family from two initial entries: a very generally formulated pattern ID 213 **ni NP-Gen ne VP** (43) and its “grandchild” construction ID 229 **ni kapli NP-Gen net** (44), where the predicative *net* ‘no’ is a fused combination of the negation marker *ne* ‘not’ and a third person singular form *est’* ‘is’ of the existential verb *byt’* ‘be’: *ne* *est’ > nest’ > net.*

(43) **ID 213:** **ni NP-Gen ne VP**

not NP-gen neg VP

*Za den’ on ne pročita-l ni stranic-y.*

during day.acc.sg he.nom neg read-pst.m not page-gen.sg

‘During that day (in that day) he did not read a single page.’

(44) **ID 229:** **(u NP-Gen) ni kapl-i NP-Gen net**

(by NP-gen) not drop-gen.sg NP-genno

*U tebja net ni kapl-i voobraženi-ja.*

by you.gen no not drop-gen.sg imagination-gen.sg

‘You do not have the least bit of imagination.’

Construction ID 229 can be analyzed as a specific instantiation of ID 213, because the anchor*kapli* of ID 229 is the lexical filler of the open slotNP-Gen of ID 213. In construction ID 229, the anchor *kapli* is also a head noun that governs another NP-Gen, a slot that can be filled with nouns denoting liquids (e.g. rain, wine) but much more often with nouns encoding feelings, emotions, and mental capacities, as illustrated in (44) with the expression *ni kapli voobraženija* *–* lit. ‘not a drop of imagination’. The noun *kaplja* ‘drop’ can be used both literally and metaphorically as a minimal quantum of fluid substance or an emotion construed as a fluid substance. Such metaphorical use is not characteristic of construction ID 213, yet it can be considered a “grandparent” of construction ID 229 in that the former clearly belongs to a lesser level of abstraction. With constructions ID 213 and ID 229 as a starting point, and having analyzed their relationship, we concluded that this family includes constructions that are more specific than ID 213 but more abstract than ID 229 and thus occupy a niche between them. These are constructions like ID 2345 **ni kapli ne VP** ‘to X not a drop’ (45) and its sisters that lexically specify the minimal quantum of an activity.

(45) **ID 2345: ni kapl-i ne VP**

not drop-gen.sg neg VP

*Ja ni kapl-i ne somneva-ju-s’.*

I.nom not drop-gen.sg neg doubt-prs.1sg-med

‘I have not a bit of doubt.’ (lit. I not a drop not doubt)

Both qualitative and quantitative analysis of corpus data led us to conclude that certain lexical fillers of the NP slot (like *kaplja* ‘drop’) in construction ID 213 **ni NP-Gen ne VP** areespecially productive and more frequent than other lexical fillers. In particular, these are nouns that refer to the smallest bits of the most basic activities: eating, drinking, speaking, motion in time and space, and dealing with money.[[13]](#footnote-13) The corresponding nouns have individual idiosyncratic restrictions on the fillers of the other slot, the VP, and form a whole range of highly frequent constructions that are both lexically specified and combinatorially idiosyncratic. Prominent minimizers include: *ni kapli* ‘not a drop’, *ni glotka* ‘not a sip’, *ni kroški* ‘not а crumb’, *ni kopejki* ‘not a kopeck’ (the smallest unit of Russian money), *ni groša* ‘not a half-kopeck piece’, *ni šagu* ‘not a step’, *ni slova* ‘not a word’, *ni minuty* ‘not a minute’, *ni sekundy* ‘not a second’, and *ni razu* ‘not a single time’, among others.[[14]](#footnote-14) Each of these minimizers can be glossed as ‘next to nothing’ and denotes the smallest bit that can be used to measure an activity. One more highly entrenched minimizer *ni čerta* ‘not a devil’ makes use of the swear word *čert* ‘devil’ and allows for an even wider range of predicates (compared to the minimizers listed above) and arguably no longer refers to a minimal quantum of an activity, being rather used as a more generalized minimizer that is in the process of turning into an adverb with idiomatic word-final stress (and even spelled sometimes as a single word).

These observations are supported by a quantitative corpus study, where we compared the overall number of attestations of these lexical fillers in the parent construction ID 213 **ni NP-Gen ne VP**. Because the constructions under scrutiny are very colloquial, we face the challenge of choosing an appropriate corpus that would represent spoken discourse. The Russian National Corpus (available at [https://ruscorpora.ru](https://ruscorpora.ru/)) is primarily comprised of written texts. We found that its Modern Subcorpus (texts created 1950-2021) provides too sparse data on some of the constructions in question. Instead, we used the Araneum Russicum Maximum corpus (available at [http://aranea.juls.savba.sk](http://aranea.juls.savba.sk/)), arguably the largest web-harvested corpus for Russian that contains 20 billion tokens (Benko & Zakharov, 2016). This corpus is a collection of texts culled from Russian websites (containing forums, blogs, social media, etc.) and therefore is more representative of spoken Modern Russian of the last two decades. As such, the Araneum Russicum Maximum corpus is recommended for studies of on-going phonological changes, newly coined neologisms and spoken language phenomena including colloquial constructions (Piperski, 2020). Table 2 lists those child instantiations of the parent ID 213 **ni NP-Gen ne VP** that turned out to be highly frequent and are currently represented as distinct entries in the Russian Constructicon. Because Russian has relatively free word order, we performed the search for two alternative sequences charateristic of the **ni NP-Gen ne VP** construction: the negated minimizer NP preceding or following the VP (columns 4 and 5 respectively).[[15]](#footnote-15) The constructions are listed according to the descending number of total hits provided in the right-most column.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Construction | Gloss | Araneum Russicum Maximum | | |
| N hits for  "ni NP-Gen  ne VP" | N hits for  "ne VP  ni NP-Gen" | Total N  of hits |
| 1690 | **ni razu ne VP** | ‘not a single time VP’ | 232 108 | 33 689 | 265 797 |
| 1777 | **ni slova ne VP** | ‘not a word VP’ | 15 936 | 31 743 | 47 679 |
| 1161 | **ni kopejki ne VP** | ‘not a kopeck VP’ | 3 795 | 12 243 | 16 038 |
| 2345 | **ni kapli ne VP** | ‘not a drop VP’ | 10 511 | 3 104 | 13 615 |
| 2346 | **ni minuty ne VP** | ‘not a minute VP’ | 5 407 | 5 271 | 10 678 |
| 2346 | **ni sekundy ne VP** | ‘not a second VP’ | 4 181 | 3 519 | 7 700 |
| 1157 | **ni čerta ne VP** | ‘not a devil VP’ | 6 405 | 555 | 6 960 |
| 1161 | **ni groša ne VP** | ‘not a half-kopeck VP’ | 280 | 1 800 | 2 080 |
| 1158 | **ni šagu ne VP** | ‘not a step VP’ | 296 | 583 | 879 |
| 1159 | **ni kroški ne VP** | ‘not а crumb VP’ | 144 | 494 | 638 |
| 1160 | **ni glotka ne VP** | ‘not a sip VP’ | 10 | 96 | 106 |
| NA | **ni stranicy ne VP** | ‘not a page VP’ | 8 | 27 | 35 |

Table 2: Raw numbers of attestations of Minimizing constructions in corpus data.

The collected data facilitate several crucial observations. First, these constructions are very different in terms of token frequency. For example, temporal minimizers*ni razu*‘not a single time’, *ni minuty*‘not a minute’, *ni sekundy*‘not a second’ are more frequently used than the minimizers that refer to food (*ni kroški*‘not а crumb’, *ni glotka*‘not a sip’). Nevertheless, we decided to represent the latter ones as distinct entries due to their cognitive relevance and higher frequency than that of many other lexical minimizers that can appear in the **ni NP-Gen ne VP** construction. In particular, in the bottom line of Table 2 we provide the number of attestations for **ni stranicy ne VP** ‘not a page VP’ that is considerably less frequent than **ni glotka ne VP** ‘not a sip VP’ and is accounted for as part of the ID 213 **ni NP-Gen ne VP** construction and not as a distinct entry (indicated with NA instead of the ID number in the left-most column). Second, the quantitative data in Table 2 reveal that in addition to idiosyncratic restrictions on the fillers of the VP slot, these child constructions have different word order preferences. Some of them (ID 2345 **ni kapli ne VP**‘not a drop VP’*,* ID 1157**ni čerta ne****VP**‘not a devil VP’) clearly prefer the VP to be preceded by the negated NP, whereas others (ID 1161 **ni kopejki ne VP** ‘not a kopeck VP’ and ID 1161 **ni groša ne VP** ‘not a half-kopeck VP’) are more often attested with the opposite word order, having the VP after the NP.

Figure 9 visualizes the family of Minimizing constructions that emerged from family-based expansion. The constructions are represented as boxes, and shading highlights the two constructions that were in this family prior to family-based expansion of the inventory. The remaining white boxes visualize the nine constructions that were added. All of them are child constructions of ID 213 **ni NP-Gen ne VP**, and vertical lines indicate the inheritance links that connect them to the parentID 213 construction. The red color highlights the fact that the lexical minimizers correspond to the fillers of the NP-Gen slot of their parent construction. The ID 213 construction can be traced to an even more abstract schema [ni X ne VP] that serves as a parent for two other families of Minimizing constructions, where X can be an adverb (e.g. ID 1451 **niskol’ko ne VP** ‘none at all VP’) or a PP (ID 1452 **ni na gramm ne VP** ‘lit. not on a gram VP’) that we leave beyond the scope of this study. The horizontal lines connect constructions ID 2345, ID 1160, ID 1159 and ID 1161 on the left of the diagram as referring to physically accessible minimizers (drop, sip, crumb, and kopeck). On the right end of the diagram the horizontal lines indicate conceptual closeness of two temporal minimizers in ID 2346 and ID 1690.

Diagram

Description automatically generated

Figure 9: Family of Minimizing constructions with double negation.

On the one hand, we have kept the ID 213 **ni NP-Gen ne VP** as a distinct entry to account for less frequent lexical fillers like *ni stranicy* ‘not a page’, *ni stroki* ‘not a line’, *ni zvuka* ‘note a sound’, *ni sleda* ‘not a trace’, *ni duši* ‘not a soul’, etc. On the other hand, we represent the two most frequent monetary minimizers *ni kopejki* ‘not a kopeck’ and *ni groša*‘not a half-kopeck’ as a single entry ID 1161 **ni kopejki/groša ne VP**, where the slash showsthem as alternative anchor words. The same principle lies behind the representation of the two most frequent temporal minimizers*ni minuty*‘note a minute’ and *ni sekundy*‘note a second’ as a single entry ID 2346 **ni minuty/sekundy ne VP**.

**5.6 Summary of family-based expansion**

Summing up, we have looked at various ways to expand constructional families, which are groups of closely related constructions with relatively homogeneous semantics and often similar syntactic (or structural) properties. We leave beyond the scope of this study expansion to constructions that share the same lexical anchors. By contrast with the five strategies of family-based expansion presented above, this method is not based on constructional families. As it turns out, the newly added constructions with the same anchor word tend to belong to very different, often unrelated, families. Some lexemes can be extremely productive in forming partially schematic constructions. For Russian, such popular anchor words include the nouns *raz* ‘one time, occasion’ (with more than thirty different constructions according to Iomdin, 2015), *vid* ‘look’, *sila* ‘force, power’, *storona* ‘side’, *vremja* ‘time’ (Iomdin, 2019), the word *čto* ‘that’ (a subordinating conjunction or relative pronoun, cf. Iomdin, 2013), and the swear word *čert* ‘devil’, among others.

**6. Conclusions**

The contribution of this article is threefold, as it has methodological, empirical, and theoretical dimensions. First, we propose and elaborate a new methodology for expanding the size of a constructicon inventory. This methodology has proved to be efficient and successful for the Russian Constructicon and can potentially serve for building or expanding constructicons of other languages. This methodology may ultimately be useful for working on languages of any structure and type, as long as the language in question has families of constructions where the family members are linked via synonymous, antonymous, syntactic, or other structural relationships. Second, we detailed the proposed methodology in terms of five concrete strategies for working with families of constructions and provided five empirical case studies of constructional families that illustrate how these strategies work in practice. Third, the theoretical contribution consists in that we explored units larger than a single construction, i.e., families of related constructions, and offered our understanding of this phenomenon based on our experience of identifying families for over 2,200 constructions in the Russian Constructicon.

In a nutshell, family-based expansion focuses on identifying families of constructions, understanding the organization of a family, and adding missing members that compete with, alternate with, or complement those constructions that are already accounted for in the database. This methodology makes it possible to significantly expand and systematize the existing inventory of constructions and represent the various types of relationships that hold constructions together and structure a constructicon. The methodology of family-based expansion is based on the analysis of both vertical (hierarchical) and horizontal (lateral, or paradigmatic) relationships among constructions. The crucial advantage of the proposed methodology is that it elaborates a constructicon in both quantitative and qualitative ways: the inventory becomes larger and is transformed from a list of collected items into a structured inventory. In other words, it entails both quantitative expansion and qualitative systematization (and harmonization) of the existing inventory of constructions.

The Russian Constructicon resource continues to grow. The next phase of inventory expansion that we are currently exploring is based on the thorough analysis of individual constructions. Writing the entry for each construction entails undertaking a small case study that necessarily involves a fine-grained analysis of corpus data. This process often reveals homonymous patterns or the closely related yet different constructions that deserve to be added as distinct entries in our database.

**References**

Aikhenvald, A. (2012). The essence of mirativity. *Linguistic Typology*, *16*, 435–485.

Benko, V. & Zakharov, V. (2016). Very Large Russian Corpora: New Opportunities and New Challenges. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”*, *15*(22), 79–93. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet.

Cappelle, B. (2006). “Particle Placement and the Case for “Allostructions.” *Constructions Online 1* (7), 1–28.

Comrie, B., Haspelmath, M., Bickel, B. (2008). The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. Max Planck Institute for Evolutional Anthropology. <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

Croft, W. (2001). *Radical construction grammar. Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Diessel, Holger. (2015). Usage-based construction grammar. In Ewa Dąbrowska and Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 295–321. Berlin: Mouton de Gruyter.

Diessel, Holger. (2019). *The grammar network: How linguistic structure is shaped by language use.* Cambridge: Cambridge University Press.

Ehrlemark, A., Johansson, R. & Lyngfelt, B. (2016). Retrieving Occurrences of Grammatical Constructions. *Proceedings of COLING 2016*, *the 26th International Conference on Computational Linguistics: Technical Papers*, 815–824. Osaka, Japan.

Endresen, A. (2019). Two origins of the prefix IZ- and how they affect the VY- vs. IZ- correlation in Modern Russian*. Russian Linguistics: International Journal for the Study of Russian and other Slavic Languages, 43*(3), 205–229.

**Endresen, A. &** A. Janda, L. A. (2020). Taking Construction Grammar one step further: Families, clusters, and networks of evaluative constructions in Russian. In M. Putnam, M. Carlson, A. Fábregas & E. Wittenberg (Eds.), Defining Construction: Insights into the Emergence and Generation of Linguistic Representations [special issue of *Frontiers in Psychology,* *11*] (pp. 1–22).

Fillmore, Ch. J. (1988). The Mechanisms of “Construction Grammar. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55.

Goldberg, A. E. (2002). Surface Generalizations: An alternative to alternations. *Cognitive Linguistics 13*(4), 327-356.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.

Goldberg, A. E. (2013). Constructionist approaches. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 15–31). Oxford: Oxford Unievrsity Press.

Goldberg, A. E. & Herbst, T. (2021). The nice-of-you construction and its fragments. *Linguistics,* *59*(1), 285–318.

Goldberg, A. E. & Jackendoff, R. (2004). The resultative as a family of constructions. *Language,* *80*(3), 532–68.

Janda, L. A. & Clancy, S. J. (2002). *The Case Book for Russian*. Bloomington, IN: Slavica Publishers.

Janda, L. A. (2015). *Cognitive Linguistics in the year 2015*. *Cognitive semantics,* *1*, 131–154.

Janda, L. A., **Endresen**, A., Zhukova, V., Mordashova, D., & Rakhilina, E. (2020). How to build a constructicon in five years: The Russian Example. In F. Brisard, T. Colleman, A. De Wit, R. Enghels, N. Koutsoukos, T. Mortelmans & M. Sol Sansiñena (Eds.), *The wealth and breadth of construction-based research* [special issue of *Belgian Journal of Linguistics*, *34*] (pp. 162–175).

Janda, L. A., **Endresen**, A., Zhukova, V., Mordashova, D. & Rakhilina, E. (Forthc.). From data to theory: an emergent classification based on the large-scale Russian constructicon. *Constructions and frames*.

Janda, L. A., Lyashevskaya, O., Nesset, T., Rakhilina, E. & F. M. Tyers. (2018). Chapter 6. A Constructicon for Russian: Filling in the Gaps. In B. Lyngfelt, L. Borin, K. Ohara & T. T. Torrent (Eds.), *Constructicography: Constructicon development across languages* [Constructional Approaches to Language 22] (pp. 165–181). Amsterdam: John Benjamins.

Iomdin, L. L. (2013). Nekotorye mikrosintaksičeskie konstrukcii v russkom jazyke s učastiem slova čto v kačestve sostavnogo komponenta [Certain microsyntactic constructions in Russian which include the word čto as a constituent element]. *Južnoslovenski Filolog,* *69*, 137–147.

Iomdin, L. L. (2015). Konstrukcii mikrosintaksisa, obrazovannye russkoj leksemoj raz [Microsyntactic constructions formed by the Russian word *raz*]. *SLAVIA časopis pro slovanskou filologii,* *84*(3), 291–306.

Iomdin, L. L. (2019). Russkie mikrosintaksičeskie elementy, motivirovannye slovom *vid*: korpusnoe issledovanie semantiki[Russian microsyntactic elements derived from the noun *vid*: a corpus study of semantics]. *Proceedings of the international conference “Corpus linguistics-2019”*, (pp. 189–202).St. Petersburg: St. Petersburg State University Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jakobson, R. (1957). *Shifters, verbal categories and the Russian verb*. Cambridge, MA: Harvard University Press.

Jespersen, O. (1924). *The Philosophy of Grammar.* London: G. Allen and Unwin. Reprinted in 2010 by Routledge.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. v. 1. Stanford: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.

Lyngfelt, B. (2018). Introduction: Constructicons and constructicography. In B. Lyngfelt, L. Borin, K. Ohara & T. T. Torrent (Eds.), *Constructicography: Constructicon development across languages* (pp.1–18).John Benjamins, Amsterdam.

Mel’čuk, I. A. (1998). *Kurs obščej morfologii. T. II. Part 2: Morfologičeskie značenija* [A course in general morphology, Vol. 2, Part 2: Morphological meanings]. Moskva: Jazyki slavjanskoj kul’tury; Wien: Wiener Slavistischer Almanach.

Nida, E. (1949). *Morphology: The descriptive analysis of words*. Ann Arbor: University of Michigan Press.

Piperski, A. (2020). Russkij jazyk i korpusnoe raznoobrazie [Russian language and corpus diversity]. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”,* *19*, 1–13.

Plungian, V. A. (1999). A typology of phasal meanings. In W. Abraham & L. Kulikov (Eds.), *Tense-aspect, transitivity, and causativity: Essays in honor of Vladimir Nedjalkov* (pp. 311–321). John Benjamins, Amsterdam.

Plungian, V. A. (2011). *Vvedenie v grammatičeskuju semantiku: Grammatičeskie značenija i grammatičeskie sistemy jazykov mira* [An introduction to grammatical semantics: Grammatical meanings and grammatical systems in the languages of the world]. Moscow: Russian State University for the Humanities Press.

Rakhilina, E. V. (Ed.). (2010). Lingvistika konstrukcij [Linguistics of constructions]. Moscow: Azbukovnik.

Ruiz de Mendoza Ibáñez, Francisco José, Alba Luzondo Oyón, Paula Pérez-Sobrino. (2017). *Constructing Families of Constructions: Analytical Perspectives and Theoretical Challenges*. Amsterdam: John Benjamins.

Schmid, H.-J. (2020). *The dynamics of the linguistic system: usage, conventionalization and entrenchment*. Oxford: Oxford University Press.

Švedova, N. Ju. (1960). *Očerki po sintaksisu russkoj razgovornoj reči* [Outlines of the syntax of conversational Russian]. Moscow: Izdatel’stvo Akademii nauk SSSR.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen. (Ed.). *Language typology and syntactic description*. *Vol. III: Grammatical categories and the lexicon* (pp. 36–149). Cambridge: Cambridge University Press.

Talmy, L. (2000). Lexicalization patterns, Surveying lexicalization patterns. In L. Talmy (Ed.). *Towards a cognitive semantics. Vol. II: Typology and process in concept structuring* (pp.21–212)*.* Cambridge, MA: MIT.

Treis, Y. (2018). Comparative Constructions: An Introduction. *Linguistic Discovery*, *16*(1), i–xxvi.

Van den Heede, M. & Lauwers, P. (2021). Het interesseert me geen ene je-weet-wel: The productivity of minimizing constructions in present-day Dutch. A talk presented at the International Conference on Construction Grammar (ICCG11), Antwerp August 18-20, Belgium.

Wittgenstein, L. (1953). *The philosophical investigations.* Oxford: Blackwell.

**Abbreviations**

1–3 – 1st–3rd person, acc – accusative, adv – adverbializer, Cl – clause, coll – collective, cop – copula, dat – dative, f – feminine, fut – future, gen – genitive, imp – imperative, indef – indefinite, inf – infinitive, ins – instrumental, intj – interjection, ipfv – imperfective, loc – locative, n – neutrum, neg – negation, nom – nominative, NP – noun phrase, m – masculine, med – middle voice, pass – passive, pfv – perfective, pl – plural, PronPers – personal pronoun, PronPoss – possessive pronoun, prs – present, pst – past, ptcl – particle, ptcp – participle, sg – singular, short – short form of an adjective, VP – verb phrase, () – optional element.

**Acknowledgments**

This work was supported by the Directorate for Higher Education and Skills, under the Norwegian Ministry of Education and Research (HK-dir, https://hkdir.no/, the grant CPRU-2017/10027), the Ministry of Science and Higher Education of the Russian Federation (the Agreement No 075-15-2020-793; 2021), and the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A03065528). We thank UiT The Arctic University of Norway for their support of our research.

**Running head:** Family-based expansion of the Russian Constructicon

1. This and all examples cited in the Russian Constructicon and in this article are edited versions of corpus examples from the Russian National Corpus ([https://ruscorpora.ru](https://ruscorpora.ru/)). [↑](#footnote-ref-1)
2. The number of constructions in a constructicon is largely a matter of how to count them, it heavily depends on the level of granularity in the representations of distinct constructions chosen by the creators of the resource. Because of the pedagogical purposes of our resource, we chose a relatively granular level of description that is optimal for practical considerations. We elaborate this point in Section 4. [↑](#footnote-ref-2)
3. In this case, rather rare in our taxonomy of semantic types of constructions, we use terminology that overlaps with terms normally employed for semantic roles. The semantic type Addressee and its subtypes Beneficiary, Maleficiary, and Audience, mentioned here, refer to the semantics of the whole constructions, and the terms indicate various ways of interaction between the participants of a situation (see Section 5.2). [↑](#footnote-ref-3)
4. All terms used in the Russian Constructicon are fully defined in the Instructions page, and are motivated by conventions of typological linguistics. [↑](#footnote-ref-4)
5. Non-existence is a semantic type that goes beyond simple negation. In particular, it contains families of constructions that encode Dissappearance of a key participant of a situation (Section 5.4) and a family of Minimizing constructions that signify absence of an action by negating its smallest portion (Section 5.5). Non-existence contains many other peculiar constructions like ID 1455 **NP na nule** (as in *immunitet na nule* ‘zero-immunity’) that go beyond grammatical negation. Attitude and Assessment are two densely populated types of Evaluative constructions that cover a wide range of semantic nuances described in (Endresen & Janda, 2020): e.g. ID 500 **NP-Nom Cop tak sebe** (as in *Kartina tak sebe* ‘The painting is so-so’) encoding negative Assessment vs. ID 482 **NP-Dat net dela do NP-Gen** (as in *Mne net dela do tvoix deneg* ‘I don’t care about your money’) encoding negative Attitude. A case study of Attitude constructions encoding Support (Section 5.3) gives a flavor of the semantic complexity of this type that goes beyong grammatical means of Modality that it might be associated with. The semantic type Salient property is also very diverse but for the sake of argument can be illustrated with a construction that features repetition of the noun: ID 1445 **takoj Noun-Nom - ~Noun-Nom** (as in *Ona takaja devočka-devočka* ‘She’s such a girly girl’). [↑](#footnote-ref-5)
6. See the English description at <https://constructicon.github.io/russian/semantic-types/> and the Russian description at <https://constructicon.github.io/russian/semantic-types-russian/>. [↑](#footnote-ref-6)
7. An example of an Assessment construction is ID 2160 **NP-Nom Cop xorošij-Short/ploxoj-Short NP-Ins**, as in *Èti mesta xoroši svoimi lesami* ‘These places are good in terms of their forests’. [↑](#footnote-ref-7)
8. An example of a Prohibitive construction is ID 231 **ni slova NP-Dat o NP-Loc!**, as in *Ni slova mame o našej poezdke!* ‘Not a word to mom about our trip!’. [↑](#footnote-ref-8)
9. An example of an Inclusive construction is ID 441 **NP, v tom čisle (i) NP**, as in *Živye suščestva, v tom čisle i rastenija, p’jut vodu* ‘Living creatures, including even plants, drink water’. [↑](#footnote-ref-9)
10. Note that this construction ID 171 later received a more general formula **NP/VP vrode NP-Gen**, as explained further in this section and represented in Figure 5. [↑](#footnote-ref-10)
11. We use here the standard terminology well-known and broadly used in typological studies of comparative constructions (Treis, 2018). [↑](#footnote-ref-11)
12. In ID 1972, *tju-tju* and *tjutju* are spelling variants of the same anchor word. [↑](#footnote-ref-12)
13. Comparable results for the Minimizing construction in Dutch are reported by Van den Heede & Lauwers (2021). [↑](#footnote-ref-13)
14. Most minimizers can also be used with diminutive suffixes and/or or with the numeral *odin* ‘one’. [↑](#footnote-ref-14)
15. The searches were conducted in Araneum Russicum III Maximum, the numbers in Table 2 are provided according to the date of access 8.09.2021. The query was formulated for specific nouns in the following manner: "ни""капли""не"[tag="V.\*"] (meaning that the combination of words "*ni kapli ne"* is followed by any verb form) for the word order "ni NP-Gen ne V", and as "не"[tag="V.\*"]"ни""капли"[tag!="N.\*"&tag!="A.\*"] (meaning any negated verb form followed by *ni kapli* without any noun or adjectival dependent elements) for the word order "ne V ni NP-Gen". [↑](#footnote-ref-15)