Constructional Approaches
to Language 22

Constructicography

Constructicon development across languages

edited by
Benjamin Lyngfelt
Lars Borin
Kyoko Ohara
Tiago Timponi Torrent

John Benjamins Publishing Company

Constructicography

Constructional Approaches to Language

ISSN 1573-594X

The series brings together research conducted within different constructional models and makes them available to scholars and students working in this and other related fields.

The topics range from descriptions of grammatical phenomena in different languages to theoretical issues concerning language acquisition, language change, and language use. The foundation of constructional research is provided by the model known as Construction Grammar (including Frame Semantics). The book series publishes studies in which this model is developed in new directions and extended through alternative approaches. Such approaches include cognitive linguistics, conceptual semantics, interaction and discourse, as well as typologically motivated alternatives, with implications both for constructional theories and for their applications in related fields such as communication studies, computational linguistics, AI, neurology, psychology, sociology, and anthropology.

This peer reviewed series is committed to innovative research and will include monographs, thematic collections of articles, and introductory textbooks.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/cal

Volume 22

Constructicography. Constructicon development across languages
Edited by Benjamin Lyngfelt, Lars Borin, Kyoko Ohara and Tiago Timponi Torrent

Constructicography

Constructicon development across languages

Edited by

Benjamin Lyngfelt
University of Gothenburg

Lars Borin
University of Gothenburg

Kyoko Ohara
Keio University

Tiago Timponi Torrent
Federal University of Juiz de Fora

John Benjamins Publishing Company
Amsterdam / Philadelphia

Funders: Riksbankens jubileumsfond (The Swedish Foundation for Humanities and Social Sciences)
Grant number P12-0076:1

DOI 10.1075/cal.22

Cataloging-in-Publication Data available from Library of Congress:

LCCN 2018014202 (PRINT) / 2018040536 (E-BOOK)

ISBN 978 90 272 0100 3 (HB)
ISBN 978 90 272 6386 5 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

CHAPTER 6

A constructicon for Russian

Filling in the gaps

Laura A. Janda[i], Olga Lyashevskaya[ii], Tore Nesset[i], Ekaterina Rakhilina[ii,iii] and Francis M. Tyers[ii]
[i]UiT The Arctic University of Norway / [ii]National Research University Higher School of Economics / [iii]Vinogradov Institute for Russian language

The Russian Constructicon project currently prioritizes multi-word constructions that are not represented in dictionaries and that are especially useful for learners of Russian. The immediate goal is to identify constructions and determine the semantic constraints on their slots. The Russian Constructicon is being built in parallel with the Swedish Constructicon and will ultimately model the entire Russian language in terms of constructions at all levels from morpheme to discourse. The contents of the Russian Constructicon will serve learners of the language, linguists researching both language-internal and typological phenomena, and will also serve language technology applications such as spell checkers and automated readability assessment tools.

Keywords: Russian, constructions, multi-word, semantic constraints, learners, language technology, typology

1. Introduction

The Russian Constructicon project has emerged organically from a milieu with sustained focus on a range of relevant theoretical and practical aims, including: construction grammar, lexical semantics, quantitative analysis of language data, and development of pedagogical materials for learners of Russian as well as language technology resources for users of Russian. While each of these ideas and undertakings approach Russian from a unique perspective, they all converge on a single challenge, namely the lack of an extensive inventory of Russian constructions. Certain kinds of constructions are represented in dictionaries and other reference works, but many types of constructions are not. This is not due to any shortcomings in such reference works, but due to the fact that their mission is fundamentally different:

DOI 10.1075/cal.22.06jan

they are not designed to deliver a full-scale inventory of the constructions that are useful for second-language learners. For example, a dictionary will not predict constructions of the type: (1) `X tak i ne Vpast` [X.NOM thus and not V.PST] 'But X didn't V after all', as in *On tak i ne ženilsja* [He.NOM thus and not marry.PST.M.SG] 'But he didn't get married after all'; and (2) *Raz i ty …* [Once and you …] '(And) before you know it you're …', as in *Raz i ty v belom plat'e* [Once and you in white.LOC.SG.N dress.LOC.SG] 'And before you know it you're wearing a white dress'.

In its present stage, the Russian Constructicon project is focused on those "missing" constructions, particularly the constructions that are most essential for learners of Russian. A prototypical characteristic of the constructions that we target is the presence of one or more "slots" (see Section 3) that can be filled with a range of words depending on the semantic restrictions of the given construction.

A concise history tracing the relationship of the Russian Constructicon project to the research agendas of its partners appears in Section 2. Section 3 details both the types of constructions that have been previously documented and those that have not, and then presents examples of the types of constructions that the project is currently collecting. Descriptions of annotation and interpretation techniques are provided in Section 4. In Sections 5 and 6 we project the benefits of the Russian Constructicon both in terms of further research it will facilitate and user applications (language technology resources) that can be built on or enhanced by the Constructicon.

2. History and partners

In Russia Construction Grammar has a long history. An important early contribution is the Meaning Text Theory proposed by Mel'čuk, Apresjan and Žolkovskij (Žolkovskij & Mel'čuk, 1965), whose lexical functions became one of the basic concepts in the Moscow Semantic School Approach. This included the analysis of constructions with so-called "light verbs", intensifiers, and lexicalized evaluative expressions. Additionally, Apresjan (1967) explored verbal government as morpho-syntactic relationships motivated by semantics. Švedova (1960) developed the idea of syntactic schemas associated with words and turned attention to patterns that are significant in spoken Russian, beginning with the use of reduplication. Another important theoretical contribution was made by Zolotova in her *Syntactic Dictionary* (2006) which gave an inventory of minimal units of Russian syntax. However, all of these works aimed at constructions at a rather abstract and generic level, focusing on the basic syntax of the language. More recently, Rakhilina's group has focused on these data in the theoretical context of Construction Grammar, analyzed a number of core and non-core Russian constructions, and showed how the constructions

are organized (Rakhilina, 2010). Saj (2008, 2014; Ovsjannikova & Saj, 2014) has led research on the syntactic periphery of Russian, on the interrelationships between lexical items and slots, as well as constructions that do not conform to core syntax. Rakhilina and Letučij (2012; Letučij & Rakhilina, 2014) have focused on what they call "quasigrammatical" constructions and the ways in which they relate to various semantic fields such as time, iterativity, and quantification. Kuznetsova (2015) and Janda and Solovyev (2009) have taken a quantitative approach to the study of Russian constructions to determine the relationships between lexical items and slots.

Despite Russian's vast size (it ranks sixth in the world in terms of total number of speakers, eighth in terms of L1 speakers; https://www.ethnologue.com/statistics/size), the Russian language lags far behind English in terms of electronic resource development. A Russian Constructicon is an important component in addressing this need. Our project approaches the building of a Russian Constructicon from the complementary perspectives of native and non-native language users, and we achieve this through collaboration between Russian and foreign researchers.

The main partners in the project are linguists at the Higher School of Economics (HSE) in Moscow (https://ling.hse.ru/en/) and their counterparts at The Arctic University of Norway (UiT) in Tromsø, namely those in the CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) research group (https://uit.no/forskning/forskningsgrupper/gruppe?p_document_id=344365), and SweCcn – a Swedish Constructicon research group at the University of Gothenburg (https://spraakbanken.gu.se/eng/sweccn).

The partners at HSE and UiT share three core features in their linguistic agendas: (1) the theoretical framework of cognitive linguistics, (2) focus on construction grammar, and (3) statistical analysis of linguistic data. All three of these features directly support the development of a Russian Constructicon as a natural outgrowth of established research traditions.

The Russian Meaning Text Theory and other semantic theories that have emerged in Russia are highly compatible with cognitive linguistics (Rakhilina, 2000, pp. 342–378). Linguists of Russian at UiT share the theoretical commitment to cognitive linguistics, and within that theoretical framework, both groups of linguists have consistently focused on construction grammar. Both groups of linguists have also applied quantitative methods to the study of Russian linguistics, and in both cases this has spilled over into computational approaches and applications (Janda, 2013; Lyashevskaya, 2016). The Russian partners play leading roles in the continuing development of the Russian National Corpus (the foremost linguistic database of Russian, http://ruscorpora.ru/, released in 2002 and under continuous expansion and development). HSE is a world leader in the development of open-source electronic resources for Russian, such as learner corpora, a corpus

of heritage Russian, corpora of dialectal and regional Russian, a tutorial in academic writing, and the semantic edition of Tolstoy's collected works (a digital humanities project). UiT has developed UDAR (Reynolds, 2016), the only full-scale open-source finite-state transducer morphological computational model of Russian that takes into account stress (the placement of accents of words, which can convey meaning differences, as in *dóma* [at.home] 'at home' and *dóm-a* [house-GEN.SG] 'house' vs. *dom-á* [house-NOM.PL] 'houses'), and is also engaged in the development of interactive and web-enhancement resources for learners of Russian.

In a very concrete sense, the Russian Constructicon project has evolved from traditional reference and electronic corpus resources. There are, of course, a myriad of dictionaries of Russian, but among these Zaliznjak 1980 stands out as a landmark work that for the first time detailed the morphological forms for all inflected words in Russian. Among major world languages, Russian is relatively morphologically complex, with large inflectional paradigms for nouns, adjectives, and verbs, features usually associated with minority languages (McWhorter, 2011; Trudgill, 2011). Zaliznjak set the standard for interpreting and modeling the morphological complexity of Russian, an essential component for most language technology resources. Zaliznjak 1980 is also a cornerstone of the Russian National Corpus (as well as other corpora of Russian), since the morphological analysis it uses, both in the portion of the corpus that is automatically tagged and in the portion that has been manually tagged, is based directly on Zaliznjak's model of Russian morphology.

Morphological analysis of Russian has made it possible to search for lexemes in a corpus (since all forms of a word can be associated with the appropriate lexeme), and this in turn has facilitated the creation of corpus-based dictionaries such as the frequency dictionary by Lyashevskaya and Sharoff (2009).

Corpus-based research set the stage for the systematic study of linguistic constructions, since it raised the issue of how to track and interpret units larger than single words. A number of resources have been developed to address this need, all of which give a firm basis for the building of a Russian Constructicon. One outcome of this line of work is the *Corpus Dictionary of Multi-Word Lexical Units* (2008, http://ruscorpora.ru/obgrams.html), composed of data on frequent collocations in the Russian National Corpus, with supplementary material from Rogožnikova's (2003) dictionary of collocations and the four-volume academy dictionary of Russian (Evgen'eva, 1999). The *Corpus Dictionary of Multi-Word Lexical Units* lists over 2900 such collocations, along with their frequency (in the Russian National Corpus as of 2008), and links to corpus examples. This inventory is broken down into five groups according to syntactic-semantic functions: (1) multi-word units functioning as prepositions like *vo imja X* [in name.ACC X.GEN] 'for the sake of X'; (2) adverbial and predicational multi-word units like *na vsjak-ij slučaj* [on any-ACC.

SG event.ACC.SG] 'just in case'; (3) parenthetic multi-word units like *s točk-i zreni-ja X* [from point-GEN.SG view-GEN.SG X.GEN] 'from X's point of view'; (4) multi-word units that function as conjunctions like *dlja togo čtob(y)* [for that.GEN so-that] 'in order that'; and (5) multi-word units that function as particles like *ne inače* [not otherwise] 'certainly'.

Another outcome of corpus investigations into units larger than the lexeme was the Russian FrameBank (http://framebank.ru/; Lyashevskaya & Kashkin, 2015). Analogous to FrameNet for English (Fillmore et al., 2008), the Russian FrameBank draws on Russian lexicographical traditions and traditional printed dictionaries (Apresjan & Pall, 1982; Sazonova, 2008). The result is a hybrid resource that integrates dictionary-style information about verbal government (e.g., valency, syntactic frames) with linguistic interpretation of corpus data. The Russian FrameBank is centered on 2700 high-frequency verbs in Russian and the constructions that they appear in, both in corpus data (100 corpus examples for each verb are fully parsed both semantically and syntactically and classified according to construction type), and according to dictionaries (which may list constructions in addition to those found in the 100 corpus examples). For example, for the verb *vzjat'* 'take', there are three examples among the 100 corpus examples of the `S.NOM V S.ACC v + S.ACC` construction, as in *Ja sam vzja-l v ruk-i mokr-uju xolodn-uju butylk-u* [I.NOM self.NOM.SG.M take-PST.M.SG in hand-ACC.PL wet-ACC.SG.F cold-ACC.SG.F bottle-ACC.SG] 'I myself took the cold wet bottle in my hands'. However, there are no corpus examples in the FrameBank sample of the type `S.NOM V S.ACC S.INS` like *On vzja-l ščit lev-oj ruk-oj* [He.NOM take-PST.M.SG shield.ACC.SG left-INS.SG.F hand-INS.SG] 'He took the shield with his left hand', although this type is attested in dictionaries and can be located in more extensive corpus searches. At the present time this research is also being extended to constructions associated with adjectives, yielding findings concerning constraints such as the limitation to predicative use for the `Adj` *na* + ACC construction in for example *On sposoben na podlost'* [He.NOM capable.M.SG.SHORT-FORM on meanness.ACC.SG] 'he is capable of meanness', and that superlative forms can have different argument structure properties than their neutral equivalents, as in *lučš-ij/sam-yj xoroš-ij v mir-e* [best-NOM.SG.M in world-LOC.SG] 'best in the world', cf. the unattested **xorošij v mire* 'good in the world'.

The FrameBank hybrid between a linguistic reference work and a portal for corpus examples is the future of the dictionary as envisioned by Atkins (1992) and Kilgarriff et al. (2006), and also leads us in the direction of a dictionary of constructions, or a constructicon. However, as we detail in the following section, there remain gaps in our coverage of the syntactic-semantic peculiarities of Russian, and the Russian Constructicon project is designed to fill those gaps.

3. Russian constructions: What's missing

Our starting point is construction grammar as outlined by Langacker (1987, 1991a–b, 2003), Croft (2001), Goldberg (1995, 2006), and Fillmore (1985; Kay & Fillmore, 1999). Although these scholars take slightly different perspectives on constructions, they all share a similar view on what constitutes a construction, namely any conventionalized pairing of form and meaning in language, at any level, from the level of the morpheme, through words and phrases, and up to the level of discourse. The meaning of each construction is emergent (Langacker, 1991b, pp. 5–6, 534), motivated by the patterns of uses of the units that appear in the construction, and also by the larger (clause- or discourse-level) constructions that a given construction appears in. Since a language is a network of interrelated constructions, a constructicon is a model of an entire language. While our ultimate goal is to create a full-scale constructicon, at present we have made strategic decisions to prioritize the types of constructions that should be collected.

The Russian Constructicon project specifically addresses the resources that are lacking for both research and pedagogy with respect to Russian constructions. Some items along the scale from morphemes to discourse are relatively well described. For example, at the word level we have traditional dictionaries, and the argument structures associated with lexemes are detailed in those sources and in the Russian FrameBank. There also exist phraseological dictionaries (Mixel'son, 1896–1912/2004; Lubensky, 1995; Bystrova, 1997; Kuz'mič, 2000; Fedosov, 2003), but these have a strong bias toward very specific types of phrases such as sayings, aphorisms, and proverbs where the relationship between the components and the semantics of the whole are particularly obscure. Take for example the phrase *kak siv-yj merin* [like gray-NOM.SG.M gelded.horse.NOM.SG], which literally means 'like a gray gelded horse', but actually describes a particularly dishonest manner of behavior (usually in relation to telling lies), roughly equivalent to 'through one's teeth' in English collocations like *He's lying through his teeth*. These are the types of entries one finds in phraseological dictionaries, but phrases like this tend to be of very low frequency – *kak sivyj merin* appears only twenty-seven times in the entire Russian National Corpus, approximately once in ten million words. Such phrases are idiosyncratic, thus rarely yielding general patterns that would be of interest to theoretical linguists, and so infrequent as to be of little use to learners of Russian as a second language. Some of these resources, while they have their merits, are themselves very skewed. For instance, Baranov et al.'s (2009) dictionary-thesaurus of Russian idioms was largely compiled from detective stories and thus overrepresents phrases used in Russian taboo expressions and swearing known as *mat*, a much stronger genre than its English correlate and decidedly inappropriate for use by second language learners in most contexts. What learners really need are

phrases that phraseological dictionaries overlook, such as *davaj ruku* [give.IMP.SG hand.ACC.SG] 'give me your hand (so that we can shake hands or so that you can help me get up)' and *skol'ko možno!* [how-much possible] (lit. 'how much is possible (for someone to X)') 'oh, for crying out loud, give it a rest already!' (used to express exasperation at excessive talk about something).

Because of its inflectional morphology, pedagogical materials for learners of Russian invest heavily in teaching paradigms and grammatical endings. This is well justified since in a very real sense, one cannot even begin to speak Russian without mastering a large portion of the grammatical inflections. It is fairly easy to succeed at speaking "bad" English for example, by merely stringing together lexemes in a largely predictable order – the result will not be idiomatic, but you can get your message across. However, speaking good English is very hard. The big hurdle for learners of English comes along when they try to master the constructions, making the need for a constructicon very obvious. For learners of Russian, it is difficult even to speak badly and be understood since all words in a sentence (except for prepositions and conjunctions and a few "particles", see Endresen et al., 2016) have to be inflected. Russian grammatical morphology has to be acquired and routinized to a high degree right from the beginning. This is a huge task and for this reason the American Council of Teachers of Foreign Languages (http://www.languagetesting.com/how-long-does-it-take) ranks Russian among the "Group III Languages" in their four-point scale (with respect to difficulty for learners whose native language is English). Group I Languages (Afrikaans, Danish, Dutch, French, etc.) are learned easily and quickly, followed by Group II Languages (Bulgarian, Dari, Farsi, German, etc.). Group IV Languages (Arabic, Chinese, Japanese, Korean) are the hardest to acquire, and note that none of those are Indo-European. Russian is the Group III Language with the largest number of speakers and the only major world language in that group (other languages in Group III are: Amharic, Bengali, Burmese, Czech, Finnish, Hebrew, Hungarian, Khmer, Lao, Nepali, Filipino, Polish, Serbo-Croatian, Sinhala, Thai, Tamil, Turkish, and Vietnamese). While Russian textbooks place considerable emphasis on inflectional morphology, they tend to have at best only sporadic coverage of constructions.

However the need to master the morphology doesn't mean that speaking good Russian is any less dependent on knowing constructions. Indeed, the morphology serves an essential role in Russian constructions. But the constructional landscape of Russian is also highly complex, and we are only beginning to explore that landscape. Especially with respect to the needs of learners, Russian constructions are woefully underdescribed.

Our Russian Constructicon project currently prioritizes the constructions that are missing from other resources. This means that we are not concerned with units at the word-level (since those are represented in dictionaries), nor with the

argument structure features of word-level units (since those are represented in dictionaries and FrameBank), nor with sayings and phrases (since those are represented in phraseological dictionaries). Our focus is on the multi-word units that are not represented in other resources and particularly on those that are most useful to learners of Russian.

We use a team strategy in our approach to constructions. Native speakers are typically blind with regard to the constructions that are challenging to learners, since to them all constructions are equally comprehensible. Non-Russian team members are needed to identify the constructions that stand between learners and Russian proficiency. However, only native Russian team members have the capacity to fully interpret and annotate constructions.

The multi-word units that we target present a range of types that vary according to the presence vs. absence both of "slots" (underdetermined portions of constructions) and of constraints on those slots. On one end of the scale are fixed expressions where all the components are obligatory and unchangeable, such as *Kto tam?* [who.NOM there] 'Who's there?' (a response to a knock at the door) or *Vot ešče!* [Look still] 'No way!'. Baranov et al. (2009) call these "situational clichés", and they can be thought of as degenerate constructions. Although technically they have no slots, they still have variables, since there often has to be something that precedes or follows them, such as the knock at the door before *Kto tam?* 'Who's there?' or the specification of what is being rejected in *Vot ešče! Ja posud-u my-t' ne bud-u.* [Look still I.NOM dishes-ACC.SG wash-INF not be.FUT-1SG] 'No way! I'm not going to wash the dishes.' There are also some constructions that approach this degenerate type because they have severe restrictions on their slots, as in *Èx ty!* [Oh you.NOM] 'Shame on you! Darn!' (said when something doesn't work out). For some speakers this is a fixed expression allowing only the second person singular (intimate) or plural pronouns *ty* and *vy* 'you', while others can also admit (usually reduplicated) names for people (*Èx Vitja, Vitja*), but other fillers are excluded.

On the other end of the scale are syntactic constructions that have almost no constraints on their slots. This type is called a "schema" by Švedova (Švedova et al., 1980) and Belošapkova (1977), and can be realized as both a simple sentence like *Kakoj* X! [What.NOM.SG.M X.NOM.SG] 'What a X!' (where the adjective *kakoj* 'what kind of' needs to have the correct inflectional ending to agree with the number and gender of whatever noun goes in the slot), or as a complex sentence like *Esli* Y, *togda* Z [If Y, then Z] 'If Y, then Z'. At present we lack inventories of both the slotless degenerate type of constructions and those that are maximally open.

However, the most interesting constructions are those that lie between these two extremes, namely those with various types of restrictions on their slots, and these can include both constructions that constitute entire sentences and those that are phrases. The tendency here is that when there are lexical constants in a

construction (items that are fixed, cf. Fillmore et al., 1988), there are also greater semantic constraints on the slots. For example, the construction *Kak u vas s* X? [How by you.GEN with X.INS] 'What's your X situation like?' has a semantic restriction on the range of words that can go in the X slot, which most often refer to essential challenges for human beings like *pitaniem* 'food', *zdorov'em* 'health', *den'gami* 'money', *pogodoj* 'weather', *nasledstvennost'ju* 'inheritance'. It is hard to come up with contexts that would support the use of other kinds of fillers, such as features of nature like *nebom* 'sky' and deverbals like *prixodom* 'arrival' in this slot. A phrase-level example is *Let* Y [year.GEN.PL Y.GEN] 'about Y years old, in his/her Y-ies', as in *let semidesjati* [year.GEN.PL seventy.GEN] 'about seventy years old, in his/her seventies'. In addition, constructions can have multiple slots, all of which have semantic restrictions and some of which can be optional. For example, there is the X *v* Y [X in Y.ACC] construction used to describe patterns on clothing, as in *jubk-a v kletočk-u* [skirt-NOM.SG in check-ACC.SG] 'a checkered skirt'. The X slot is filled with a noun that refers to an article of clothing (which may be singular or plural), and the case marking depends on the role of that noun in the larger sentence. The Y slot refers to a type of pattern, usually additionally marked with a diminutive suffix containing *k*, such as *poloska* 'stripe', *gorošek* 'polka dot', *cvetoček* 'flower'. An option is to also specify the color(s) of the pattern by inserting one or more adjectives designating colors before Y.

Our current task in building the Russian Constructicon is to study such syntactic fragments and their interpretations, to map out the range of Russian constructions, and work out the semantic restrictions on their slots.

4. Status of the project and examples from the Russian constructicon

Textbooks of Russian and texts that represent or approximate spoken language (for example children's stories and films as well as the prose of certain writers such as Sergej D. Dovlatov) are good sources for the type of constructions we are targeting. At present over six hundred constructions have been entered in *A Constructicon for Russian* at https://spraakbanken.gu.se/karp/#?mode=konstruktikon-rus. This site uses the same architecture as the Swedish Constructicon and thus preserves all the search and other features of that constructicon and is designed to be comparable across languages.

A full entry in the Russian Constructicon can include up to five elements: NAME, DEFINITION, STRUCTURE, EXAMPLES, and COMMENT, as shown in Figure 1.

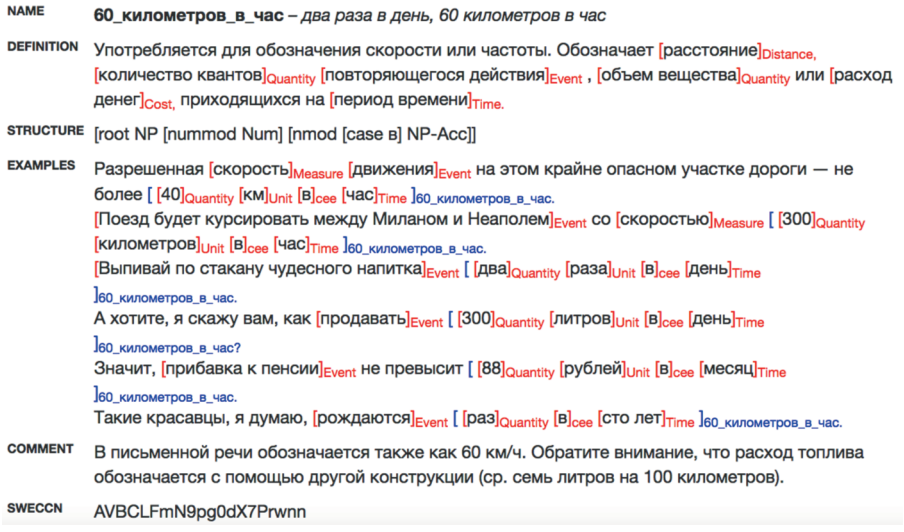

Figure 1. Example entry in the Russian constructicon

The NAME of a construction may either represent an example or be more schematic, depending on the construction. In Figure 1, the NAME is `60_kilometrov_v_čas` 'sixty kilometers an/per hour', and a second example is also supplied: *dva raz-a v den'* [two.NOM time-GEN.SG in day.ACC.SG] 'two times a/per day'. The NAME is used in the EXAMPLES section, appearing in blue with square brackets to show where the construction begins and ends.

The DEFINITION of the construction describes its semantics, with tags for the elements. In our example in Figure 1, the DEFINITION (translated from the Russian) is: Used to designate speed or frequency. Designates the [distance]$_{\text{Distance}}$, [number of units]$_{\text{Quantity}}$ of a [repeated action]$_{\text{Event}}$, the [volume of a substance]$_{\text{Quantity}}$ or the [expenditure of money]$_{\text{Cost}}$, that occurs over a [period of time]$_{\text{Time}}$. The tags appear in red and are used also in the EXAMPLES with square brackets so that is it easy to keep track of correspondences. The tags and the definition aim to capture the semantic restrictions on the slots of the construction. Thus, for example, we see that there is an Event (named outside the construction) that involves a Quantity, usually expressed with a numeral and a Unit in relation to a period of Time.

The STRUCTURE of this construction is rendered in Universal Dependency Grammar as: [root NP [nummod Num] [nmod [case v] NP-Acc]. This means that there are two noun phrases and the preposition *v* 'in' in the construction. The first noun phrase can contain a numeral and a noun phrase quantified by that numeral. The second noun phrase is governed by the preposition in the accusative case.

The EXAMPLES for this construction are as follows:

(1) *Razrešenn-aja skorost' dviženij-a na èt-om*
allowed-NOM.SG.F speed.NOM.SG movement-GEN.SG on that-LOC.SG.M
krajne opasn-om učastk-e dorog-i ne bolee
extremely dangerous-LOC.SG.M portion-LOC.SG road-GEN.SG not more
40 km v čas
40.GEN km.GEN.PL in hour.ACC.SG
'The speed limit for that extremely dangerous part of the road is not more than 40 km per hour.'

(2) *Poezd bud-et kursirova-t' meždu Milan-om i*
train.NOM.SG be.FUT-3SG shuttle-INF between Milan-INS.SG and
Neapol-em so skorost'ju 300 kilometr-ov v čas
Naples-INS.SG with speed-INS.SG 300 kilometer-GEN.PL in hour.NOM.SG
'The train will shuttle between Milan and Naples at a speed of 300 kilometers per hour.'

(3) *Vypivaj po stakan-u čudesn-ogo napitk-a*
drink.IMP.SG along glass-DAT.SG marvelous-GEN.SG.M beverage-GEN.SG
dva raz-a v den'
two.ACC time-GEN.SG in day.ACC.SG
'Drink this marvelous beverage twice a day.'

(4) *A xoti-te, ja skaž-u vam, kak prodava-t'*
and want-PRS.2PL I.NOM tell-PRS.1SG you.DAT how sell-INF
300 litr-ov v den'
300.ACC liter.GEN.PL in day.ACC.SG
'And if you want, I will tell you how to sell 300 liters a day.'

(5) *Znači-t, pribavka k pensi-i ne*
mean-PRES.3SG supplement.NOM.SG to pension-DAT.SG not
prevysi-t 88 rublej v mesjac
exceed-PRS.3SG 88.ACC ruble.GEN.PL in month.ACC.SG
'In other words, the pension supplement will not exceed 88 rubles per month.'

(6) *Tak-ie krasavc-i, ja dumaj-u,*
Such-NOM.PL.M handsome.man-NOM.PL I.NOM think-PRS.1SG
roždaj-ut-sja raz v sto let
be.born-PRS.1SG-REFL time.NOM.SG in hundred.ACC year.GEN.PL
'Such handsome men, I think, are born once in a hundred years.'

The COMMENT for this entry is: "In writing it also appears as 60 km/h. Note that the use of fuel is designated by a different construction (compare seven liters in 100 kilometers)."

Although at present we are focusing on multi-word units, our aim is to model the entire Russian language in terms of constructions. To this end, existing resources (dictionaries, argument structure information from FrameBank) will be integrated into the Russian Constructicon, and the scope of the project will be extended to include both units smaller than a word (morphemes, derivational morphology) and concatenation of constructions into larger discourse units. For now, the entries are given in Russian, though in the future users will be able to get the definitions and comments also in other languages, such as English. In keeping with the pedagogical aims of the Russian Constructicon, materials and resources for learners will also be developed and integrated into this project.

5. Further research facilitated by the Russian constructicon

Of course a large research investment will be made in the definition of Russian constructions and the semantic restrictions on their slots, and corpus linguistic techniques will play an important role in that research. However, the Russian Constructicon itself will also serve as a research tool. There are many directions which that research may take. For example, to date there has been very little research on typological comparisons of constructions across languages. Rakhilina and colleagues (Rakhilina & Majsak, 2007; Rakhilina et al., 2012; Rakhilina & Plungian, 2013) have pioneered typological work on the lexical semantics of certain domains (aquamotion, pain, speed), but such typological comparisons could be extended both in terms of the syntax and the semantics of constructions. Following this lead, it would be possible to take an onomasiological approach, starting from general types of meanings such as negation and indefiniteness, and examine how these meanings are expressed by constructions.

Within Russian, various kinds of classifications of constructions will reveal systematic grammatical patterns and also facilitate research as well as access to examples through our interface. For example, both syntactic and semantic classifications can be developed. This will make it possible to discover the relationships among constructions in what could be called "construction families" similar to the family of `Subject-Auxiliary Inversion` constructions in English (Goldberg, 2006, Chapter 8). Some preliminary work on paradigmatic relations among constructions has been attempted (Janda & Divjak, 2008), but only at a very schematic level (specifying the grammatical case of the arguments of verbs). Constructions can be grouped paradigmatically according to the part of speech that serves as their core (nouns, verbs, etc.). To our knowledge, no systematic study of the syntagmatic co-occurrence patterns of Russian constructions has been attempted, and this is

a very complex dimension, since constructions can be nested within each other, overlap, or be contiguous, even across sentence boundaries.

Another line of research that will benefit both description and pedagogy is the behavior of grammatical categories in constructions. Many constructions have slots for verbs, and in Russian all verbs express either perfective or imperfective aspect, referring to the way in which an event is understood, roughly as either a complete whole or as an unbounded situation. The category of aspect in Russian is among the most challenging grammatical concepts for learners of Russian. It is extremely difficult both for linguists and for language teachers to explain when to select a perfective or an imperfective verb. Textbooks devote considerable space to "rules" for using aspect, but nearly all such rules admit exceptions. These rules present various "triggers" for use of aspect, such as: "use perfective aspect in the presence of *uže* 'already'", or "use imperfective in the presence of *vsegda* 'always'". However, the triggers for such rules are actually fairly rare in authentic texts: Reynolds (2016) finds that these triggers co-occur with only about 2% of verbs in a corpus. While the triggers are good indicators of aspect (yielding 98% correct guesses according to rules), they aren't plentiful enough to be useful. In other words, by focusing on a small number of coarse-grained triggers, we are failing both as linguists to fully describe the phenomenon and as instructors to give our students adequate guidance. The Russian Constructicon will make it possible to investigate the parameters of less clear-cut cases. For example, a search in the Russian National Corpus reveals that 75% of the verbs that appear directly after *čtoby* 'in order to' are perfective. While 75% is a strong trend, it is not very reliable. We need more detail on exactly what kinds of constructions and which verbs influence the choice of aspect. The Russian Constructicon already shows promise, by identifying constructions where imperfective verbs are preferred after *čtoby* 'in order that', such as `X sozdan, čtoby` [X.NOM.SG created.NOM.SG.M in.order.that] 'X was made in order to' as in *avtomobil' sozdan, čtoby na nem ezdi-t'* [automobile.NOM.SG created.NOM.SG.M in.order.that on it.LOC ride-INF] 'the automobile was made to be ridden' and *sliškom* Y, *čtoby* [too Y in.order.that] 'too Y to' as in *ja sliškom ustal, čtoby sraža-t'-sja* [I.NOM too tired.M.SG in.order.that fight-INF-REFL] 'I'm too tired to fight'). Details like these can be used to calibrate more precise rules. And this kind of research can be extended to other grammatical categories and parts of speech. In this way, the Russian Constructicon provides added value for both researchers and learners.

6. Applications served by the Russian constructicon

In addition to serving linguistic research and pedagogical needs, the Russian Constructicon has important implications for the development of language technology applications for Russian. Many types of constructions present challenges for computational processing even in morphologically tagged corpora; this is especially true for multi-word units that can be discontinuous and contain variable slots. A full-scale inventory of Russian constructions can improve the standard resources of language technology such as spell checkers and machine translation. The density and complexity of constructions are one indicator of the readability of texts that has until now remained beyond the reach of language technology (see the overview in Vajjala, 2015). Comparison of the constructions present in texts rated for readability can serve as training material for machine learning that will make it possible to automatically and accurately gauge appropriate reading materials for both native Russian schoolchildren and second language learners. Interactive learning and web enhancement tools (Meurers et al., 2010) can also be designed to focus on the task of mastering Russian constructions.

Acknowledgments

This article is an outcome of a research project funded by the Norwegian Centre for International Cooperation in Education (CPRU-2017/10027) and by UiT, The Arctic University of Norway, and of a research project implemented as part of the project № 16-18-02071 from the Russian Science Foundation and the project RFBR- 17-04-00517a from the Russian Foundation for Basic Research.

References

Apresjan, Ju. D. (1967). *Èksperimental'noe issledovanie semantiki russkogo glagola* [Experimental research on the semantics of the Russian verb]. Moscow: Nauka.

Apresjan, Ju. D., & Pall, È. (1982). *Russkij glagol – vengerskij glagol* [Russian verb – Hungarian verb]. Budapest: Tankönyvkviadó.

Atkins, S. (1992). Tools for computer-aided corpus lexicography: the Hector project. In F. Kiefer, G. Kiss & J. Pajsz (Eds.). *Papers in Computational Lexicography: Complex'92*, Hungarian Academy of Sciences, Budapest (pp. 1–60). Budapest: Akadémiai Kiadó.

Baranov, A. N., Voznesenskaja, M. M., Dobrovol'skij, D. O., Kiseleva, K. L., & Kozerenko, A. D. (2009). *Frazeologičeskij ob"jasnitel'nyj slovar' russkogo jazyka* (Phraseological explanatory dictionary). Moscow: ÈKSMO.

Belošapkova, V. A. (1977). *Sovremennyj russkij jazyk. Sintaksis* [Contemporary Russian. Syntax]. Moscow: Vysšaja škola.

Bystrova, E. A. (1997). *Učebnyj frazeologičeskij slovar'* [Academic phraseological dictionary]. Leningrad: Prosveščenie.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Endresen, A., Janda, L. A., Reynolds, R., & Tyers, F. M. (2016). Who needs particles? A challenge to the classification of particles as a part of speech in Russian. *Russian Linguistics.* doi:10.1007/s11185-016-9160-2.

Evgen'eva, A. P., Ed. (1999). *Malyj akademičeksij slovar'* [Small academy dictionary]. Moscow: Russkij jazyk.

Fedosov, I. V. (2003). *Frazeologičeskij slovar' russkogo jazyka* [Phraseological dictionary of Russian]. Moscow: JuNVES.

Fillmore, C. (1985). Syntactic intrusions and the notion of grammatical construction. *Proceedings of the Berkeley Linguistics Society* 11, 73–86.

Fillmore, C., Kay, P., & O'Connor, C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *let alone*. *Language* 64, 501–538. doi:10.2307/414531

Fillmore, C. J., Bernal, E. & DeCesaris, J. (2008). Border conflicts: FrameNet meets construction grammar. *Proceedings of the XIII {EURALEX} International Congress*. Barcelona: Universitat Pompeu Fabra. http://www.icsi.berkeley.edu/proceeds/2008/speech/Euralex2008/docum/10001.pdf

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

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

Janda, L. A. (Ed.) (2013). *Cognitive linguistics: The quantitative turn. The essential reader*. Berlin: De Gruyter Mouton. doi:10.1515/9783110335255

Janda, L. A., & Divjak, D. (2008). Ways of attenuating agency in Russian. In A. Siewierska (Ed.), *Impersonal constructions in grammatical theory*, a special issue of Transactions of the Philological Society 106, 138–179.

Janda, L. A., & Solovyev, V. (2009). What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20, 367–393. doi:10.1515/COGL.2009.018

Kay, P., & Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. *Language* 75,1–34. doi:10.2307/417472

Kilgarriff, A., Rundell, M., & Uì Dhonnchadha, E. (2006). Efficient corpus creation for lexicography: Building the new corpus for Ireland. *Language resources and evaluation* 40, 127–152. doi:10.1007/s10579-006-9011-7

Kuz'mič, V. (2000). *Žgučij glagol – slovar' narodnoj frazeologii* [Burning verb – a dictionary of folk phraseology]. Moscow: Zelenyj vek.

Kuznetsova, J. (2015). *Linguistic profiles: Going from form to meaning via statistics* (= Cognitive Linguistics Research 53). Berlin: De Gruyter Mouton. doi:10.1515/9783110361858

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Vol. I. Theoretical prerequisites. Stanford: Stanford University Press.

Langacker, R. W. (1991a). *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter. doi:10.1515/9783110857733

Langacker, R. W. (1991b). *Foundations of cognitive grammar*. Vol. II. Descriptive application. Stanford: Stanford University Press.

Langacker, R. W. (2003). Constructional integration, grammaticization, and serial verb constructions. *Language and Linguistics* 4, 251–278.

Letučij, A. B. & Rakhilina, E. V. (2014). Načal'naja stadija grammatikalizacii značenij glagol'noj množestvennosti v kvazigrammatičeskix konstrukcijax [Early stage of grammaticalization of the meaning of vebal multiplicity in quasi-grammatical constructions]. In A. Moldovan & S. Tolstaja (Eds.), *Slavjanskoe jazykoznanie. XV Meždunarodnyj s"ezd slavistov. Minsk, 21–27 avgusta 2103 g. Doklady rossijskoj delegacii* [Slavic linguistics. XV International Congress of Slavists. Minsk August 21–27, 2013. Contributions by the Russian participants] (pp. 432–451). Moscow: Indrik.

Lubensky, S. (1995). Random House Russian-English Dictionary of Idioms. New York: Random House.

Lyashevskaya, O. N. (2016). *Korpusnye instrumenty v grammatičeskix issledovanijax russkogo jazyka*. Corpus tools for Russian grammar studies. Moscow: LRC Publishing House.

Lyashevskaya, O., & Kashkin, E. (2015). FrameBank: a database of Russian lexical constructions. In M. Yu. Khachay, N. Konstantinova, A. Panchenko, D. I. Ignatov, & G. V. Labunets (Eds.), Analysis of images, social networks and texts. Fourth international conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised selected papers (= Communications in computer and information science, Vol. 542) (pp. 337–348). Springer.

Lyashevskaya O. N., & Sharoff, S. A. (2009). *Častotnyj slovar' sovremennogo russkogo jazyka (na materiale Nacional'nogo Korpusa Russkogo Jazyka)* [Frequency dictionary of modern Russian (based on the Russian National Corpus)]. Moscow, Azbukovnik.

McWhorter, J. H. (2011). *Linguistic simplicity and complexity. Why do languages undress?* Berlin: De Gruyter Mouton. doi:10.1515/9781934078402

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., & Ott, N. (2010). Enhancing authentic web pages for language learners. In *Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-5) at NAACL-HLT 2010* (pp. 10–18). Los Angeles.

Mixel'son, M. I. (1896–1912/2004). *Bol'šoj tolkovo-frazeologičeskij slovar'* [Big explanatory phrazeological dictionary]. Moscow: ETS Publishing House.

Ovsjannikova, M. A. & Saj, S. S. (2014). Raznye formy, obščie mexanizmy: al'ternacija odno- i dvuxob"ektnoj diatez u russkix glagolov mental'noj sfery [Different forms, common mechanisms: the alternation of single- and double-object diathesis among Russian verbs expressing mental events]. In S. S. Saj, M. A. Ovsjannikova & S. A. Oskol'skaja (Eds.), *ACTA LINGUISTICA PETROPOLITANA. Trudy Instituta lingvističeskix issledovanij RAN [Works of the Institute for Linguistic Research of the Russian Academy of Sciences] Russkij jazyk: grammatika konstrukcij i leksiko-semantičeskie podxody [Russian: construction grammar and lexico-semantic approaches]* (pp. 248–291). St. Petersburg: Nauka.

Rakhilina, E. (2000). *Kognitivnyj analiz predmetnyx imen: Semantika i sočetaemost'* [Cognitive analysis of concrete nouns: Semantics and collocation]. Moskva: Russkie slovari.

Rakhilina, E. (Ed.) (2010). *Lingvistika konstrukcij* [The linguistics of constructions]. Moscow: Azbukovnik.

Rakhilina, E. V. & Majsak, T. A. (2007). Glagoly dviženija i naxoždenija v vode: leksičeskie systemy i semantičeskie parametry (Verbs of motion and position in water: lexical systems and semantic parameters). In T. A. Kajsak & E. V. Rakhilina (Eds.), *Glagoly dviženija v vode: leksičeskaja tipologija* [Verbs of motion in water: lexical typology] (pp. 27–75). Moscow: Indrik.

Rakhilina, E. V., Reznikova, T., & Bonch-Osmolovskaya, A. (2012). Towards a typology of pain predicates. *Linguistics* 50, 421–465.

Rakhilina, E. V. & Letučij, A. B. (2012). Russkie konstrukcii s vremennym značeniem: o granicax nastojaščego vremeni [Russian constructions with temporal meaning: on the borders of present tense]. In V. I. Zabotkina (Ed.), *Präsens: Sbornik naučnyx trudov* [Präsens: Anthology of scientific works] (pp. 224–242). Moscow: OLMA Media Grupp.

Rakhilina, E. V. & Plungian, V. (2013). Time and speed: Where do speed adjectives come from? *Russian linguistics* 37, 347–359. doi:10.1007/s11185-013-9117-7

Reynolds, R. J. (2016). *Russian natural language processing for computer-assisted language learning*. Doctoral Dissertation, UiT The Arctic University of Norway.

Rogožnikova, R. P. (2003). *Tolkovyj slovar' sočetanij, èkvivalentnyx slovu* [Explanatory dictionary of collocations that are equivalent to a word]. Moscow: AST, Astrel'.

Saj, S. S. (2008). *K tipologii antipassivnyx konstrukcij: semantika, pragmatika, sintaksis* [Toward a typology of antipassive constructions: semantics, pragmatics, syntax]. PhD dissertation. St. Petersburg: Russian Academy of Sciences, Institute of Linguistic Research.

Saj, S. S. (2014). Leksičeskie mečanismy grammatičeskogo drejfa: konstrukcii s dativnym sub"ektom v russkom jazyke XVIII–XXI vekov [The lexical mechanisms of grammatical drift: constructions with a dative subject in Russian in the XVIII–XXI centuries]. In S. Ju. Dmitrenko & N. M. Zaika (Eds.), *ACTA LINGUISTICA PETROPOLITANA. Trudy Instituta lingvističeskix issledovanij RAN [Works of the Institute for Linguistic Research of the Russian Academy of Sciences] Studia typologica octogenario Victori Khrakovskij Samuelis filio dedicate* (pp. 568–610). St. Petersburg: Nauka.

Sazonova, I. K. (2008). *Tolkovo-grammatičeskij slovar' russkix pričastij* [Explanatory and grammatical dictionary of Russian participles]. Moscow: AST-PRESS.

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

Švedova, N. Ju., Arutjunova, N. D., Bondarko, A. V., Ivanov, V. V., Lopatin, V. V., Uluxanov, I. S., & Filin, F. P. (1980). *Russkaja grammatika. II: Sintaksis* [Russian grammar. II: Syntax]. Moscow: Nauka.

Trudgill. P. (2011). *Sociolinguistic typology. Social determinants of linguistic complexity*. Oxford. Oxford University Press.

Vajjala, S. (2015). *Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications*. PhD thesis, University of Tübingen.

Zaliznjak, A. A. (1980). *Grammatičeskij slovar' russkogo jazyka* [Grammatical dictionary of Russian]. Moscow: Russkij jazyk.

Zolotova, G. A. (2006). *Sintaksičeskij slovar': Repertuar èlementarnyx edinic russkogo sintaksisa* [Syntactic dictionary: The repertoire of elementary units of Russian syntax]. Moscow: Editorial URSS.

Žolkovskij, A. K., & Mel'čuk, I. A. (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza [On a possible method and means for semantic synthesis]. *Naučno-texničeskaja informacija* [Scientific and Technical Information], 5, 23–28.