What goes around, comes around: Cases that keep me going

Laura A. Janda, UiT The Arctic University of Norway

Abstract:
I am fascinated by the treasure trove of meanings tucked away in the grammatical morphemes that many people think of as mere functional fillers. As a student, the Slavic case endings baffled me, then later delighted me with their complex stories about trajectories, time, benefit and harm, labels, and so much more. Some twenty years ago I was satisfied that I had cracked that code, and after writing some articles and a couple of textbooks on the topic, I moved on. But the cases came back to me again and again. In this article, I tell the story of how my work on case semantics later helped to inspire three further projects: two major online resources, the Strategic Mastery of Russian Tool and the Russian Constructicon, and an analysis of president Putin's portrayal of Russia, Ukraine, and NATO. At first glance it might seem that this line of research is rather shallow and merely descriptive, however digging into case semantics reveals some deep philosophical issues concerning the relationship of meaning to grammar, the assumptions inherent in linguistic reference works, the representation of paradigms in the minds of speakers, and the ways in which we can measure grammatical norms and deviation.

Keywords: Slavic languages, case, semantics, cognitive linguistics, construction grammar, political discourse

1. Introduction

Theoretical linguistics, applied linguistics, and language pedagogy are sometimes kept separate from each other, on the assumption that they are very different pursuits. Here I present case semantics as a red thread that has led through a series of projects I have undertaken that link these three disciplines to each other. Section 2 presents the meanings of grammatical cases, using Cognitive Linguistics as the theoretical framework and the Russian case system as the material basis. In Section 3 I turn to the distribution of grammatical case in corpus data and a pedagogical resource created to use this data to provide strategic input for language learners. Case never occurs in a vacuum, always hosted by words and embedded in constructions, but the majority of grammatical constructions cannot be deduced from traditional reference works, a fact that motivated the building of the Russian Constructicon, which is the topic of Section 4. In Section 5 I venture into analysis of political discourse through the distribution and meanings of grammatical case. Conclusions are offered in Section 6.

2. Case Semantics: Linguistic theory and description

After studying Russian for three years in the 1970s, I realized I had a problem. I knew a lot of words, and I could parse just about any sentence, but I was still often stumped about what a given sentence meant. A big part of my problem was the meanings of the Russian (and mutatis mutandis Slavic) cases. As a student I was perplexed by the seemingly random long lists of prepositions and verbs I was assigned to memorize for each case. It was clear to me already then that the grammars I was reading couldn't be telling the whole story. Later, when I had a steady job, I tackled what I assumed were the

hardest cases, the Dative and Instrumental (Janda 1993). Little did I suspect that the Genitive (Janda 1999) and Accusative (Janda 2000) cases would offer plenty of challenges as well. Even the Nominative and Locative were not trivial, and they rounded out the set for two textbooks that I co-authored (Janda and Clancy 2002 and 2006).

2.1 Theoretical issues

On one level I was doggedly picking apart the nitty-gritty details of grammatical case, considered by some linguists to be a syntactic phenomenon devoid of meaning. On another level I was confronting some basic philosophical tenets of linguistics, namely the role of meaning in grammar, and my appreciation for the form-meaning relationship continued to grow. It is common to think of a language as consisting of a lexicon – a set of words that contain the meanings, and a grammar that shows how the words are combined. From the perspective of Cognitive Linguistics, the lexicon and the grammar are not separate entities, but parts of a single system, or as Langacker (2008: 15) describes it: "lexicon, morphology, and syntax form a continuum." In this system, all units have both form and meaning, although the meanings of syntactic expressions tend to be relatively more schematic and polysemous than those of lexemes. More specifically with regard to my research agenda, Langacker (2008: 95) states that the "basic grammatical classes are semantically definable".

While on the face of it the claim that grammatical categories invoke meanings might seem surprising to some, there are several types of evidence that support grammatical meaning: a) typological variation in how functions are expressed, and b) the internal structure of cognitive categories shared across lexicon and grammatical categories.

Many functions are expressed grammatically in some languages, but lexically in others, and often the very same function can be expressed both ways even in the same language and even simultaneously in a single utterance. Here are just a few examples of how synthetic grammar and analytic use of lexemes compete in the same semantic domains. The functions of the grammatical cases we find in Slavic languages can be expressed by means of adpositions in languages without grammatical case. For example, many uses of the Slavic Genitive case can be rendered with the English preposition *of*, as in Russian *načalo fil'ma* 'the beginning of the movie', and many uses of the Slavic Dative case can be rendered as English *to*, as in *dat' graždanam nadeždu* 'give hope to the citizens.' In English *we travel by car*, but in Czech *jedeme autem* using the Instrumental case, sometimes redundantly augmented by a preposition: *jedeme s autem*. And throughout the Slavic languages the meaning of the Locative case is supplemented by prepositions. Some might object that adpositions "don't count" as lexemes because they are merely "function words", but it is not hard to find examples where we need undisputed lexemes to translate the meaning of a grammatical case, as for example Russian *u nego kepka blinom* 'he has a hat that looks like a pancake', where the Instrumental case points to what the hat looks like. This blurring of how functions are expressed synthetically vs. analytically is by no means limited to the meanings of cases. Definiteness expressed by the English article *the* can be expressed by suffixes in Bulgarian, as in *kniga-ta* 'the book', and by either an article or a suffix or even both at once in Norwegian *denne bok-en* 'the/that book'. Verbal categories of tense, aspect, and mood likewise admit both grammatical and lexical expression. It seems that about

half of the languages of the world lack an inflectional future (cf. Dryer and Haspelmath 2013; WALS Feature 67A), and about 40% lack an inflectional past tense (cf. WALS Feature 66A). While in some languages these roles are taken on by auxiliary verbs and one could debate whether auxiliary verbs are mere "function words" or full lexemes, in some languages you need an adverb to express tense, as in North Sámi, where *ihttin* 'tomorrow' or some other temporal expression is needed to specify future. Languages like Slavic that express aspect grammatically are in the minority in the world (cf. WALS Feature 65A); most languages resort to adding in lexemes or whole phrases with meanings like 'finished', 'completely', 'continually', 'was in the habit of' when there is a need to make aspectual meaning clear. Even adjectives are not exempt from such variation, for we find that comparative and superlative meanings can be produced both by affixes and by lexemes; compare synthetic Persian *zibâ-tar* [beautiful-COMPARATIVE], *zibâ-tar-in* [beautiful-COMPARATIVE-SUPERLATIVE] with analytic English equivalents 'more beautiful', 'most beautiful'. Virtually every grammatical category reveals similar examples where the same function can be expressed either synthetically with grammatical morphemes or analytically with lexemes. In other words, there seems to be no clear boundary separating grammatical from lexical meaning in terms of form.

There is likewise no clear boundary between grammar and lexicon in terms of the internal structure of meaning categories. If the main purpose of language is to convey meaning, perhaps it is not surprising that grammar and lexicon jointly participate in this task. And if we cannot definitively distinguish grammatical meaning from lexical meaning, then perhaps the next question, is: how does meaning work? Here I lean upon scholarship reaching back to Eleanor Rosch (1973a and 1973b). Meaning is not "out there" in the world, but is rather a cognitive construct created by human beings based on their perception of realia. Through her research on categorization, Rosch discovered that human beings do not operate in terms of Aristotelian categories defined by sets and boundaries, but by what she termed "radial categories". Radial categories are structured around a central prototype (or cluster of prototypes) with extensions radiating from that prototype. Rosch famously showed that English speakers have a radial category for 'bird': prototypical birds are small, feathered and fly, like robins and sparrows, whereas chickens (with limited flying ability and used as food) are les prototypical, while ostriches and penguins are peripheral. Likewise, apples are a prototypical fruit, while lemons are less so, and avocadoes are quite peripheral. While grammatical meanings are typically more schematic, they can also involve a polysemous radial category structured around a prototype. Janda et al. (2013) present the meanings of the prefixes that signal Perfective aspect in Russian, many of which display an internal radial structure. For example, the prefix *raz-* has a prototypical meaning APART manifest especially when used with motion verbs, as in *razojtis'* 'disperse, walk away in different directions.' This meaning can be extended to apply specifically to the edges of a two-dimensional item, with SPREAD as the meaning in *raskatat'* 'roll out (dough), or a three-dimensional item, with SWELL as the meaning in *razdut'* 'inflate'. Further metaphorical extension yields the meaning EXCITEMENT, as in *razgorjačit'sja* 'heat up, irritate'. The examples presented in Section 2.2 aim to reveal the structures of the case meanings of the Russian cases.

Close examination of case meanings confirms the tenet of Cognitive Linguistics that grammar and lexicon are not distinctly separate, but constitute a continuum, all parts of which contribute to the mission of conveying meaning. Although grammatical meaning may be more abstract and schematic than lexical meaning, meaning at all points along the continuum is a cognitive construct in which prototypical meanings motivate extensions to more peripheral ones.

2.2 Overview of the case meanings, with Russian as an example

Because the details are important to support the theoretical points made above and to motivate the projects described below in Sections 3-5, I will walk through the meanings of all six of the Russian grammatical cases. In the heading introducing each case, I will identify a schematic meaning that summarizes the abstract overall idea expressed by the case and then briefly present a network of between one and four meaning nodes, each cited in small caps, that form the core of the case's meaning. I will point out how the meanings are linked to each other in a relationship of structured polysemy, and I will also give some indications of the further metaphorical and metonymic extensions of these meanings. This is a very condensed version of the contents of this line of research; for a fuller exposition of these meanings, see Janda 1993, 1999, 2000, and Janda and Clancy 2002.

2.2.1 Nominative: Identification

The Nominative case has two central meanings. NOMINATIVE: A NAME can point at an item, be used to call someone, or serve as the grammatical subject. NOMINATIVE: AN IDENTITY is associated with verbs meaning 'be' in formulations meaning 'X is Y' (as in *Ivan **xorošij student*** 'Ivan is **a good student**').

2.2.2 Genitive: Backgrounding with respect to a proximate item

The Genitive case establishes the relationship of a focused entity (a trajector) to something that is backgrounded (a landmark marked with the Genitive). GENITIVE: A SOURCE references a point of departure further specified by prepositions meaning 'from' (*iz*, *s*, *ot*, plus *iz-za* 'from beyond', metaphorically extended to mean 'because of' and *iz-pod* 'from beneath') as well as verbs expressing withdrawal (like *izbegat'* 'avoid', *bojat'sja* 'be afraid of'). This meaning is extended metaphorically to other domains such as time (*s **detstva*** 'since **childhood'**), cause (*smert' ot **razryva serdca*** 'death due to **heart attack**'), and human relationships (*iz **milosti*** 'out of **charity**'). GENITIVE: A GOAL references the opposite maneuver, further specified by prepositions (like *do* 'up to, until', *dlja* and *radi* 'for', *protiv* 'against') and verbs and adjectives expressing (mostly metaphorical) approach (like *ždat'* 'wait for', *želat'* 'wish'). This meaning is extended metaphorically to other domains such as time (*do **svidanija*** 'until **we meet again**') and purpose (*dlja **rešenija*** 'in order **to solve**'). GENITIVE: A WHOLE references the existence of something as a part of a larger unit or collection. This meaning motivates uses of the Genitive case that translate as 'of' and expressions of possession (*ošibka **prezidenta*** '**the president's** mistake') or color (*galstuk **belogo cveta*** 'a **white** tie') in English. This meaning is also associated with complex prepositional phrases (*v kačestve **polnopravnyx učastnikov*** 'in the capacity of **full-fledged participants**') as well as quantification by numerals (*sto **studentov*** 'one hundred **students**'), and in partitive expressions (*vypit' **čaju*** 'drink **some tea**'). GENITIVE: A REFERENCE locates an item with

respect to a landmark in domains of space (like *u* 'by, at'), time (like calendar dates, as in **četvertogo** ijulja '**the fourth** of July'), comparison (*god budet lučše* ***predyduščego*** 'this year will be better than **the previous one**'), and absence (*bez* 'without').

2.2.3 Dative: Interaction

The Dative case encodes the capacity of an entity to interact with its surroundings, by receiving objects, absorbing experiences, or exerting equal or superior strength. DATIVE: A RECEIVER is used primarily to mark the indirect object (*učitel' podaril* ***studentu*** *knigu* 'the teacher gave **the student** a book'), including with verbs of communication (*otvetit'* ***komu-to*** 'answer **someone**') and payment (*zaplatit'* **komu-to** 'pay **someone**'). DATIVE: AN EXPERIENCER is associated with words denoting harm (*mešat'* 'hinder, annoy'), benefit (*služit'* 'serve'), belonging to (*prinadležat'* 'belong to'), and needing (*trebovat'sj*a 'be necessary to'). DATIVE: A COMPETITOR expresses the capacity of the Dative entity as compared to another entity that is either equal (*protivostojat'* 'withstand') or lesser in strength or influence (*poddavat'sja* 'give in to'), and is associated with the prepositions *k* 'toward' and *po* 'along'.

2.2.4 Accusative: Direction

The Accusative case signals a path toward a destination, or merely the endpoint of that path. ACCUSATIVE: A DESTINATION marks a direct object (*učitel' kupil* ***knigu*** 'the teacher bought **a book**'), which is a metaphorical version of the destination meaning, and is associated with metaphorical extensions to domains such as time (*v* ***ponedel'nik*** 'on **Monday**'), purpose (*otvet na* ***ego vopros*** 'the answer to **his question**'), change of state (*inogda ljubov' perexodit v* ***nenavist'*** 'sometimes love turns into **hatred**'), and mathematics (*v* ***četyre*** *raza* '**quadr**upled'). In the spatial domain, the path referenced by the Accusative case is further specified by prepositions such as *v* 'into', *na* 'onto', *za* 'beyond', *pod* 'under'. ACCUSATIVE: A DIMENSION measures a distance or size in the domain of space (*rasstojanie v* ***dva*** *kilometra* 'a distance of **two** kilometers'), or a duration in the domain of time (*interval v* ***dve*** *nedeli* 'an interval of **two** weeks'). ACCUSATIVE: AN ENDPOINT is primarily associated with the domains of space and time as specified by both prepositions (such as *v* and *za*, both indicating the end of a distance or duration, as in *za* ***odnu nedelju*** 'in/by the end of **a week**') and postpositions (such as *nazad* 'ago').

2.2.5 Instrumental: Accessory

The Instrumental case expresses "an accessory for something else" (Janda & Clancy 2002: 19). INSTRUMENTAL: A MEANS expresses a conduit for an action, such as a path that facilitates motion (as in *idti* ***lesom*** 'go **through/by means of the forest**') or an instrument that makes an action possible (as in *rezat' xleb* ***nožom*** 'slice bread **with a knife**'). This meaning is metonymically extended to include use with verbs signifying control (*zavedovat'* 'manage') and evaluation (*vostorgat'sja* 'be delighted with'), and to the agent in a passive construction (*kniga pročitana* ***studentom*** 'the book read **by the student**'). INSTRUMENTAL: A LABEL is used with verbs denoting being, becoming, and seeming, as in *koška javljaetsja* ***mlekopitajuščim*** 'a cat is **a mammal**'. INSTRUMENTAL: AN ADJUNCT occurs with the preposition *s* 'with' and expresses companionship. INSTRUMENTAL: A LANDMARK signifies peripheral locations without contact with the prepositions *nad* 'above', *pod* 'under', *pered* 'in front of', *za* 'behind', and *meždu* 'between'.

2.2.6 Locative: Location

The Locative case has only one meaning, LOCATIVE: A PLACE, which identifies locations in space or other domains, such as time (*v* ***ètom godu*** '**this year**', *pri* ***kommunizme*** 'during **the time of communism**') and states of being (*v* ***vostorge*** 'in **ecstasy**'). The meaning of the Locative case is always further specified by prepositions *v* 'in', *na* 'on', *pri* 'at', *o* 'about', *po* 'after'.

2.3 A coherent account of case semantics

While many of the details of case meanings listed above may seem trivial, their consolidation into a coherent system serves both theoretical and pedagogical purposes. This analysis brings a mass of disparate details together in a clear and elegant model. The model furthermore neatly predicts the use of case with novel vocabulary. For example, the borrowed verb *dirižirovat'* 'conduct (a musical group)' governs the Instrumental case, following the model of a group of native Russian verbs meaning 'manage, govern, lead' such as *rukovodit'* that govern the Instrumental. Similarly, the borrowed adjective *izomorfnyj* 'isomorphic' governs the Dative case, following the model of native Russian adjectives like *ravnyj* 'equal' that govern the Dative. And this model can be directly implemented in the classroom, for it is much more tractable than a long and scattered list of seemingly unmotivated contexts for one case or another that must be memorized. The model gives students a meaningful scaffold on which to build their understanding of grammar.

Just knowing the meanings of the cases, however, is not enough for a student to gain a secure grasp of Russian grammatical case. One also has to connect the cases to the morphemes that express them, as well as to the specific words and contexts in which the cases typically appear. Together with able teams of colleagues I have had the opportunity to build two resources to address these needs: The Strategic Mastery of Russian Tool (SMARTool) and the Russian Constructicon.

3. The Strategic Mastery of Russian Tool (SMARTool)

For decades I made beginning Russian students rehearse inflectional paradigms. I would write out the paradigm for a word on the board and have the students call out one form after another, then I would erase a couple of the forms, and make the students call out the forms again, and I would repeat this until the students were calling out the entire paradigm from memory in front of a blank chalkboard. I assumed that memorization of paradigms was necessary to equip students with inflectional forms in a way that mimicked the capacity of native speakers. Surely, I reasoned, all native speakers have somewhere in their internal grammars the entire paradigms of all words. But once large digital corpora started becoming available in the early 2000s, I began to suspect that I might not be right. Later, an experiment (Janda and Tyers 2018) proved me wrong. The results of this experiment inspired the creation of the Strategic Mastery of Russian Tool, called the "SMARTool" for short.

3.1 The Distribution and Learnability of Inflected Forms

A striking characteristic of all corpus data is the skewed distribution of items. The frequencies of words follow Zipf's Law (1949), a power law according to which the

second most frequent word is only one-half as frequent as the most frequent word, the third most frequent word is one-third as frequent, and so on, with a long tail of words that appear only once. The latter are known as hapaxes, which constitute one half of the total unique lexemes in a corpus. The very existence of so many hapaxes undermines the notion of paradigms: these words by definition cannot be represented in all their forms. And it is not just hapaxes that call the existence of paradigms into question: Zipf's Law applies also to inflected forms, meaning that even high frequency words have skewed distributions of forms within their paradigms. This fact has important implications for the understanding of paradigms, and indeed for the question of whether paradigms are a cognitive reality or just a convenience constructed by linguists and language pedagogues.

In an inflected language like Russian, nouns, verbs, and adjectives all have large numbers of inflected forms. Even a small vocabulary of a few thousand words represents over 100,000 potential forms. But the vast majority of those forms are rarely, if ever used, so one wonders whether we can assume that they are all in the heads of native speakers either.

The largest available corpora of Russian already exceed the volume of the lifetime exposure of a native speaker to their language. If we use a corpus as a proxy for such exposure, we can measure the skew in the distribution of inflected forms. In other words, we can estimate the frequency of various paradigm forms in the input that a native speaker would encounter. However, we don't need to measure from the largest corpora because Zipf's Law scales up: the proportions are stable even as corpus size grows. And this is fortunate because it means that we can use smaller "gold standard" corpora annotated for disambiguation of syncretic forms that yield reliable data on inflection.

When we examine corpus data, we find that even among high frequency words only about 10% of inflected forms are encountered frequently; the remainder are absent or rare. The percentage of lexemes in a word class that are attested in all paradigm forms depends upon the size of the paradigm, and this number decreases dramatically as the size of the paradigm increases. For the small paradigm of English nouns with only two forms – Singular and Plural – only 24% of nouns are found in both forms in a corpus. Norwegian marks both number and definiteness on nouns, meaning that there are four forms in the paradigm, but we find only 3% of nouns in all paradigm forms in a corpus. With a bigger paradigm like that of Estonian nouns with 28 forms, the number of nouns attested in all forms in a corpus is vanishingly small, approaching zero. Russian has a moderate-sized noun paradigm of twelve forms if we combine the second Locative (as in *v snegu* 'in the snow') with the Locative, the second Genitive (as in *čaju* 'some tea') with the Genitive, and the second Accusative (*pojti v soldaty* 'join the ranks of soldiers') with the Accusative and leave aside the "new" vocative (*Svet!* 'Sveta!'). Only 0.06% of nouns appear in the full set of paradigm forms in a Russian corpus (see more on this research in Janda & Tyers 2018).

In light of this distribution, it is reasonable to ask: how can Russian inflection be learned? Francis Tyers and I ran a machine-learning experiment that tested two possible

answers to this question: learning by exposure to full paradigms vs. learning by exposure to only the lemma and the single most frequent inflected form of each word. Our experiment is explained in full detail in Janda and Tyers 2018, so I offer only an abbreviated description here. We ranked nouns, verbs, and adjectives according to their frequency in a corpus, and took the 5400 most frequent lexemes (this was the ceiling set by a threshold for frequency and available data), dividing them into groups of 100, starting from the highest frequency items. Aside from the fact that the full paradigms model got to see the whole paradigm of each word, whereas the single forms model saw only the most frequent form, the experiment was the same for both tests. First the two models were trained on the top 100 words, then each model was given just the lemmas of the second 100 (unseen) words as a test. The test was to produce a specific inflected form (actually the most frequent form for that lemma) given only the lemma for each of the 100 previously unseen words. The machine's guesses were recorded and scored for accuracy. Then the second 100 words were added to the training data and the third 100 words were used to test both models. And then the third 100 words were added to the training data and the fourth 100 words were used to test both models. This procedure was iterated until we ran out of data at the 54th trial. The results were remarkable. Whereas both models performed poorly in the first few iterations, by the time they reached the sixteenth iteration, the single forms model surpassed the full paradigms model, which it consistently outperformed both in terms of overall accuracy and in terms of the egregiousness of errors (measured as Levenshtein distance between an error and the correct form).

In sum, the machine found it easier to master Russian inflection when learning only the most frequent word forms than when learning entire paradigms. The single forms model made fewer errors and the errors it did make were not as bad. This finding is consistent with a usage-based cognitively plausible model of morphological inflection. Given this outcome, it was clear to me that I needed to make a radical change in the way I taught inflection. If learning inflection by means of entire paradigms was too hard for a computer and entire paradigms are not reflected in corpus data, I shouldn't be asking my students to learn that way. Corpus data would play a major role in creating a new learning resource, making it possible to discover exactly what forms are most frequent for each lexeme.

3.2 Building and Using the SMARTool

Inspired by our experiment and funded by a grant from the Norwegian Directorate for Higher Education and Skills, I set about creating the Strategic Mastery of Russian Tool (SMARTool) together with a team of colleagues and students at UiT The Arctic University of Norway, the Higher School of Economics in Moscow, and the University of Helsinki: Radovan Bast (UiT), Tore Nesset (UiT), Francis Tyers (HSE), Mikhail Kopotev (UH), Valentina Zhukova (HSE), Elizaveta Kibisova (HSE), Svetlana Sokolova (UiT), Evgeniia Sudarikova (HSE), Ekaterina Rakhilina (HSE), Olga Lyashevskaya (HSE), and James McDonald (UiT). The SMARTool is freely available to the public without any password or login at: https://smartool.github.io/smartool-rus-eng/ and all data and code is stored open-source on github. A subset of the SMARTool, called SMARTool for Min russiske reise (https://smartool.github.io/min-russiske-reise/) serves just the A1 vocabulary broken down according to the lessons in our introductory inline course materials (a free

MOOC available at https://open.uit.no/courses/course-v1:UiT+mrr+2023/about). The building process and functions of the SMARTool are summarized here; for more details see Janda 2019.

The guiding principles for this project were that: 1) machine learning indicates that focus on the most frequent word forms is the best path to full mastery of inflectional morphology, and 2) language technology resources make it possible to identify the most frequent word forms and the grammatical constructions and collocations that motivate their use. In other words, our aim was to make learning of inflection maximally strategic by focusing on authentic usage. Of course language teachers have always focused on certain forms and contexts that are commonly encountered, but this has been based on intuition. For the first time we would do this in a scientific way, designing a resource based on empirical evidence.

We aggregated from textbooks a vocabulary of over 3000 inflected words, consisting of nouns, verbs, and adjectives and representing the CEFR (Common European Frame of Reference) language proficiency levels A1, A2, B1, and B2. In the research for Janda & Tyers 2018, we had learned that even high frequency words tend to appear commonly in only three or fewer inflected forms, so our goal was to discover which forms were the most strategic for each of the 3000 vocabulary items. For this task we turned to the SynTagRus corpus, a "gold standard" corpus which offers 100% manually corrected disambiguation of forms. For most words we collected the three most frequent forms, but if fewer than three forms accounted for over 90% of the attestations of a word, then we collected only those forms. For example, over 90% of the attestations of the noun *sentjabr'* September are either the Genitive Singular *sentjabrja* or the Locative Singular *sentjabre*, so we collected only those two forms. Once we had collected the most strategic inflected forms, we needed to identify their typical contexts in order to show how they are used. We consulted a variety of corpora (primarily the RNC and the Collocations Colligations Corpora at http://cococo.cosyco.ru/) to find representative example sentences that we then edited as necessary for the various levels. Finally we designed a user-friendly website.

In the SMARTool, a user first chooses the appropriate proficiency level (A1 through B2, or "all levels") and then selects the vocabulary to focus on through one of three filters: topic, analysis, and dictionary. All searches return words represented by their three or fewer most frequent inflected forms presented in example sentences. The user can click a button to show English translations of the sentences and can click another button for audio of each sentence. The "Search by dictionary" button returns a list of words at the given level. The "Search by topic" button offers a menu of topics, such as *vremja (time)*, *eda (food)*, and *životnye/rastenija (animals/plants)*, and users can toggle through all the items in the given category. For example, under *eda (food)*, one finds the word *sous* 'sauce' and these three sentences with the top three most common inflected forms of the word:

Vasja prigotovil kuricu v slivočnom souse. (Loc.Sing)
'Vasya cooked a chicken in a creamy sauce.'
Ljuboe mjaso on ljubit est' s soevym sousom. (Ins.Sing)

‘He likes to eat all kinds of meat with soy sauce.’
Lučše vsego on gotovil tomatnyj sous. (Acc.Sing)
‘Best of all he could cook tomato sauce.’

The “Search by analysis” button is handy for finding words and contexts for specific combinations of grammatical categories. For example, if one wants to find the most strategic words for learning the Dative Plural at the A2 level, the SMARTool returns these items in corpus-inspired example sentences: *pričinam* ‘reasons’, *sapogam* ‘boots’, *sportsmenam* ‘athletes’, *stroiteljam* ‘builders’, *šaxmatam* ‘chess’. If at the B2 level one searches for Perfective Gerunds, one gets a longer list of items including *ogloxnuv* ‘deafened’, *ogljanuvšis’* ‘(after) taking a look around’, *posočustvovav* ‘feeling sorry for’.

While the SMARTool provides information on the most likely combinations of all grammatical categories for each word, case is perhaps the most prominent category, since it relates to two of the three parts of speech in the SMARTool – nouns and adjectives – and one of those, nouns, is by far the most common part of speech, both in corpora of Russian and proportionately also in the SMARTool. Therefore, a major strength of the SMARTool is the way it represents case usage.

Another resource inspired by the research in Janda and Tyers 2018 has been created for Czech: GramatiKat (Kováříková et al. 2023; https://korpus.cz/gramatikat/). The GramatiKat interface allows users to view the distribution of morphological case both as a baseline (i.e., for all lexemes of a given part of speech) and for individual lexemes. GramatiKat opens the way for researchers to gauge differences in grammatical distributions between a reference corpus and target texts both overall and at the level of specific lexemes.

Of course it is one thing to build a resource and quite another thing to get students to actually use it. To this end we have devised a secondary resource with exercises to engage students with the SMARTool: https://smartool.github.io/exercises/. The SMARTool exercises are of two types, Treasure Hunt and Story Time, designs that emerged from work with a student focus group. Both types of exercise can be part of self-study, assigned as homework, or used in group work in a classroom.

The Treasure Hunt prompts the learner to use a SMARTool search function to gather data to help them to find the answer to a question. The questions range across levels of proficiency and probe various topics relating to patterns that students might not otherwise notice on their own, such as:

- Most Russian words beginning in *a-* or è- are foreign borrowings.
- The word rossijskij ‘Russian’ is used to describe items connected to Russia as a state (like *pasport* ‘passport’ *and Federacija* ‘Federation’) but *russkij* ‘Russian’ is used to describe items connected to the Russian language, culture, and ethnic identity (like *alfavit* ‘alphabet’, *literatura* ‘literature’).
- The prepositions *na* ‘on(to)’ and *s* ‘from’ are used with large open spaces or events, while other places use the prepositions *v* ‘in(to)’ and *iz* ‘from’.

Story Time trains learners to compose texts on various topics, using vocabulary, grammatical constructions, and collocations modeled in the SMARTool. For example, a B1 learner is asked to write 2-3 connected sentences on the topic of *zdorov'e (health)* using a given set of SMARTool vocabulary items, and in the SMARTool the student also finds examples of how these words are used in sentences with specific collocations and grammatical contexts:

- *prinimat'* 'take': + *lekarstvo* 'medicine'
- *operacija* 'operation': + *na* 'on' + Locative; + *provoditsja pod* občšim *narkozom* 'is conducted under general anesthesia'
- *želudok* 'stomach': *u* 'by' + Genitive + *bolit* 'hurts' + ('X has a stomach ache'); bol' 'pain' + *v* 'in' + Locative ; *rasstrojstvo* 'upset' + Genitive
- *analiz* 'analysis, test': + *krovi* 'blood'; *rezul'taty* 'results' + Genitive

An enduring theme of our work with the SMARTool has been that inflectional morphology doesn't happen in a vacuum; it is part of a bigger ecosystem of context involving word-specific preferences for both collocations and grammatical constructions. The lack of adequate resources to address this ecosystem motivated us to undertake another project, namely the building of the Russian Constructicon.

3.2 The Russian Constructicon

Like the work on case semantics, this project grew out of a frustration with existing resources. According to Construction Grammar (Goldberg 2006), an entire language can be described in terms of the form-meaning pairings that constitute grammatical constructions, but the vast majority of constructions are not represented in reference works. An example of the multiword constructions that are underrepresented is NP-Dat Copula *daleko do* NP-Gen, as in *Tebe daleko do lučšego rabotnika* 'You are by far not the best worker (lit. To you it is far to the best worker)'. The Russian Constructicon (Janda et al. 2018; https://constructicon.github.io/russian/) is an attempt to fill this gap, and our online resource currently provides semantic and syntactic descriptions, examples, and much more for over 4000 Russian constructions on a website that is free, open to the public, and searchable according to a large number of parameters. Case semantics play a role in a large portion of Russian multiword grammatical constructions, and the initial inventory of the Russian Constructicon was based on my earlier work on case semantics, and then later expanded through various methods (Janda et al. 2021). The Russian Constructicon is a multipurpose resource, designed to serve linguists as well as learners and teachers of Russian, and has spawned further publications (Endresen and Janda 2020; Janda et al. 2023; Janda, Endresen and Zhukova 2024; Zhukova and Janda 2024; Rakhilina et al. 2022).

3.2.1 Theoretical and practical arguments for a constructicon

Linguists traditionally describe languages in terms of a lexicon and the rules of basic grammar that operate on lexemes. The theoretical framework for the Russian Constructicon project, Construction Grammar (Fillmore and Kay 1999, Croft 2001, Tomasello 2003, Fried & Östman, 2004, Goldberg, 2006), however, takes a very different approach to language description by taking the construction as the basic (but not elementary) unit of language and claiming that an entire language can be described in terms of an interconnected system of constructions. Goldberg (2013: 17) defines

constructions as “conventional, learned form-function pairings at varying levels of complexity and abstraction”. This definition is intentionally very broad; it recognizes all language structures as constructions. At the extremes of the two dimensions of complexity and abstraction are items that are readily recognized by traditional linguistics. Examples of highly complex constructions are entire discourse structures such as an interview or a short story. The minimum of complexity is a simplex item with only one unit, and these can be found at both ends of a continuum from concrete to abstract. The concrete simplex items of language are individual words and morphemes, like the Russian adverb *daleko* ‘far’ and the preposition *do* ‘to’, and these are represented in dictionaries. In constructions, we term such items “anchors”. The abstract simplex items of language are bits of core grammar such as the subject of a sentence or the object of a preposition and are defined by the grammatical categories they express, such as case, tense, etc. In constructions we call such items “slots” and refer to the lexemes that fill slots as “fillers”. These abstract slots belong to the core syntax typically described in a grammar. In our construction above there are two NP slots, each with a case value (Dative and Genitive), as well as a copula for which the tense is not specified. Between these extremes there are thousands of essential multi-word expressions comprised of one or more anchors and/or fillers, the vast majority of which are not represented in traditional reference works. While all of these items, both the extremes and the multi-word expressions, are constructions, Construction Grammar tends to focus primarily on the multi-word expressions in an attempt to fill this gap.

A constructicon is a collection of the constructions of a language. While it is perhaps not feasible to create a resource that would contain all of the constructions of a language, the Russian Constructicon project takes seriously the tenet that this is in principle possible, resulting in the largest existing constructicon for any language, currently with over 4000 constructions. The patterns that emerge from this large-scale constructicon make it possible to trace the relationships that hold among constructions and the contexts in which various phenomena exist. For example, rather than investigating reduplication in isolation, it is now possible to extract the subset of grammatical constructions that have repeated elements and reveal their relationships to the rest of the Russian Constructicon (Janda, Endresen and Zhukova 2024). The Russian Constructicon is not a list. We find that “no construction is an island”; the Russian Constructicon is an interconnected system of thousands of constructions in which lexicon and grammar are fully integrated. Particularly striking are the arrays of semantic connections that join constructions into groupings across all levels, from the most local families of (nearly) synonymous constructions to the most abstract high-level semantic classes. We additionally find a variety of syntactic affinities across constructions, as well as links based on morphology, and the use of specific anchor and filler lexemes (for more on the systematic relationships among constructions, see Zhukova and Janda 2024).

3.2.2 Building and using the Russian Constructicon

The Russian Constructicon is an ongoing team effort that has involved collaboration between faculty and students at both UiT The Arctic University of Norway and the Higher School of Economics in Moscow. Some of my most prominent collaborators are:

Radovan Bast (UiT), Anna Endresen (UiT), Daria Mordashova (HSE, MGU), Ekaterina Rakhilina (HSE), Valentina Zhukova (UiT), and at least forty students over a period of nearly a decade have contributed. The Russian Constructicon project has received financing from the Norwegian Directorate for Higher Education and Skills, the Ministry of Science and Higher Education of the Russian Federation, and the National Research Foundation of Korea. The Russian Constructicon is a free open-source resource available without registration or password.

Case semantics has played a major role in the Russian Constructicon from the very beginning, when our first collection of constructions was derived from the pages of the Case Book for Russian (Janda and Clancy 2002). Since then we have employed a variety of methods, including manual collection from reading texts and scripts, semiautomatic collection of frequent multiword collocations, and intuitive probing of native speakers' competence to fill out families of (nearly) synonymous constructions (for details on this process, see Endresen et al. To Appear). And since nearly every grammatical construction contains a noun phrase or an adjective or a participle (i.e., something that can be inflected for case), grammatical case figures prominently in the entire Russian Constructicon.

When a user opens the Russian Constructicon page, they find a window where they can browse over 4000 constructions. From this homepage is possible to filter constructions by typing in specific anchor words or slot tags, as well as selecting a proficiency level (from A1 to C2). For example if we type in (using Cyrillic) the word *daleko* 'far', we find eleven constructions with that anchor word, among them the construction mentioned above. When we click on that construction, we get this information (here additionally annotated with information in square brackets, and with all Cyrillic rendered in Latin transcription, and translations of Russian text):

473 [an ID number used internally by developers]
NAME NP-Dat Cop daleko do NP-Gen [the name of the construction]
Tebe daleko do lučšego rabotnika. [a short recognizable illustration of the construction, here: You are far from being the best worker]

DEFINITION (Russian) [most constructions come with a definition in Russian, some also have Norwegian and English definitions, this is still under development]
Konstrukcija oboznačaet, čto [učastnik situacii]$_{\text{Participant}}$ ili [ob"ekt]$_{\text{Theme}}$ ne obladaet dostatočnymi kačestvami i nedostatočno xoroš, čtoby byt' kak [ètalon]$_{\text{Standard}}$. Konstrukcija osnovana na sravnenii i soderžit ocenočnuju xarakteristiku vozmožnostej ili kačestv [učastnika]$_{\text{Participant}}$ ili [ob"ekta]$_{\text{Theme}}$ kak značitel'no ustupajuščix vozmožnostjam ili kačestvam [togo ètalona, s kotorym oni sravnivajutsja]$_{\text{Standard}}$. Kak esli by govorjaščij sčital, čto rasstojanie ot učastnika ili ob"ekta do ètalona očen' veliko.
[The construction indicates that the [participant in the situation]$_{\text{Participant}}$ or [object]$_{\text{Theme}}$ does not possess sufficient qualities and is not good enough to be like the [standard]$_{\text{Standard}}$. The construction is based on comparison and contains an evaluative characteristic of the abilities or qualities of the [participant]$_{\text{Participant}}$ or [object]$_{\text{Theme}}$ as significantly inferior to the abilities or qualities of the [standard with

which they are compared]$_{Standard}$. It is as if the speaker believes that the distance from the participant or object to the standard is very great.]

EXAMPLES [Five corpus examples are given, but here we show only one]

1. Vidite li, delo v tom, čto [gubernatoru oblasti]$_{Participant}$ eščë daleko do [prezidenta]$_{Standard}$.
   [You see, the point is that the regional governor is far from being the president.]

CEFR LEVEL: A2

When the user clicks to get additional information, they can find: equivalent constructions in Norwegian and English; common fillers; the semantic and syntactic types of the construction; the syntactic function, structure, and part of speech of the anchor; the dependency structure of the name of the construction and its illustration; the communicative type of the construction (e.g., Declarative); a usage label (e.g., Colloquial); a comment (often citing closely-related constructions); and references to relevant scholarly works.

The Advanced search page of the Russian Constructicon allows the user to filter constructions according to all parameters for which constructions are tagged: semantic types, semantic roles, morphology, syntactic type of construction, syntactic function of anchor, syntactic structure of anchor, part of speech of anchor, and CEFR level. Among other things, this makes it possible to search for constructions that involve each of the grammatical cases.

On the Daily dose page a user can choose a proficiency level and receive five randomly selected grammatical constructions to train on.

The Statistics page shows graphs and raw numbers for the distribution of syntactic types of constructions, syntactic functions of anchors, and semantic types of constructions. The graph for semantic type can be modified to show only the distribution for a selected syntactic type. This page also lists the ten most frequent anchor words for each of three parts of speech: verbs (starting with *znat'* 'know', *govorit'* 'say', *xotet'* 'want'), nouns (starting with *vremja* 'time', *delo* 'thing, case', *raz* 'time'), and adjectives (starting with *ravnyj* 'equal', *xorošij* 'good', *polnyj* 'full').

The site has an Instructions page to guide the user through all the terms and conventions and an About page that describes the project. There is also a YouTube channel with instructional videos about the project: https://www.youtube.com/channel/UC8q-_F8c8bx9gl7fYET1-dQ.

Several spinoff projects are under development, including constructicons for Ukrainian, Persian, and Hill Mari. Since the code is open-source and publicly available it is possible for researchers who wish to create constructicons for other languages to reuse and adapt our model.

As with the SMARTool, a further challenge is to make the Russian Constructicon more accessible to language learners. To this purpose we have created a related resource, called Construxercise!: https://constructicon.github.io/construxercise-rus/. This resource facilitates hands-on learning of Russian constructions through exercises aimed at a strategic group of 57 Russian highly frequent discourse constructions that students can use to structure a discourse by doing things like introducing a topic, clarifying a point, giving an example, adding information, expressing an opinion, asking someone for their opinion, hedging, drawing a conclusion, etc. The constructions that support these skills are presented in twelve lessons on topics like Introducing yourself, Getting a job, Getting around, Holiday celebrations. Construxercise! is conceived of as a multifunctional resource that serves the needs of different types of users and offers educational materials that can be used as either a central or complementary teaching component, either in class or for self-guided study.

4. Analysis of political discourse: Putin makes his case

Of course, it is not only linguists and language learners who use case. Grammatical case is a feature of over two-thirds of the world's languages, used by all speakers of those languages, and it makes sense to ask what role case is playing particularly in the most powerful of those speakers. This brings us to Putin and a question that has bothered me for a long time: Why is Putin so popular? Why do Russians find him convincing? Public opinion polls (see https://www.levada.ru/en/ratings/, https://media.fom.ru/fom-bd/d46pi2024.pdf) have consistently shown Putin's approval rating at between 60% and 90% over the past quarter century (at the time of this writing it stands between 82% and 87%). Although opinion polls carried out in Russia are not entirely reliable, certainly this means that there are a lot of Russians who stand behind their leader. Putin is not a brilliant orator, as anyone who has watched his hours-long speeches can attest, but maybe there is something in the way he delivers his messages that makes them compelling to his listeners. In Janda et al. 2022 we looked just at how Putin uses grammatical case, and found consistent deviations from Russian norms. This research was carried out in collaboration with Masako Fidler (Brown University), Václav Cvrček (Charles University), and Anna Obukhova (UiT) and funded by a grant from the Norwegian Research Council (https://threat-defuser.org/).

Our research is based on four assumptions. The first is that a corpus of a language of a sufficient size can serve as a proxy for the linguistic experience and expectations of native speakers. A corpus is perhaps an imperfect representation but it is the closest thing we have to a model of the input that gives a native speaker their special competence in a language, their conscious and unconscious knowledge of the norms of their language. Second: speakers are known to be sensitive to deviations from these norms. Third: while words can be consciously chosen, grammar less under conscious control and more systematic. Fourth: grammar and meaning are joined in a semantic continuum; grammar is not just empty scaffolding.

In a nutshell, our idea was to compare Putin's use of grammatical case with what we find in a corpus of Russian and analyze the deviations for how they support his political messages. To this end, we performed the first extension of Keyword Analysis to a new methodology we call "Keymorph Analysis". Keyword Analysis (cf. Scott 1996) is a well-

established method widely used in corpus-assisted discourse analysis. Keyword Analysis focuses on the distribution of words, identifying as "keywords" those that are unusually frequent in a target text as compared to a reference corpus. In this way keywords reveal the "aboutness" of a text. However, Keyword Analysis has mostly been performed on English, which has little morphology and no grammatical case. We created the first proof-of-concept for Keymorph Analysis using as our target text Putin's speeches during a three-week period leading up to and following the full-scale invasion of Ukraine in February 2022 (34,720 tokens), and as our reference corpus the Russian InterCorp portion of the Czech National Corpus (www.korpus.cz, 20.1 million tokens).

We examined Putin's use of case with three words: *Rossija* 'Russia', *Ukraina* 'Ukraine', and *NATO* 'NATO'. These three words occur a total of 395 times in the Putin target text and 7801 times in the reference corpus. All attestations of these words in the target text were manually annotated for the precise case meaning expressed. The relevant case meanings that appeared most often with these words in both the target text and the reference corpus are the following:

- Nominative: agent (subject); label
- Genitive: agent or patient; possession
- Dative: potential agent (usually human)
- Accusative: patient (direct object); destination
- Instrumental (with preposition *s*): collaborator
- Locative: a place

We found that Putin's use of grammatical case with the three nouns deviates significantly from the case usage observed in the reference corpus, and that Putin's usage strongly underpins his political message. *Rossija* 'Russia' is statistically overrepresented in the Genitive and Dative cases, *Ukraina* 'Ukraine' is overrepresented in the Genitive case but underrepresented in all other cases, and *NATO* 'NATO' is overrepresented in the Accusative case and strongly underrepresented in the Dative and Instrumental cases.

Closer examination of the specific case meanings the Putin uses are more revealing. *Rossija* 'Russia' is represented as a dynamic agent (Nominative subject of transitive verbs), a collaborator (Instrumental case), a victim that has been treated unfairly (Accusative), and as a humanized entity that inspires empathy (Dative). *Ukraina* 'Ukraine' by contrast plays a passive role (Nominative subject with stative verbs), is manipulated (Accusative) and dehumanized (Dative severely underrepresented), is not a collaborator (Instrumental severely underrepresented), and is merely a location or region (use of *na* 'on' + Locative and Genitive). *NATO* 'NATO' is similarly dehumanized and not seen as an agent (Nominative) or a collaborator (Instrumental). NATO's signature role in Putin's narrative is as a future destination for Ukraine (Accusative, Locative). In sum, Putin depicts Russia as a dynamic, agentive, foregrounded actor, a reliable partner for collaboration, but also the victim of unfair geopolitical maneuvers. Ukraine, by contrast, is dehumanized, relatively static, and backgrounded, often merely a territorial location rather than a state. NATO appears primarily as the label for an untrustworthy organization and as a destination for Ukraine.

One year after the full-scale invasion of Ukraine, on February 21, 2023 Putin delivered a speech to the Federal Assembly (10,538 tokens), which news media declared to be "more of the same". However, in a further analysis we found some important shifts in Putin's message conveyed by grammatical case. In this speech Putin emphasized the great potential of Russia's self-sufficient economy and the ways that Russia has been unfairly targeted by the West. Ukraine was mentioned only twelve times in this speech, referred to mainly as Russia's "historical territories" and the West's "Anti-Russia". NATO was no longer depicted as the destination of Ukraine, but instead foregrounded as an aggressor.

We have demonstrated that Keymorph Analysis can complement Keyword Analysis and other traditional methods of discourse analysis. Over- and underrepresentation of grammatical cases can be identified by measuring deviations from corpus norms. This method of analyzing grammatical case reveals the roles of social actors in a discourse, and can be used not only by linguists, but also in the disciplines of the social sciences. While one's choice of words is deliberate and conscious, grammatical case is obligatory and serves as a second channel for signaling the roles notions have in a discourse. We reason that consistent deviation from grammatical norms likely has an impact on hearers, driving home a message like a steady drumbeat. Our results invite further comparisons, for example of Putin with other politicians, and with messages in various types of manipulative texts.

5. Conclusion

This story of linguistic theory and its application to language pedagogy is both a professional one and a personal one. The meanings of grammatical case that so frustrated me as a student have inspired an enduring fascination that leads in many directions at once. No matter where I turn, the cases keep coming back to me. This research agenda has supported the core tenet of Cognitive Linguistics that grammar has meaning. I have learned that native speakers probably don't have a full set of paradigms in their heads; instead they most likely triangulate from many smaller partly overlapping subsets of paradigms comprised of the most common forms for individual words. We made a resource to reflect this finding and I changed my pedagogical approach accordingly. We have filled in the some of the gaps between what we find in dictionaries and grammar books with descriptions of thousands of multi-word constructions. And we have used the statistical distribution of grammatical case to probe the ideological messages of Vladimir Putin.

References

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Dryer, Matthew S. and Martin Haspelmath, (eds.) 2013. *WALS Online (v2020.4)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591.

Endresen, Anna, Laura A. Janda. 2020. "Taking Construction Grammar One Step Further: Families, Clusters, and Networks of Evaluative Constructions in Russian. In Mike Putnam, Matthew Carlson, Antonio Fábregas, Eva Wittenberg, eds. *Defining*

Construction: Insights into the Emergence and Generation of Linguistic Representations (= special issue of *Frontiers in Psychology* 11), 1–22. https://doi.org/10.3389/fpsyg.2020.574353

Endresen, Anna, Laura A. Janda, Ekaterina Rakhilina, Valentina Zhukova, Daria Mordashova. To Appear. From a list to a network: Growing the Russian Constructicon via research on families of constructions. In: *Constructing constructicons*. John Benjamins.

Fillmore, Charles J. and Paul Kay. 1999. *Construction Grammar.* Berkeley: University of California.

Fried, M., & Östman, J.-O. 2004. Construction grammar: A thumbnail sketch. In M. Fried & J.-O. Östman (Eds.), *Construction grammar in a cross-language perspective* (pp. 11–86). John Benjamins.

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

Goldberg, Adele. 2013. Constructionist approaches. *Construction Grammar Handbook*. Oxford University Press volume, Thomas Hoffmann and Graeme Trousdale (eds.), 15–31.

Janda, Laura A. 1993. *A Geography of Case Semantics: The Czech Dative and the Russian Instrumental* (=*Cognitive Linguistics Research,* v. 4). Berlin: Mouton de Gruyter.

Janda, Laura A. 1999. "Peircean semiotics and cognitive linguistics: a case study of the Russian genitive," in *The Peirce Seminar Papers*, ed. by Michael Shapiro. New York/Oxford: Berghahn Books, 441-466.

Janda, Laura A. 2000. "A cognitive model of the Russian accusative case," in *Trudy meždunarodnoj konferencii Kognitivnoe modelirovanie*, No. 4, part I, ed. by R. K. Potapova, V. D. Solov'ev and V. N. Poljakov. Moscow: MISIS, 20-43.

Janda, Laura A. 2019. "Businessmen and Ballerinas Take Different Forms: A Strategic Resource for Acquiring Russian Vocabulary and Morphology". *Russian Language Journal* 69, 175-193.

Janda, Laura A. and Steven J. Clancy. 2002. *The Case Book for Russian*. Bloomington, IN: Slavica.

Janda, Laura A. and Steven J. Clancy. 2006. *The Case Book for Czech*. Bloomington, IN: Slavica.

Janda, Laura A., Anna Endresen, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nesset, Svetlana Sokolova. 2013. *Why Russian aspectual prefixes aren't empty: prefixes as verb classifiers*. Bloomington, IN: Slavica Publishers.

Janda, Laura A., Francis M. Tyers. 2018. “Less is More: Why All Paradigms are Defective, and Why that is a Good Thing”. *Corpus Linguistics and Linguistic Theory* 14(2). doi.org/10.1515/cllt-2018-0031.

Janda, Laura A., Olga Lyashevskaya, Tore Nesset, Ekaterina Rakhilina, Francis M. Tyers. 2018. “Chapter 6. A Constructicon for Russian: Filling in the Gaps”. In Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, & Tiago Timponi Torrent (Eds.), *Constructicography: Constructicon development across languages* [*Constructional Approaches to Language* 22], 165-181. Amsterdam: John Benjamins Publishing Co. doi 10.1075/cal.22.06jan

Janda, Laura A., Anna Endresen, Valentina Zhukova, Daria Mordashova, Ekaterina Rakhilina. 2021. “How to build a constructicon in five years: The Russian Example”. In: Frank Brisard, Timothy Colleman, Astrid De Wit, Renata Enghels, Nikos Koutsoukos, Tanja Mortelmans, and María Sol Sansiñena (eds.), *The Wealth and Breadth of Construction-Based Research = Belgian Journal of Linguistics* 34: 161–173.

Janda, Laura A., Masako Fidler, Václav Cvrček, Anna Obukhova. 2022. “The case for case in Putin’s speeches”. *Russian Linguistics*. https://doi.org/10.1007/s11185-022-09269-2.

Janda, Laura A., Anna Endresen, Ekaterina Rakhilina, Valentina Zhukova, Daria Mordashova. 2023. From data to theory: an emergent semantic classification based on a large-scale constructicon. *Constructions and Frames* 15:1, 1–58. https://doi.org/10.1075/cf.00066.jan.

Janda, Laura A., Anna Endresen and Valentina Zhukova. 2024. Typology of reduplication in Russian: constructions within and beyond a single clause. Valentina Zhukova. In M. Kopotev and K. Kwon (Eds.), *Constructions with lexical repetitions in East Slavic languages* (Volume 384 in the series *Trends in Linguistics. Studies and Monographs [TiLSM]*), 71–96. 2024. De Gruyter Mouton. https://doi.org/10.1515/9783111165806-003.

Janda, Laura A., Anna Endresen, Ekaterina Rakhilina, Valentina Zhukova, Daria Mordashova. 2023. From data to theory: an emergent semantic classification based on a large-scale constructicon. *Constructions and* Frames15:1, 1–58. https://doi.org/10.1075/cf.00066.jan.

Kováříková, D., M. Beneš, L. Smejkalová, O. Kovářík. 2023. The Structuralist tradition meets empirical data: Corpus data enhancing the Czech Internet Language Reference Book. *Word Structure* 16:2–3, 233–257. https://www.euppublishing.com/doi/10.3366/word.2023.0230.

Langacker, Ronald W. 2008. *Cognitive Grammar.* Oxford: Oxford University Press.

Rakhilina, Ekaterina; Zhukova, Valentina; Demidova, Daria; Kudryavtseva, Polina; Rozovskaya, Gloria; Endresen, Anna; Janda, Laura Alexis. 2022. Фразеология в

ракурсе «Русского конструктикона». *Trudy Instituta Russkogo Iazyka imeni V.V. Vinogradova*; Volume 2 (32), 13–44. https://doi.org/10.31912/pvrli-2022.2.2

Rosch, Eleanor. 1973a. Natural categories. *Cognitive Psychology* 4, 328–350.

Rosch, Eleanor. 1973b. On the internal structure of perceptual and semantic categories. In Timothy E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, 111–144. New York: Academic Press.

Scott, M. (1996). *Wordsmith tools*. London: Oxford University Press/Lexical Analysis Software. https://lexically.net/wordsmith/.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Approach*. Cambridge: Harvard University.

Zhukova, Valentina and Laura A. Janda. 2024. Russian grammar as a constructicon: Beyond a list. In *Russian Grammar Updated,* a special issue of *Russian Linguistics* 48. https://doi.org/10.1007/s11185-024-09298-z.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.