RUSSIAN
JOURNAL OF LINGUISTICS

2022 Volume 26 No. 4

"Meaning ⇔ Text" Theory in the Linguistic Universe
In honour of Igor MEL'ČUK

Теория «Смысл ⇔ Текст» в лингвистической вселенной
К юбилею И.А. МЕЛЬЧУКА

RUSSIAN JOURNAL OF LINGUISTICS

Aims and Scope

The Russian Journal of Linguistics is a peer-reviewed international academic journal publishing research in Linguistics and related fields. It is international with regard to its editorial board, contributing authors and thematic foci of the publications.

The aims of the journal:

♦ to promote scholarly exchange and cooperation among Russian and international linguists and specialists in related areas of investigation;
♦ to disseminate theoretically grounded research and advance knowledge pertaining to the field of Linguistics developed both in Russia and abroad;
♦ to publish results of original research on a broad range of interdisciplinary issues relating to language, culture, cognition and communication;
♦ to cover scholarly activities of the Russian and international academia.

As a Russian journal with international character, it aims at discussing relevant intercultural/linguistic themes and exploring general implications of intercultural issues in human interaction in an interdisciplinary perspective. The most common topics include *language and culture, comparative linguistics, sociolinguistics, psycholinguistics, cognitive linguistics, pragmatics, discourse analysis, intercultural communication, and theory and practice of translation*. In addition to research articles, the journal welcomes book reviews, literature overviews, conference reports and research project announcements.

The Journal is published in accordance with the policies of *COPE* (*Committee on Publication Ethics*) **http://publicationethics.org**. The editors are open to thematic issue initiatives with guest editors.

Further information regarding notes for contributors, subscription, open access and back volumes is available at **http://journals.rudn.ru/linguistics**.

E-mail: **lingj@rudn.ru**

Цели и тематика

Журнал Russian Journal of Linguistics – периодическое международное рецензируемое научное издание в области междисциплинарных лингвистических исследований. Журнал является международным как по составу редакционной коллегии и экспертного совета, так и по авторам и тематике публикаций.

Цели журнала:

♦ способствовать научному обмену и сотрудничеству между российскими и зарубежными лингвистами, а также специалистами смежных областей;
♦ знакомить читателей с новейшими направлениями и теориями в области лингвистических исследований, разрабатываемых как в России, так и за рубежом, и их практическим применением;
♦ публиковать результаты оригинальных научных исследований по широкому кругу актуальных лингвистических проблем междисциплинарного характера, касающихся языка, культуры, сознания и коммуникации;
♦ освещать научную деятельность как российского, так и международного научного сообщества.

Будучи международным по своей направленности, журнал нацелен на обсуждение теоретических и практических вопросов, касающихся взаимодействия культуры, языка и коммуникации. Особый акцент делается на междисциплинарные исследования. Основные рубрики журнала: *язык и культура, сопоставительное языкознание, социолингвистика, психолингвистика, когнитивная лингвистика, прагматика, анализ дискурса, межкультурная коммуникация, теория и практика перевода*. Кроме научных статей публикуется хроника научной жизни, включающая рецензии, научные обзоры, информацию о конференциях, научных проектах.

Перечень отраслей науки и групп специальностей научных работников в соответствии с номенклатурой ВАК РФ: Отрасль науки: 10.00.00 – филологические науки; Специальности научных работников: 10.02.01 – русский язык, 10.02.04 – германские языки, 10.02.05 – романские языки, 10.02.19 – теория языка, 10.02.20 – сравнительно-историческое, типологическое и сопоставительное языкознание.

Журнал строго придерживается международных стандартов публикационной этики, сформулированных в документе *COPE* (*Committee on Publication Ethics*) **http://publicationethics.org**.

Правила оформления статей, архив и дополнительная информация размещены на сайте: **http://journals.rudn.ru/linguistics**.

Электронный адрес: **lingj@rudn.ru**

"Meaning ⇔ Text" Theory in the Linguistic Universe
In honour of Igor MEL'ČUK

CONTENTS

Теория «Смысл ⇔ Текст» в лингвистической вселенной
К юбилею И.А. МЕЛЬЧУКА

СОДЕРЖАНИЕ

Research article / Научная статья

"Meaning ⇔ Text" Theory and the linguistic universe of Igor Mel'čuk

Svetlana IVANOVA🆔 and Tatiana LARINA🆔✉

¹Pushkin Leningrad State University, St. Petersburg, Russia
²RUDN University, Moscow, Russia
✉larina-tv@rudn.ru

Abstract
The introductory article to the special issue dedicated to Igor Mel'čuk summarizes the main tenets of his "Meaning ⇔ Text" theory, and outlines its contribution to the development of diverse areas of modern linguistics. This theory can be characterized as a formalized, semantically oriented, multilevel, structural, functional and global model of language which explains the way the speaker generates the meaning embodied in the text. Both the article and the volume as a whole show the relevance of this theory for the development of semantics, grammar, pragmatics, typology and lexicography and highlight its theoretical and practical implications for linguistic studies and interdisciplinary research. We then briefly present the articles in this issue. Some of them were written directly in line with the "Meaning ⇔ Text theory" and were influenced by Mel'čuk's ideas; others were the result of the stimulus that this theory, like its author, gave to the comprehension of the complex issues arising from the description of various levels of the language system. But all of them are related to Igor Mel'čuk, both as a linguist and a personality, and they are dedicated to him.
Keywords: *Meaning ⇔ Text theory, Mel'čuk, semantics, linguistic meaning, formalization, functional model*

Теория «Смысл ⇔ Текст» и лингвистическая вселенная И. А. Мельчука

С.В. ИВАНОВА🅳, Т.В. ЛАРИНА🅳✉

¹Ленинградский государственный университет имени А.С. Пушкина, Санкт-Петербург, Россия
²Российский университет дружбы народов, Москва, Россия
✉larina-tv@rudn.ru

Аннотация

В вводной статье к специальному выпуску, посвященному И.А. Мельчуку, обобщены основные положения разработанной им теории «Смысл ⇔ Текст» и показан ее вклад в развитие различных направлений современной лингвистики. Данная теория предлагает формализованную, семантически ориентированную, многоуровневую, структурно-функциональную и глобальную лингвистическую модель языка и объясняет, как говорящий порождает воплощенный в тексте смысл. В статье, как и в выпуске в целом, показана актуальность рассматриваемой теории для настоящего этапа развития теории языка, включая семантику, грамматику, прагматику, типологию и лексикографию, подчёркивается ее теоретическое и прикладное значение для лингвистических и междисциплинарных исследований. Далее кратко представлены публикации номера, объединенные идеями, в которых заметно влияние работ И.А. Мельчука. Часть статей написана непосредственно в русле теории «Смысл ⇔ Текст», другие явились результатом стимула, которым послужили как данная теория, так и ее автор, при осмыслении сложных вопросов структурно-семантического описания языка. Объединяющим началом для всех публикаций выпуска является научное творчество и личность Игоря Александровича Мельчука, которому этот выпуск и посвящается.

Ключевые слова: *Теория Смысл ⇔ Текст, Мельчук, семантика, языковое значение, формализация языка, функциональная модель*

1. Introduction

This issue is dedicated to the internationally renowned linguist Igor Mel'čuk, the founder of the "Meaning ⇔ Text" linguistic theory and one of the pioneers of machine translation. His scholarly life is impressive for its outstanding results and the grand scale of problems discussed and solved. He worked for 20 years at the Institute of Linguistics of the Russian Academy of Sciences, and then another 32 at the University of Montreal, Canada. Since then he has continued to work actively, as Honorary Doctor of the Institute of Linguistics of the Russian Academy of Sciences, as Emeritus Professor at the University of Montreal and a member of the Royal Society of Canada. He lectures, advises young researchers, writes books and articles. So far, he has 305 published articles and 49 books, and work on the next book has begun. Thus, along with a personal anniversary, Igor Mel'čuk celebrates

his professional anniversary – 65 years of continuous research and academic creativity.

With his innovative ideas and devotion to science, he has inspired and continues to inspire his colleagues, novice researchers and students both in Russia and internationally, as his linguistic universe has no boundaries, and his human qualities such as kindness, openness and sparkling humour fascinate anyone who has had the privilege of interacting with him.

We invited some of Igor's longstanding friends, with whom he started and continued his career, his colleagues and students, as well as admirers and disciples of his research work to contribute to this issue. Together with them, we have prepared this intellectual and emotional gift to him. On behalf of the editorial board of the journal, our authors and readers, we wholeheartedly congratulate Igor Mel'čuk on his ninetieth birthday and wish him good health, inexhaustible optimism and new creative achievements.

In this issue, we intend to summarize the contribution Mel'čuk has made to semantics, to the development of the theory of meaning and the methodology of text analysis, as well as to trace the influence of his "Meaning ⇔ Text" theory on the further advancement of text linguistics, the theory and practice of machine translation, lexicography, typology, computational and corpus linguistics. In our introductory article, we will briefly dwell on the main ideas of this theory and outline the directions it has indicated for many fields of language study.

2. Meaning ⇔ Text Theory

2.1. Meaning ⇔ Text linguistic model: its highlights

The Meaning ⇔ Text Theory (MTT) was elaborated by Igor Mel'čuk about 60 years ago and became a primary focus for Mel'čuk as a researcher. Though the theory appeared in the late 1960s, it has not lost its relevance. Since then, different scholars from various schools of linguistics worldwide have been using its postulates, methodology or notions in their research. The theory explains the way the speaker generates the meaning embodied in the text which is understood as a meaningful unity. Many other domains of linguistic research have benefited from MTT, and its findings and approaches are applied in diverse studies.

Since its origins linguistics has sought answers to several major questions: How does language convey meaning? How do linguistic resources combine to produce it? To answer these questions, on the one hand, it is necessary to define what language is, as an entity, and how it is designed. On the other, it is crucial to account for the way meaning is rendered in communication which is viewed, first and foremost, as the transfer of meaning. Mel'čuk's theory concerns both questions, and builds an integral approach that unites the intrinsic features of the language and the way they are responsible for rendering meaning. MTT combines both strands in an integrated vision of the linguistic construction of meaning via text.

The theory describes a natural language which is understood as a logical device (Mel'čuk 2021: 10), i. e., one designed according to the laws of logic. From our temporal perspective, it seems possible that the idea of such a device may have been influenced by the ideas of the 60s, when a profound interest in cybernetics and the power of machines permeated all branches of scholarly research. This "device" (or machine) provides transition from the meaning that the speaker wants to render to the text which serves as its carrier. Though it is mentioned that the device works both ways – from meaning to text for the speaker and vice versa for the listener, MTT prioritizes text generation (in other words, speaking) over text interpretation (comprehension), reflected in the name of the theory: from meaning to text, which is why we have Meaning ⇔ Text Theory and despite the special sign of a two-way transition, the meaning in the name still comes first.

Another idea popular at the time MTT was born is that of modelling objects to understand how they work. Mel'čuk devised a model of a natural language which is understood as "a set of rules encoded in the brains of its speakers that establish a correspondence between meanings of **L**(anguage) and their expression, or texts of **L**(anguage)" (Mel'čuk 2020: 4). Since language is a set of rigorous rules that establish correspondences between meanings and possible connected expressions that could surface in text, this system of rules could serve as a model of language.

This model is formalized via a system of formal, symbolically represented rules. These decode the process of exteriorization of deep level dependences among linguistic units through trees of dependence and the description of lexical functions. Mel'čuk also mentions essential requirements for these rules: they must be logically consistent, unambiguous, and transparent in terms of rendering content (Gladky & Mel'čuk 1969: 9) as well as "parsimonious and elegant in their form" (Mel'čuk 2020: 5). In Mel'cuk's words, "linguistic meanings (in the technical sense of the term) appear as formal symbolic objects called semantic representations [SemRs], and texts—as phonetic representations [PhonRs]" which could be expressed by way of a formula $\{SemR_i\} \Leftarrow language \Rightarrow \{PhonR_j\} \mid i \neq j, 0 < i, j \leq \infty$ (Mel'cuk 2021: 9). Several kinds of formalism are used: "semantic networks for representing meanings of sentences and lexical units; dependency trees for representing the syntactic structure of sentences; lexical functions for representing lexical relations; and rules of various types for representing semantic operations" (Mel'čuk & Milićević 2020: xvii).

This model gives weight to the semantic component of meaning. Mel'čuk states that semantics "is not "just another component" of a linguistic system: it occupies a special place because language is above all a communication tool – that is, a means for conveying meaning" (Mel'čuk & Milićević 2020: 6). Thus, the semantic component constitutes the core of language description, as semantic properties of linguistic expressions largely predetermine their syntactic behavior as well as influence their morphology (Mel'čuk & Milićević 2020: 6)[1]. Consequently,

---

[1] This statement may seem obvious at first glance, but we should remember that one of the most widespread theories of language today, generative grammar and its connected ideologies, consider

this model establishes close ties between vocabulary and grammar. According to the Meaning ⇔ Text Theory, semantics is viewed as a system of rigorous formal rules since language *per se* is represented as a set of rules that establish correspondences between meanings and their possible expressions. Another idea behind the semantic model is the possibility of meaning decomposition, which assumes that complex meanings may be represented via simpler ones (cf. Anna Wierzbicka's notion of "semantic primitives" [Wierzbicka 2021], see also [Mel'čuk 1989]). Mel'čuk claims that semantic representation is based on a limited number of elementary semes which in some way resemble, and may be called, "atoms of meaning" (Mel'čuk 1999: 58).

As may be seen, this model is multilevel or, in Mel'čuk's words, stratificational (= multi-stratal) (Mel'čuk & Milićević 2020: 10), for language is understood as a multi-level entity. The author singles out a hierarchy of levels including semantic (meaning), syntactic, morphological, phonological and phonetic / graphic (text). At each major representational level except the semantic, a Meaning-Text Model recognizes two sub-levels which are called 'deep' and 'surface'. Thus, we can speak about a deep-syntactic representation and a surface-syntactic representation, and so on. Mel'čuk provides the following explanation for these sub-levels: "A deep sub-level is oriented towards the meaning – that is, towards the content the Speaker wants to express; its task is to explicitly reflect the relevant informational distinctions. A surface sub-level is oriented towards the text – that is, towards the form in which the content is expressed; its task is to explicitly reflect all relevant formal distinctions" (Mel'čuk 2020 & Milićević: 16).

In terms of the structural features of each layer, semantics is modelled by a network (more than one way in to each point, more than one way out), while the final output, a phonological string, is modelled by a line (one way in and one way out from each point). The intermediate, syntactic layer is modelled by trees, which, appropriately, have just one way in to each point, but more than one way out. In its overall architecture, then, MTT is intended to parallel the way meanings – ultimately related to concepts held in the networked structure of neurons in the human brain – pass via morphosyntactic structures (trees) to the linear string of phonological production.

The model is also structural as language itself and its units are understood as structures. Thus, semantics is described in terms of semantic structures which are represented as networks with nodes. Sentences and their semantics are represented

---

language in a fundamentally different way: not as a tool for expressing meanings, but as a formal mechanism for generating an infinite set of syntactically correct structures. These can later be "voiced" and "interpreted" with the help of a semantic (or formal-logical) component. In this regard, Mel'čuk's position has not changed since his earliest works: if, for his generative opponents, language is primarily a "syntactic machine", the products of which can later be assigned meaning, then in his conception, language has always been seen as a "translating machine", i.e. a system of multi-level transformations that ensure the transition from meaning to text. This brings MTT closer to what are commonly known as "functional" language models today, although it differs from these latter in a number of respects.

by way of deep and surface syntactic structures (Mel'čuk 2021: 12–13). The same goes for morphology. The structural approach makes the theory applicable for machine translation and text generation.

This model is functional: it is based on MTT, which serves as a framework for functional models of languages with its use of various formal systems: semantic networks, syntactic trees, lexical functions, paraphrasing rules, and so on. The strength of functional models lies in their ability to model an object that is hidden from direct observation. Such hidden properties of language may only be discerned by rigorous analytical processes. Functional models represent the behavior of the object under analysis, which accounts for their usefulness in understanding the inner workings of the brain, appropriately described as a "black box" (Mel'čuk 2012a: 18). In fact, a functional model reveals correspondences between a set of meanings and a set of texts, accounting for the way a speaker expresses meanings via texts and, conversely, extracts meanings from texts (Mel'čuk & Milićević 2020: 8).

This model is global, for it works at all levels of language as a system, showing how language functions from meaning to text. The model's aim is to make it possible to compute correct expressions for any meaning in English, and vice versa; to establish links not only between linguistic units and their meanings but also between English sentences and their meanings (Mel'čuk & Milićević 2020: 5).

Finally, this approach is synthesis oriented. Though the meaning-text correspondence is bi-directional, Mel'čuk claims that linguistics should study and describe language functioning in the Meaning-to-Text direction, as natural language is mostly about speaking rather than understanding (Mel'čuk 2021: 8–9). Linguistic synthesis, or text production, is more important for linguistics than analysis, or text understanding. This is because text comprehension involves, to a degree, the understanding of the subject matter treated, i.e. it presupposes the understanding of extralinguistic reality, a factor that may distort the whole picture. Mel'čuk posits that "going from a given meaning to all the texts that can express it liberates the researcher from huge "non-linguistic" difficulties leaving him face-to-face with the language in its purest state" (Mel'čuk 2021: 9–10). Another argument in favor of the priority of synthesis over analysis is the idea that before any text is analyzed it should be synthesized, which accords with the idea that text synthesis is primary to analysis

2.2. Meaning ⇔ Text Theory: major guidelines to linguistic semantics

Mel'čuk affirms that meaning and its description will be a considerable task for many generations of linguists (Mel'čuk 1999, 2012, Mel'čuk & Milićević 2020). The Meaning ⇔ Text theory attempted to present his own contribution to the solution to the problem. It offers a system of notions to describe lexical and syntactic semantics. Starting with a general understanding of meaning as "informational content that can be verbalized in the given language" (Mel'čuk & Milićević 2020: 4), Mel'čuk introduces a whole range of notions which have since

been used in linguistic semantics to reveal the meaning of a language unit. Such notions as valency, actants, diatese, sentential semantics were born within this theory and have been successfully used by its followers. The correlation of all these terms constitutes an integrated system aimed at the description of linguistic semantics.

Another essential detail for modern semanticists is Mel'čuk's treatment of meaning. He sees a difference between linguistic (= "shallow") non-pragmatic, non-extralinguistic, non-encyclopedic meaning and real (= "deep") meaning (Mel'čuk & Milićević 2020: 73). To explain these contrasts, Mel'čuk analyzes linguistic meaning as including propositional, communicative and rhetorical components (Mel'čuk & Milićević 2020: 77). Propositional meaning is the semantic content of a language unit and can be described by means of logical propositions; communicative meaning deals with the speaker's communicative intentions, while rhetorical meaning involves the speaker's stylistic intentions; whether the utterance is neutral, formal, colloquial, poetic, ironic, etc. (ibid.).

Mel'čuk is consistent in his claim that meaning is atomically built and extremely well-structured. Its description makes use of logical atoms, which are predicates and names of objects (Mel'čuk 1999: 62). This approach was used in Lexicography, and the "Explanatory Combinatorial Dictionary of Contemporary Russian" (Mel'čuk & Zholkovsky 1984) serves as a vivid example of its application.

2.3. Meaning ⇔ Text Theory: heuristics and application

The Meaning ⇔ Text Theory paved the way for various avenues of linguistic research. Within MTT an artificial semantic language was elaborated (Mel'čuk 1974) which enriched lexicography as well as provided an integral description of language. This artificial semantic language is abstract, since it was designed for a machine which has no knowledge of what any linguistic expression might mean. Thus, the formal language is strict, representing the structure of meaning. Another outcome of MTT for computational linguistics and machine translation is its building of an inventory of basic semes.

Meaning ⇔ Text Theory serves as a lens for other issues which Mel'čuk studies, and it found application in many areas. When Mel'čuk describes the range of problems that semantic studies border on and affect, he actually outlines the frontiers of his own theory and its impact on various fields of linguistic research. A renewed approach to lexical semantics, as well as producing findings about the semantic and syntactic aspects of linguistic units, contributed to the development of lexicology and lexicography as its applied counterparts. This culminated, on the one hand, in the expansion of modern semantics, which enriched natural language processing (e.g., machine translation, automatic text generation, mathematical linguistics), and on the other, led to advances in the field of language learning and teaching, as in both spheres the main focus is the transference of meaning (Mel'čuk & Milićević 2020: 7).

Another of Mel'čuk's influential contributions concerns his idea of a language model and its application, not only in linguistics but also in interdisciplinary research. He himself sees a high practical utility of his model in at least three technological and social domains: natural language processing; teaching and learning languages, and manufacturing reference books such as dictionaries, pedagogical grammar books and manuals (Mel'čuk 2021: 28).

To sum up, it is useful to cite Igor Mel'čuk's description of the perspectives of his theory, when he affirms that "the only reliable key to human thinking, in all its complexity, is natural language. Without a profound understanding of how language is functioning in our psyche, there will be no good understanding of information processing by the human brain. That is why functional models of language, and MTM (Meaning-Text Model) in particular, have nowadays acquired a special significance. Linguistics must take a place of honour among the "hard" sciences, and functional models, which embody the typical scientific approach to complex phenomena, will make their contribution" (Mel'čuk 2021: 28). This theory still serves as a framework and a lens for many other issues which Mel'čuk studies.

3. Outline of contributions to the issue

The issue is opened by Igor Mel'čuk. His article "Russian reduplicative surface-syntactic relations in the perspective of general syntax" is another gift to our readers (see Mel'čuk 2018a), and is part of a larger project aimed at creating an inventory of superficial-syntactic relations in the world's languages, started 60 years ago (Mel'čuk 1962: 47–49). Since then, several lists of reduplicative surface-syntactic relations have been compiled in Russian, English, French, German, and other languages (e.g., Mel'čuk 2015: 444–453, 2016: 184–194, 2018b, etc.). The purpose of this article is to supplement the existing lists and identify special Russian reduplicative superficial syntactic relations that appear exclusively in syntactic idioms based on reduplications. As a result of the study carried out within the framework of the "Meaning ⇔Text" Theory, 7 types of surface-syntactic relations in the Russian language were identified, which belong to the field of phraseology and have not yet received sufficient attention from scholars.

Anna Wierzbicka continues to develop the idea that, despite the phenomenal diversity of languages and cultures, which is often absolutized by researchers, there are basic elements of human thinking that are present as separate words (or signs) in all human languages (see e.g. Goddard & Wierzbicka 1994, Wierzbicka 1992, 2020, 2021). The concepts they designate are equally perceived by people regardless of the language they speak. According to Leibniz, they represent "the alphabet of human thought" (see Wierzbicka 1972: 6). In this issue, she explores the universal concepts of "I" and "you". She uses an approach to the study of meaning, developed on the basis of "semantic primitives", known under the acronym NSM (Natural Semantic Metalanguage) (see Goddard & Wierzbicka 1994, Wierzbicka 2021 and others). She argues that "YOU" and "I" ("I" and

"THOU") are fundamental elements of human thought, present as distinct words (or signs) in all human languages.

Wierzbicka's ideas of a universal metalanguage have been successfully tested by a number of scholars in typologically different languages (see Gladkova & Larina 2018) and have proven to be effective in explicating culture-specific words and concepts that are difficult to translate (e.g. Wierzbicka 1997, 1999, 2014, Levisen & Waters 2017, Bromhead & Ye 2020). In this issue, Mark Durie uses the NSM to explicate four Arabic words that name key Islamic concepts without equivalents in English. Emphasizing the importance of understanding the meaning of keywords when studying another culture, the author notes that they are inevitably distorted during translation, as they fall under the influence of another language. The study shows the effectiveness of NSM in explicating Arabic key words and overcoming their resistance to translation into English, as well as their role in providing an adequate understanding of the key concepts of another culture, including religious ones.

Natural semantic metalanguage (NSM) is also used as a method of semantic analysis by Anna Gladkova who considers changes in the modern Russian lexicon, namely, the development of new meanings for the words *zhest'* (from the primary meaning of *zhest'* 'tin' as a type of metal), *zhestkij* 'hard/tough/firm' and *zhestko* 'firmly/toughly'. She proposes NSM explications of these words, identifies connection of their meanings with the cultural themes of 'emotionality', 'not being in control' and 'strait talk', and notes the increased salience of these words in modern Russian.

If the semantic metalanguage refers to the study of the internal processes of language, to what is often hidden from direct observation or internalized, the studies of other levels of the language system, also inspired by Mel'čuk's ideas, refer to the units on the surface of language use, i.e., exteriorized. Thus, Nicholas Evans' study concerns interactive phonetics, which can be further integrated into the sound system. The latter either expands the phonological system by incorporating new phonemes into itself, or expands the phonotactic possibilities of phonemes occurring in other phonotactic positions. In support of his argument the author draws on phonemes restricted to interactive contexts in the Papuan language Nen, as well as phonemes with restricted phonotactic distributions (English) or morphological distributions (Bininj Kunwok, an Australian language of the Gunwinyguan family). He proposes aims to expand our understanding of this kind of phoneme by considering specific vocal gestures involved in interactional settings.

A wide range of complex issues related to the field of morphology, semantics, pragmatics and typology is considered in the article "Perfective, performative and present: Some non-standard combinations in Slavic and beyond" (Plungian, Rakhilina & Reznikova). Based on representative empirical material from the Russian National Corpus, which includes the parallel corpora containing examples from literary texts with their translations into other European languages, the authors

explore a special type of context in which negative perfective forms acquire the meaning of "prospective present". The authors argue that in some cases, these constructions correspond to speech acts and can be considered as a kind of performative. Such quasi-performative constructions expand our understanding of the meanings that present perfective forms can acquire.

Several articles in this issue of the journal present the results of research carried out at the intersection of semantics, pragmatics and lexicography. Barrios Rodriguez analyzes the difficulties of lexicographic processing of formulemes that arose when compiling the Spanish electronic dictionary *Diretes*. Noting that within the framework of the "Meaning ⇔ Text" theory, pragmatems have been formalized on the basis of lexical functions, the author argues the need for their formalization on the basis of pragmatic functions. Drawing on the results of the analysis of 200 formulemes, she concludes that pragmatic functions can be useful not only for formalizing formulemes and pragmatems in other languages, but also for revising the list of illocutionary verbs, from the perspective of Phraseology and Lexicography.

Olga Solopova and Tamara Khomutova present the first results of a larger project of an explanatory combinatorial dictionary of English conflict lexis. It is based on the corpus of texts on political discourse, and is generally consistent with Mel'čuk's Meaning ⇔ Text theory. The article discusses the principles for creating such a dictionary, and provides an example of a dictionary entry whose microstructure includes semantic, phonological and co-occurrence zones.

Valentina Apresjan and Boris Iomdin explore Russian interrogative-relative pronouns (wh-words) as a lexicographic type which requires a unified description. They draw on the principles of the Moscow School of Semantics (see Apresjan 2005, 2009, Apresjan, Sannikov & Boguslavskii 2010, Apresjan, Iomdin & Boguslavskii 2012, Boguslavskii 1996, among many others), which largely continue and develop the ideas and principles of Mel'čuk's "Meaning ⇔ Text" Theory. They include a multi-level description of language, a focus on semantic and combinatorial properties of linguistic items, and acknowledge the validity of the dictionary as a tool for linguistic research. The study focuses on semantic, syntactic and co-occurrence properties of a number of Russian interrogative pronouns. The results suggest a deep motivating connection between the semantic properties of interrogative pronouns and their combinatorial abilities.

The category of Russian aspect is the focus of Solovyev, Bochkarev & Bayrasheva, which is not surprising as Mel'čuk has made significant contributions to the formalization and understanding of this complex grammatical category. Drawing on some of his ideas, the authors continue the discussion of whether verbs in aspectual pairs are in inflectional or derivational relations. They propose two new approaches that compare the quantitative characteristics of aspectual pairs of Russian verbs of both types. The results of the study suggest that the suffixal and affixal methods of forming aspectual pairs have an equal status in terms of their inflectional or derivational classification.

4. Conclusion

In this introductory article we aimed to summarize the main tenets of "Meaning ⇔ Text" theory founded by Mel'čuk and outline its contribution to the development of diverse areas of modern linguistics. We must admit that this is a very brief overview which cannot cover the linguistic universe of Mel'čuk. The articles in this issue confirm that the principles of the "Meaning ⇔ Text" Theory, which focuses primarily on language synthesis, speech production and the understanding of meanings that are encoded in text, have been further developed by Mel'čuk and his followers, and continue to be used in all types of linguistic research.

No doubt, the theory will inspire new generations of linguists and acquire new followers, in their eternal quest for the essence of language and the workings of the minds that use it. It is our hope that, after becoming familiar with the articles published in this issue, our readers will turn to the books and articles written by Igor Mel'čuk and enjoy the harmony and depth of his ideas.

Once again, we conclude with wishing Igor many more years of joy, happiness and creativity!

RU

1. Введение

Этот выпуск посвящен всемирно известному лингвисту — Игорю Александровичу Мельчуку, создателю лингвистической теории «Смысл ⇔ Текст» и одному из пионеров машинного перевода. Научная жизнь Игоря Александровича впечатляет и восхищает своей результативностью и масштабностью. 20 лет он проработал в Институте языкознания Российской академии наук, а затем еще 32 года в Монреальском университете, Канада. Все последующие годы и до сих пор он продолжает активно трудиться, являясь Почетным доктором Института языкознания РАН, Заслуженным профессором Монреальского университета и членом Королевского общества Канады. Он читает лекции, консультирует молодых ученых, пишет статьи и книги. На данный момент на его счету 305 научных статей и 49 книг, работа над следующей продолжается. Таким образом, вместе с замечательным личным юбилеем Игорь Александрович Мельчук отмечает свой профессиональный юбилей – 65 лет непрерывного научного поиска и научного творчества.

Своими новаторскими идеями и преданностью науке он вдохновлял и продолжает вдохновлять коллег, начинающих ученых и студентов как в России, так и за ее пределами, поскольку его лингвистическая вселенная не знает границ, а его прекрасные человеческие качества – доброта, открытость, искрометный юмор – завораживают каждого, кому посчастливилось с ним общаться.

В этот выпуск мы пригласили давних друзей Игоря Александровича, с которыми он начинал, а затем продолжал свой творческий путь, коллег

и учеников, а также почитателей и ценителей его научного творчества. Вместе с ними мы преподносим ему этот интеллектуальный и эмоциональный подарок. От имени редколлегии журнала, наших авторов и читателей мы от всей души поздравляем Игоря Александровича Мельчука с замечательным юбилеем и желаем ему доброго здоровья и новых творческих достижений!

В данном выпуске мы хотим обобщить вклад И.А. Мельчука в семантику, в разработку теории значения и методологию анализа текста, а также обозначить влияние его теории «Смысл ⇔ Текст» на дальнейшее развитие теории текста, теории и практики машинного перевода, методологии лингвистического анализа, лексикографии, типологии, компьютерной лингвистики и корпусной лингвистики. В нашей вводной статье мы коротко остановимся на основных идеях данной теории и наметим направления, которые она обозначила для многих сфер изучения языка как системы и языка в его функционировании.

2. Теория «Смысл ⇔ Текст»

2.1. Лингвистическая модель «Смысл ⇔ Текст»: отправные точки

Теория «Смысл ⇔ Текст» (ТСТ) была разработана И.А. Мельчуком около 60 лет назад и стала визитной карточкой для него как исследователя. Данная теория объясняет, как говорящий порождает смысл, воплощенный в тексте, который трактуется как смысловое единство. Хотя теория появилась в конце 1960-х гг., она не утратила своей актуальности и по сей день. Со времени появления ее постулаты, методология и предложенные в ней понятия используются в трудах ученых самых разных лингвистических школ и направлений. Несомненно то, что достижения ТСТ повлияли на развитие различных областей лингвистических исследований, а ее результаты и принятый подход применяются как в теории, так и на практике.

С самого первого дня лингвистика искала ответы на свои главные вопросы: как язык передает значение и что в языке отвечает за передачу значения. Чтобы ответить на эти вопросы, с одной стороны, необходимо определить, что такое язык как сущность и как он устроен. С другой стороны, крайне важно учитывать способ передачи значения, поскольку коммуникация — это, прежде всего, передача значения. Теория И.А. Мельчука ставит оба эти вопроса и строит целостный подход, объединяющий структуру языковой единицы и значение, которое эта единица передает. ТСТ связывает обе стороны знака воедино и дает ответ на вопрос о том, как язык порождает смысл через тексты.

Теория «Смысл ⇔ Текст» имеет дело с естественным языком, который понимается как логическое устройство (Mel'čuk 2021: 10), т.е. устройство, спроектированное по законам логики [2]. Это «устройство» (или машина)

---

[2] Сейчас мы можем предположить, что идея устройства или машины могла появиться под влиянием вдохновляющих идей 60-х годов: глубокий и неподдельный интерес к кибернетике,

обеспечивает переход от смысла, который хочет передать говорящий, к тексту, который служит носителем смысла. Хотя упоминается, что устройство работает в обоих направлениях — от смысла к тексту со стороны говорящего и в обратном направлении, когда речь идет о слушателе, ТСТ отдает приоритет генерации текста (другими словами, говорению) по сравнению с интерпретацией текста (т.е. пониманием), что находит свое отражение в названии теории: она именуется «Смысл ⇔ Текст», и несмотря на особый знак двустороннего перехода, смысл в названии всё-таки стоит первым.

Еще одна идея, которая была популярна в то время, когда зародилась ТСТ, — это идея моделирования объектов, позволяющая понять принцип функционирования исследуемого объекта. Идея модели была транспонирована на естественный язык, выступающий как набор правил, закодированных в сознании его носителя, которые устанавливают соответствие между значениями языка и их выражением, то есть текстами (Mel'čuk & Milićević 2020: 4). Исходя из этого, данная система правил может служить моделью языка.

Эта модель формализована. Соответственно, все правила представлены в формализованном виде. Они декодируют процесс экстериоризации зависимостей глубинного уровня между языковыми единицами посредством деревьев зависимостей и установления лексических функций. Правила должны соответствовать определенным требованиям: они должны быть логически непротиворечивыми, однозначными и прозрачными в плане передачи содержания (Gladky & Mel'čuk 1969: 9), а также «экономными и изящными по форме»[3] (Mel'čuk & Milićević 2020: 5). Так, утверждается, что «языковые значения (в техническом смысле этого термина) предстают как формальные символические объекты, называемые семантическими представлениями [SemR], а тексты — как фонетические представления [PhonR]», что может быть выражено формулой $\{SemR_i\} \Leftarrow$ язык $\Rightarrow \{PhonR_j\} \mid i \neq j, 0 < i, j \leq \infty$ (Mel'čuk 2021: 9). В целом используется несколько видов формализации: «семантические сети для представления значений предложений и лексических единиц; деревья зависимостей для представления синтаксической структуры предложений; лексические функции для представления лексических отношений и разного рода правила представления семантических операций» (Mel'čuk & Milićević: xvii).

Модель учитывает семантический компонент, а значит, она семантически ориентирована. Семантика не является «просто еще одним компонентом» языковой системы: она занимает в ней особое место, поскольку язык — это прежде всего средство коммуникации, то есть средство передачи смысла

пронизывающий все области научных исследований, искренняя вера в формализацию и в силу машинного интеллекта.

[3] Здесь и далее перевод наш. – *С.И.* и *Т.Л.*

(Mel'čuk & Milićević: 6)[4]. Таким образом, семантический компонент составляет ядро языкового описания, поскольку семантические свойства языковых выражений во многом предопределяют их синтаксическое поведение, а также влияют на их морфологию (Mel'čuk & Milićević 2020: 6). Следовательно, эта модель устанавливает тесную связь между лексикой и грамматикой. Кроме того, необходимо подчеркнуть, что согласно теории «Смысл ⇔ Текст» сама семантика рассматривается как система строгих формальных правил, поскольку язык представлен как набор правил, устанавливающих соответствия между значениями и их возможными выражениями. Еще одна идея семантической модели заключается в возможности декомпозиции смыслов, которая предполагает, что сложные значения могут быть представлены через более простые значения (ср. исходящую из близких предпосылок теорию «семантических примитивов» Анны Вежбицкой [Wierzbicka 2021], см. также [Mel'čuk 1989]). И.А. Мельчук утверждает, что семантическая репрезентация базируется на ограниченном числе элементарных сем, которые могут быть названы «атомами смысла» (Mel'čuk 1999: 58).

Помимо всего перечисленного, эта модель является многоуровневой или, говоря словами И.А. Мельчука, стратификационной (= многослойной) (Mel'čuk & Milićević 2020: 10), поскольку язык понимается как многоуровневая сущность. Автор теории выделяет иерархию уровней, включающую семантический (смысловой), синтаксический, морфологический, фонологический и фонетико-графический (текст). На каждом основном репрезентативном уровне, кроме семантического, смысло-текстовая модель распознает два подуровня, называемых глубинным и поверхностным. Таким образом, можно говорить о глубинно-синтаксическом представлении и о поверхностно-синтаксическом представлении и т.д. И.А. Мельчук дает следующее объяснение этим подуровням: «Глубинный подуровень ориентирован на смысл, то есть на содержание, которое Говорящий хочет выразить; его задача состоит в том, чтобы ясно передать соответствующие различия информации. Поверхностный подуровень ориентирован на текст, то есть на форму, в которой выражается содержание; его задача – эксплицитно отразить все соответствующие формальные различия» (Mel'čuk & Milićević 2020: 16).

---

[4] Это утверждение на первый взгляд может показаться очевидным, но напомним, что одна из наиболее распространенных сегодня теорий языка, генеративная грамматика (и близкие к ней идеологии), рассматривает язык принципиально иначе: не как инструмент для выражения смыслов, а как формальный механизм порождения бесконечного множества синтаксически правильных структур, которые в дальнейшем могут быть «озвучены» и «интерпретированы» с помощью сематического (или формально-логического) компонента. В этом отношении позиция И.А.Мельчука не менялась с самых ранних работ: если для его генеративных оппонентов язык – это в первую очередь «синтаксическая машина», продуктам работы которой можно впоследствии приписать значение, то в его идеологии язык всегда рассматривался как «транслирующая машина», т.е. система многоуровневых преобразований, обеспечивающих переход от смысла к тексту. Это сближает ТСТ с тем, что сегодня принято называть «функциональными» моделями языка (хотя и от этих последних она в ряде отношений отличается).

Эта модель также является структурной, поскольку сам язык и его единицы понимаются как структуры. Таким образом, семантика описывается в терминах семантических структур, которые представлены в виде сетей с узлами. Предложения и их семантика представлены в виде глубинных и поверхностных синтаксических структур (Mel'čuk 2021: 12–13). То же самое относится к морфологии. Структурный подход придает данной теории прикладной характер: ее можно использовать в практике машинного перевода и генерации текста.

Эта модель функциональна – в ее основе лежит ТСТ, на основе которой и строятся функциональные модели языков с использованием различных формальных систем: семантических сетей, синтаксических деревьев, лексических функций, правил перефразирования и т.д. Сильной стороной функциональных моделей является возможность моделировать объект, скрытый от прямого наблюдения. И человеческий разум, и язык в целом нуждаются в специальных исследовательских процедурах и методиках, чтобы обнаружить их скрытые свойства. Функциональные модели представляют анализируемый объект в его функционировании, поэтому их можно использовать для понимания внутренней работы мозга, который до сих пор сохраняет статус «черного ящика» (Mel'čuk 2012a: 18). По сути, функциональная модель выявляет соответствия между набором значений и набором текстов и раскрывает то, как говорящий выражает значения с помощью текстов и, наоборот, извлекает значения из текстов (Mel'čuk & Milićević 2020: 8).

Эту модель можно по праву охарактеризовать как глобальную, поскольку она работает на всех уровнях языковой системы. Она показывает, как функционирует язык по направлению от значения к тексту и устанавливает связи не только между лексическими единицами и их значениями, но и между предложениями и их значениями.

Наконец, этот подход ориентирован на синтез. Хотя смысловой текст является двунаправленным, И.А. Мельчук утверждает, что лингвистика должна изучать и описывать языковое функционирование в направлении «Смысл – Текст» в соответствие с этапами порождения текста (Mel'čuk 2021: 8–9). Лингвистический синтез, или, другими словами, производство текста важнее для языкознания, чем анализ или понимание текста. Это можно объяснить тем, что понимание текста предполагает в определенной степени понимание предмета, что включает понимание внеязыковой действительности, фактора, который может исказить всю картину. Если лингвисты идут от значения ко всем выражающим его текстам, они игнорируют внеязыковые факторы и исследуют язык как таковой. Кроме того, любой текст перед тем, как проводится анализ, должен быть синтезирован, а это является еще одним доказательством того, что синтез текста первичен по отношению к анализу (Mel'čuk 2021: 9–10).

2.2. Теория «Смысл ⇔ Текст»: основные ориентиры в описании лингвистической семантики

И.А. Мельчук подчеркивает, что как само значение языковой единицы, так и его описание являются сложной задачей для многих поколений лингвистов (Mel'čuk 1999, 2012, Mel'čuk & Milićević 2020). Теория «Смысл ⇔ Текст» предлагает ее решение. В рамках разработанной теории предложена система понятий для описания лексической и синтаксической семантики. Начиная с общего понимания значения как «информационного содержания, которое может быть вербализировано на данном языке» (Mel'čuk & Milićević 2020: 4), И.А. Мельчук вводит целый ряд понятий, которые широко используются в лингвистической семантике для выявления значения языковой единицы. К этим понятиям относятся валентность, актанты, диатеза, сентенциальная семантика. Соотношение всех этих терминов составляет целостную систему, направленную на описание языковой семантики.

Еще один момент, важный для современных семасиологов, – это трактовка значения в работах И.А. Мельчука. Он различает, с одной стороны, лингвистическое (= «поверхностное»), то есть непрагматическое, неэнциклопедическое значение и, с другой стороны, реальное (= «глубинное») значение (Mel'čuk & Milićević 2020: 73). Вместе с тем чтобы компенсировать эти ограничения на структуру значения и далее – смысла И.А. Мельчук анализирует языковое значение как включающее в себя пропозициональный, коммуникативный и риторический компоненты (Mel'čuk & Milićević 2020: 77). Пропозициональное значение представляет собой смысловое содержание языковой единицы и может быть описано с помощью логических пресуппозиций; коммуникативное значение связано с коммуникативными намерениями говорящего, а риторическое значение – со стилистическими намерениями говорящего: является ли высказывание нейтральным, формальным, разговорным, поэтическим, ироническим и т.д. (Там же).

И.А. Мельчук последователен в утверждении, что значение построено атомарно, очень хорошо структурировано и его описание использует логические атомы, которые являются предикатами и именами объектов (Mel'čuk 1999: 62). Этот подход был перенесен в лексикографию, и «Толковый комбинаторный словарь современного русского языка» (Mel'čuk & Zholkovsky 1984) может служить ярким примером его применения.

2.3. Теория «Смысл ⇔ Текст»: эвристика и прикладные аспекты

Теория «Смысл ⇔ Текст» предопределила развитие разных направлений лингвистических исследований. В рамках ТСТ был разработан искусственный семантический язык (Мельчук 1974), обогативший как лексикографию, так и описание языка в целом. Этот искусственный семантический язык носит абстрактный характер, поскольку рассчитан на машину, которая не может угадать, что означает то или иное выражение. Предлагаемый формальный

язык отличается строгостью и нацелен на репрезентацию структуры значения. Еще одним практическим результатом применения ТСТ в компьютерной лингвистике и машинном переводе является создание списка основных сем.

Теория «Смысл ⇔ Текст» наметила пути решения тех проблем, которые находятся как в поле внимания И.А. Мельчука, так и стоят на повестке дня других исследователей. Когда И.А. Мельчук описывает и изучает широкий круг проблем, находящихся на границе семантики и других областей языкознания, а также смежных дисциплин, он фактически очерчивает контуры своей собственной теории и ее потенциальные возможности. Новый подход к лексической семантике, а также предложенное описание семантических и синтаксических аспектов языковых единиц способствовали развитию лексикологии и лексикографии как прикладных аналогов лексической семантики. Все это привело к расширению современной семасиологии, которая уточнила процедуры обработки естественного языка (например, машинный перевод, автоматическая генерация текста, математическая лингвистика), а с другой стороны, внесла вклад в изучение и преподавание языка, поскольку в обеих сферах основной задачей является передача значения (Mel'čuk & Milićević 2020: 7).

Еще одним важным вектором, который наметил И.А. Мельчук, является возможность применения языковых моделей не только для лингвистических, но и междисциплинарных исследований. Сам он видит высокую практическую ценность своей модели, по крайней мере, в трех технологических и социальных областях: автоматическая обработка естественного языка; преподавание и изучение языков; создание словарей, учебников и пособий (Mel'čuk 2021: 28).

Нельзя не присоединиться к позиции И.А. Мельчука, когда он, резюмируя перспективы своей теории, говорит, что естественный язык является единственным надежным ключом к человеческому мышлению (Mel'čuk 2021: 28). Без глубокого понимания того, как язык функционирует в нашей психике, невозможно понять, как человеческий мозг обрабатывает информацию. Вот почему функциональные модели языка и, в частности, ТСТ (теория «Смысл ⇔ Текст») в настоящее время имеют совершенно особое значение. Как отмечает И.А. Мельчук, «лингвистика должна занять почетное место среди «точных» наук, и свой вклад внесут функциональные модели, воплощающие естественнонаучный подход к сложным явлениям» (Там же). В завершение можно со всей уверенностью сказать, что теория «Смысл ⇔ Текст» служит основой для решения многих сущностных вопросов, к которым обращается И.А. Мельчук и его многочисленые последователи.

3. Краткий обзор статей выпуска

Открывает выпуск сам юбиляр. Статья И.А. Мельчука «Русские редупликативные поверхностно-синтаксические отношения в аспекте общего синтаксиса» – еще один подарок нашим читателям (см. Mel'čuk 2018a). Данное

исследование является частью большого проекта, нацеленного на описание поверхностно-синтаксических отношений в языках мира, начало которому было положено еще в (Mel'čuk 1962: 47–49). С тех пор было составлено несколько списков редупликативных поверхностно-синтаксических отношений в русском, английском, французском, немецком и других языках (см., например, Mel'čuk 2015: 444–453, 2016: 184–194, 2018b и др.). Цель данной статьи – дополнить имеющиеся списки и определить специальные русские редупликативные поверхностно-синтаксические отношения, которые выступают исключительно в синтаксических идиомах, основанных на редупликациях. В результате исследования, проведенного в рамках теории «Смысл ⇔ Текст», выделено 7 типов поверхностно-синтаксических отношений в русском языке, которые относятся к области фразеологии и до сих пор не получили должного внимания исследователей.

Анна Вежбицкая продолжает развивать идею о том, что, несмотря на феноменальное разнообразие языков и культур, которое часто абсолютизируется исследователями, существуют базовые элементы человеческого мышления, присутствующие как отдельные слова (или знаки) во всех языках мира (Goddard & Wierzbicka 1994, Wierzbicka 1992, 2020, 2021 и др.). Обозначаемые ими понятия одинаково воспринимаются людьми независимо от языка, на котором они говорят. По словам Лейбница, они представляют собой «алфавит человеческой мысли» ("the alphabet of human thought") (Wierzbicka 1972: 6). Используя подход к изучению смысла, разработанный на основе «семантических примитивов» и известный под аббревиатурой NSM (Natural Semantic Metalanguage), в русском переводе – ЕСМ (Естественный Семантический Метаязык) (Goddard & Wierzbicka 1994, Wierzbicka 2021 и др.), А. Вежбицкая рассматривает универсальные концепты «Я» и «Ты». Результаты проведенного исследования позволяют утверждать, что они являются базовыми элементами человеческого мышления и присутствуют во всех языках, что в очередной раз свидетельствует о фундаментальном единстве человечества.

Идеи Анны Вежбицкой об универсальном метаязыке были успешно апробированы рядом ученых на типологически различных языках (см. Gladkova & Larina 2018) и доказали свою эффективность при толковании культурно-специфичных слов и понятий, являющихся весьма трудными для перевода (Wierzbicka 1997, 1999, 2014, Goddard & Wierzbicka 1994, Levisen & Waters 2017, Bromhead & Ye 2020 и др.). В данном выпуске Марк Дьюри (Duri 2022) использует метод NSM (ЕСМ) для экспликации четырех арабских слов, называющих ключевые понятия ислама, которые не имеют эквивалентных соответствий в английском языке. Подчеркивая важность понимания ключевых слов при изучении другой культуры, автор отмечает, что при переводе они неизбежно искажаются, поскольку попадают под влияние другого языка. В статье показана эффективность семантического анализа культурно-специфичных арабских слов с применением ЕСМ, который позволяет

устранить эти искажения и обеспечить адекватное понимание ключевых понятий другой культуры, в том числе религиозных.

Естественный семантический метаязык (ЕСМ) использует в качестве метода семантического анализа и Анна Гладкова, которая рассматривает изменения в современном русском лексиконе, а именно – появление новых значений у слов *жесть, жесткий* и *жестко*. В своем исследовании она предлагает ЕСМ толкования данных слов, выявляет связь их значений с культурными темами «эмоциональность», «отсутствие контроля» и «откровенный разговор» и отмечает возросшую значимость этих слов в современном русском языке.

Если семантический метаязык относится к исследованию внутренних процессов языка – того, что часто скрыто от прямого наблюдения, иными словами, интериоризовано, то исследования других уровней языковой системы, которые также отталкиваются в определенном роде от идей И.А. Мельчука, обращаются к структурам, служащим экстериоризации языковых единиц в речи. Так, Николас Эванс (Evans 2022) предлагает идею интерактивной фонетики, которая может в дальнейшем интегрироваться в звуковую систему и либо расширять фонологическую систему за счет включения в нее новых фонем, либо расширять фонотактические возможности фонем, встречающихся в других фонотактических позициях. При этом в качестве аргумента автор указывает на ограниченные интерактивными контекстами фонемы в папуасском языке нэн, а также на фонемы с ограниченной фонотактической дистрибуцией (английский) или морфологической дистрибуцией (бининдж кунвок, австралийский язык гунвинигуанской семьи). Цель автора – расширить наши представления о фонемах с ограниченной дистрибуцией посредством привлечения специфических вокальных жестов, возникающих в условиях интеракции.

Широкий круг важных вопросов, относящихся к области морфологии, семантики, прагматики и типологии, рассматривается в статье «Перфектив, перформатив и настоящее время: некоторые нестандартные комбинации в славянских и других языках» (В.А. Плунгян, Е.В. Рахилина, Т.И. Резникова). На репрезентативном эмпирическом материале Национального корпуса русского языка (НКРЯ), включающем примеры из художественных текстов с их переводами на другие европейские языки, авторы рассматривают особый тип контекстов, в которых формы перфективов с отрицанием получают значение «проспективного настоящего». По мнению авторов, в некоторых случаях эти конструкции соответствуют речевым актам и могут считаться своего рода перформативами. Как показывает исследование, подобные квази-перформативные конструкции расширяют наши представления о значениях, которые могут приобретать перфективные формы настоящего времени.

Несколько статей выпуска представляют результаты исследований, выполненных на пересечении семантики, прагматики и лексикографии. Барриос Родригес (Barrios Rodriguez 2022) анализирует трудности лексикографической обработки формулем, возникшие при составлении испанского электронного словаря Diretes. Отмечая, что на сегодняшний день в рамках

теории «Смысл ⇔ Текст» прагматемы подвергались формализации на основе лексических функций, автор доказывает необходимость их формализации на основе прагматических функций. Результаты анализа 200 формулем позволили сделать вывод о том, что прагматические функции могут быть полезны не только для формализации формулем и прагматем в других языках, но и для пересмотра списка иллокутивных глаголов, что важно для фразеологии и лексикографии.

О.А. Солопова и Т.Н. Хомутова представляют первые результаты большого проекта по созданию англоязычного толково-комбинаторного словаря конфликтогенной лексики на материале корпуса текстов политического дискурса. В качестве методологических принципов построения словаря использованы положения теории «Смысл ⇔ Текст» И.А. Мельчука. В статье обсуждаются разработанные авторами принципы создания словаря, а также приведен пример словарной статьи, микроструктура которой включает семантическую, фонологическую и сочетаемостную зоны.

Опираясь на принципы Московской семантической школы (см. Apresjan 2005, 2009, Apresjan, Sannikov & Boguslavskii 2010, Apresjan, Iomdin & Boguslavskii 2012, Boguslavskii 1996 и др.), которые во многом продолжают и развивают идеи и принципы теории «Смысл ⇔ Текст», В.Ю. Апресян и Б.Л. Иомдин рассматривают русские вопросительно-относительные местоимения как лексикографический тип, который требует единого словарного описания. Авторы выделяют важнейшие принципы теории «Смысл ⇔ Текст», которые были унаследованы Московской семантической школой (МСШ), – это «установка на многоуровневое языковое описание, внимание к семантическим и комбинаторным свойствам языковых единиц, ценность словаря как инструмента лингвистического исследования» (Apresjan & Iomdin 2022). В работе описываются общие семантические, синтаксические и сочетаемостные свойства многих русских вопросительных местоимений. Результаты исследования свидетельствуют о наличии глубокой мотивирующей связи между семантическими свойствами вопросительных местоимений и их комбинаторными способностями

Не обошли вниманием авторы выпуска и русский вид, значительный вклад в формализацию и понимание природы которого внес И.А. Мельчук. Опираясь на изложенные им представления об этой сложной грамматической категории, В.Д. Соловьев, В.В. Бочкарев и В.Р. Байрашева продолжают дискуссию о том, находятся ли глаголы в аспектуальных парах в словоизменительных или словообразовательных отношениях. В статье предложены два новых подхода, сопоставляющих квантитативные характеристики аспектуальных пар русских глаголов обоих типов. На основе результатов проведенного исследования авторы приходят к выводу, что суффиксальный и аффиксальный способы образования аспектуальных пар имеют равный статус с точки зрения их классификации как словоизменительных или словообразовательных.

4. Заключение

В данной статье мы ограничились целью обобщить основные положения теории «Смысл ⇔ Текст», разработанной И.А. Мельчуком, и хотели показать ее вклад в развитие различных направлений современной лингвистики. Мы признаем, что это очень краткий обзор, в котором невозможно отразить всю лингвистическую вселенную И.А. Мельчука.

Как свидетельствуют статьи этого выпуска, принципы теории «Смысл ⇔ Текст», ориентированные, прежде всего, на языковой синтез, речеобразование и понимание тех смыслов, которые закодированы в тексте, получают дальнейшее развитие, уточняются, углубляются и используются в изучении всех уровней языка. Несомненно, теория «Смысл ⇔ Текст» будет и дальше вдохновлять исследователей в их вечных поисках сущности языка и того, как человеческий разум использует его. Мы надеемся, что, ознакомившись со статьями, опубликованными в этом выпуске, наши читатели обратятся к научному творчеству И.А. Мельчука и получат интеллектуальное удовольствие от стройности и глубины его идей.

От имени редколлегии журнала, всех наших авторов и читателей мы еще раз поздравляем Игоря Александровича с юбилеем и желаем доброго здоровья, неиссякаемого оптимизма и новых творческих достижений!

REFERENCES

Apresjan, Yurij D. 2005. *Pravila vzaimodeistviya znachenii i slovar'* [The rules of interaction of meanings and the dictionar]. *Russkii yazyk v nauchnom osveshchenii* 1 (9). 7–45. (In Russ.)

Apresjan, Yuriy D. 2009. *Issledovaniya po semantike i leksikografii*. V. 1. Paradigmatika [Studies in semantics and lexicography. Vol 1. Paradigmatics]. Moscow: Yazyki slavyanskikh kul'tur Publ. (In Russ.)

Apresjan, Yurii, Vladimir Sannikov & Igor' Boguslavskii. 2010. *Teoreticheskie problemy russkogo sintaksisa. Vzaimodeistvie grammatiki i slovarya* [Theoretical problems of Russian syntax. Interaction of grammar and vocabulary]. Moscow: Yazyki slavyanskikh kul'tur Publ. (In Russ.)

Apresjan, Yurii D., Leonid L. Iomdin & Igor' M. Boguslavskii. 2012. *Smysly, teksty i drugie zakhvatyvayushchie syuzhety* [Meanings, texts and other exciting stories]. Moscow: Yazyki slavyanskikh kul'tur. (In Russ.)

Apresjan, Valentina Yu. & Boris Iomdin. 2022. Russian interrogative pronouns as a lexicographic type. *Russian Journal of Linguistics* 26 (4). 1078–1013. https://doi.org/10.22363/2687-0088-30714 (In Russ.)

Barrios Rodríguez, María Auxiliadora. 2022. Meaning and function of Spanish formulemes and pragmatemes vs. illocutionary verbs. *Russian Journal of Linguistics* 26 (4). 1031–1049. https://doi.org/10.22363/2687-0088-31597

Boguslavskii, Igor' M. 1996. *Sfera deistviya leksicheskikh edinits* [Scope of Lexical Units]. Moscow: Yazyki russkoi kul'tury. (In Russ.)

Durie, Mark. 2022. Semantic decomposition of four Quranic words. *Russian Journal of Linguistics* 26 (4). 937–969. https://doi.org/10.22363/2687-0088-30779

Evans, Nicholas. 2022. Pushing the boundaries: Marginal phonemes and dialogic interaction. *Russian Journal of Linguistics* 26 (4). 995–1011. https://doi.org/10.22363/2687-0088-32349

Gladky, Alexey V. & Igor Mel'chuk. 1969. *Elementy matematicheskoi lingvistiki* [Elements of mathematical linguistics]. Moscow: Nauka. (In Russ.)

Bromhead, Helen & Zhengdao Ye (eds.). 2020. *Meaning, Life and Culture.* Canberra: ANU Press.

Gladkova, Anna & Tatiana Larina. 2018. Anna Wierzbicka, language, culture and communication. *Russian Journal of Linguistics* 22 (4). 717–748.

Goddard, Cliff & Anna Wierbicka (eds.). 1994. *Semantic and Lexical Universals: Theory and Empirical Findings.* Amsterdam: John Benjamins.

Levisen, Carstens & Sophia Waters (eds.). 2017. *Cultural Keywords in Discourse*. Amsterdam: John Benjamins.

Mel'čuk, Igor. 1962. Ob algoritme sintaksicheskogo analiza yazykovykh tekstov (obshchie printsipy i nekotorye itogi) [An algorithm for syntactic analysis of texts (general principles and some conclusions)]. *Mashinnyi perevod i prikladnaya lingvistika* 7. 45–87. (In Russ.)

Mel'čuk, Igor. 1974. *Opyt teorii lingvisticheskikh modelei "Smysl ⇔ Tekst". Semantika, sintaksis* [*Outline of a "Meaning ⇔ Text" Linguistic Model Theory. Semantics, Syntax*]. Moscow: Nauka. (In Russ.)

Mel'čuk, Igor. 1989. Semantic primitives from the viewpoint of the Meaning-Text linguistic theory. *Quaderni di Semantica* 10. 65–102.

Mel'chuk, Igor. 1999. *Opyt teorii lingvisticheskikh modeley "SMYSL ⇔ TEKST"* [Outline of a "Meaning ⇔ Text" Linguistic Model Theory]. Moskva: Shkola "Yazyki russkoy kul'tury". (In Russ.)

Mel'chuk Igor. 2012a. *Yazyk: ot smysla k tekstu* [Language: From meaning to text]. Moscow: Yazyki slavyanskoy kul'tury. (In Russ.)

Mel'čuk, Igor. 2012b. *Semantics: From Meaning to Text*. Vol. 1. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, Igor. 2013. *Semantics: From Meaning to Text*. Vol. 2. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, Igor. 2015. *Semantics: From Meaning to Text*. Vol. 3. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, Igor. 2016. *Language: From Meaning to Text*. Moscow / Boston: Academic Studies Press.

Mel'čuk, Igor. 2018a. Anna Wierzbicka, semantic decomposition, and the Meaning – Text Approach. *Russian Journal of Linguistics* 22 (3). 521–538. https://doi.org/10.22363/2312-9182-2018-22-3-521-538

Mel'čuk, Igor. 2018b. Genitive adnominal dependents in Russian: Surface-syntactic relations in the N→N$_{GEN}$ phrase. *Voprosy yazykoznaniya* 4. 25–46.

Mel'čuk, Igor 2021. *Ten Studies in Dependency Syntax.* Mouton de Gruyter.

Mel'čuk, Igor & Jasmina Milićević. 2020. *An Advanced Introduction to Semantics. A Meaning-Text Approach*. Cambridge: Cambridge University Press.

Mel'čuk, Igor A. & Zholkovsky Alexander K. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Vienna: Wiener Slawistischer Almanach.

Wierzbicka, Anna. 1972. *Semantic Primitives*. Frankfurt: Athenäum.

Wierzbicka, Anna. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1997. *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1999. *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge: CUP.

Wierzbicka, Anna. 2014. *Imprisoned in English: The Hazards of English as a Default Language*. Oxford: Oxford University Press.

Wierzbicka, Anna. 2020. Addressing God in European languages: Different meanings, different cultural attitudes. *Russian Journal of Linguistics* 24 (2). 259–293. https://doi.org/10.22363/2687-0088-2020-24-2-259-293

Wierzbicka, Anna. 2021. Semantic Primitives, fifty years later. *Russian Journal of Linguistics* 25 (2). 317–342. https://doi.org/10.22363/2687-0088-2021-25-2-317-342

Bionotes:
Svetlana IVANOVA is Dr Habil. in Philology, Professor, Head of the Department of Romance and Germanic Philology and Linguodidactics of Pushkin Leningrad State University. Her research interests include cultural linguistics, text linguistics, English grammar, discourse analysis, and media linguistics. She has over 240 publications in Russian and English including monographs, book chapters and articles in peer-reviewed journals.
e-mail: svet_victoria@mail.ru
https://orcid.org/0000-0002-0127-9934

Tatiana LARINA is Doctor Habil., Full Professor at RUDN University. Her research interests and publications deal with cross-cultural pragmatics, cultural linguistics. communicative ethnostyles, and (im)politeness theory. She has over 200 publications in Russian and English which comprise monographs, book chapters and articles in peer-reviewed journals, including *International Review of Pragmatics, Intercultural Pragmatics, Journal of Politeness Research,* among others.
e-mail: larina-tv@rudn.ru
https://orcid.org/0000-0001-6167-455X

Сведения об авторах:
Светлана Викторовна ИВАНОВА – доктор филологических наук, профессор, заведующий кафедрой романо-германской филологии и лингводидактики Ленинградского государственного университета имени А.С. Пушкина. Сфера ее научных интересов – лингвокультурология, лингвистика текста, английская грамматика, дискурс-анализ, медиалингвистика. Имеет более 240 публикаций на русском и английском языках, включая монографии, учебные пособия, главы в книгах и статьи в рецензируемых научных журналах.
e-mail: svet_victoria@mail.ru
https://orcid.org/0000-0002-0127-9934

Татьяна Викторовна ЛАРИНА – доктор филологических наук, профессор кафедры иностранных языков филологического факультета РУДН. Ее исследовательские интересы и публикации связаны с кросс-культурной прагматикой, коммуникативной этностилисикой и теорией (не)вежливости. Имеет более 200 публикаций на русском и английском языках, включая монографии, главы в книгах и статьи в рецензируемых научных журналах, в том числе в *International Review of Pragmatics, Intercultural Pragmatics, Journal of Politeness Research* и др.
e-mail: larina-tv@rudn.ru
https://orcid.org/0000-0001-6167-455X

Research article / Научная статья

Russian reduplicative surface-syntactic relations in the perspective of general syntax

Igor MEL'ČUK ✉

Observatoire de linguistique Sens-Texte,
Université de Montréal, Montréal, Canada
✉igor.melcuk@umontreal.ca

To the dear memory of ANDRÉ CLAS (1933–2022), the founder and first chair of the Department of Linguistics and Translation at the University of Montreal.

Abstract

The paper considers lexical reduplications in Russian in the perspective of general syntax. The goal is to define and fully characterize special Russian surface-syntactic relations [SSyntRels], that is, the reduplicative SSyntRels, which appear exclusively in syntactic idioms formed by lexical reduplications. The syntactic operation **REDUPL** is defined, and several reduplicative SSyntRels are introduced. A deductive calculus thereof is proposed, based on three parameters concerning the correlations between the reduplicate and the reduplicand: the reduplicate is anteposed/postposed (with respect to the reduplicand); is in contact/is not in contact (with the reduplicand); represents an exact/inexact copy (of the reduplicand); eight reduplicative SSynt-Rels are theoretically possible. The notion of syntactic idiom (a non-compositional multilexemic expression having a non-segmental signifier) is formulated and illustrated: e.g., the sentence Mne_Y $prazdnik_X$ ne v $prazdnik_{L'(X)}$ lit. 'To me the feast is not into a feast' = 'I cannot enjoy the feast', which implements the syntactic idiom [X to.Y] ˹BE NOT INTO L'(X)˺ 'X cannot be enjoyed by Y'. Six reduplicative SSyntRels of Russian and one of English are described. These SSyntRels are conceived as a fragment of a general inventory of SSyntRels in the world languages.

Keywords: *typology, syntax, dependencies, phraseology, syntactic idioms, Russian*

Русские редупликативные поверхностно-синтаксические отношения в аспекте общего синтаксиса

И.А. МЕЛЬЧУК ⬤ ✉

Обсерватория лингвистики «Смысл-Текст»,
Монреальский университет, Монреаль, Канада
✉igor.melcuk@umontreal.ca

Аннотация

Русские лексические редупликации (удвоения) рассматриваются в аспекте общего синтаксиса. Цель статьи — определить и полностью охарактеризовать специальные русские редупликативные поверхностно-синтаксические отношения [ПСинтО], а именно — редупликативные ПСинтО, которые выступают исключительно в синтаксических идиомах, основанных на редупликациях. Определяется синтаксическая операция редупликации **REDUPL**. Вводятся редупликативные ПСинтО и предлагается дедуктивное исчисление таких ПСинтО, основанное на трех параметрах, задающих соотношения между редупликатом и редупликандом: редупликат предшествует редупликанду/следует за ним; находится/не находится в контакте с редупликандом; представляет собой точную/неточную копию редупликанда. Тем самым теоретически возможны 8 редупликативных ПСинтО. Формулируется и иллюстрируется понятие синтаксической идиомы — некомпозиционного многолексемного выражения с несегментным означающим. Например, предложение *Мне*$_Y$ *праздник*$_X$ *не в праздник*$_{L'(X)}$ 'Я не могу наслаждаться праздником' является реализацией синтаксической идиомы [Y-у X] ˹БЫТЬ НЕ В L′(X)˺ 'Y не может наслаждаться X-ом'. Полностью описываются шесть русских редупликативных ПСинтО и одно английское редупликативное ПСинтО. Эти ПСинтО рассматриваются как фрагмент инвентаря ПСинтО, встречающихся в языках мира.

Ключевые слова: *типология, синтаксис, зависимости, фразеология, синтаксические идиомы, русский язык*

1. Stating the problem

The present study is part of a huge research task: compiling a general inventory of surface-syntactic relations [SSyntRels] in world languages.

☛ Technical terms are printed, on the first mention, in Helvetica.

The final goal of such an endeavor is a list of SSyntRels found in as many languages as possible—something similar to an inventory of language sounds (or of phonemes), of grammatical cases or of grammatical moods and tenses encountered in various languages. Since each language has its own set of SSyntRels, a general inventory can only be a set-theoretical sum (= the union) of particular SSyntRel lists established empirically for particular languages.

The first step towards the declared goal was taken 60 years ago: the paper Mel'čuk 1962: 47–49 presented a tentative list of 31 SSyntRels for Russian; this list was reproduced in Mel'čuk 1963: 491–493. Since then, several lists of SSyntRels for different languages were published:

— For Russian: Mel'čuk 1974: 221–235, 2012: 135–144, 2018a and Iomdin 2010.

— For English: Mel'čuk & Pertsov 1987: 85–156, Mel'čuk 2012–2015: vol. 3, 444–453 and 2016: 184–194.

— For French: Iordanskaja & Mel'čuk 2009, Poiret & Liu 2019.

— For German: Zangenfeind 2012.

— For Spanish: Bolshakov 2002.

In all probability, there are more such lists that I simply am not aware of.

NB A universal inventory of syntactic dependency relations based on syntactic dependency tree banks for over 70 languages, known as Stanford Universal Dependencies [SUDs], is proposed in Marneffe & Manning 2008 and Marneff *et al.* 2014; see also https://universaldependencies.org. However, SUDs are "ideologically" incompatible with SSyntRels discussed here: UDs are not really syntactic relations—they merge semantic and syntactic dependencies, the whole system being adapted for computer processing of texts. The theoretical framework and the methodology for SUDs and for SSyntRels are so different that a comparison would require a special study. The paper Gerdes *et al.* 2018 proposes a modification of SUDs, making them closer to a linguistically valid inventory of SSyntRels.

The studies Mel'čuk 2015–2016 and 2021b: 31–116 present an attempt at inventing the SSyntRels needed to describe the surface-syntactic structures [SSyntSs] of utterances in several languages of various types. This inventory needs, of course, extension and sharpening, that is, additions and modifications coming from different languages. The present paper offers one such addition: the Russian reduplicative SSyntRels, that is, the SSyntRels that are used exclusively in constructions produced by the operation of syntactic reduplication, called **REDUPL**. The paper's goal is thus to formally describe the Russian reduplicative SSyntRels.

A formal description is only possible within a predefined formal framework, and in what follows such a framework is the Meaning-Text approach (Mel'čuk 1974, 2012–2015, 2016 and 2021b, Mel'čuk & Milićević 2020); a sufficient familiarity with the corresponding notions and formalisms on the part of the reader is assumed.

> **NB** In order to facilitate the task of the reader, the paper is supplied with Appendix 1 (some crucial linguistic notions) and Appendix 2 (Russian surface-syntactic relations mentioned in the paper); here is a list of the abbreviations and notations used:

ATTR	: **ATTRIBUTIVE** deep-syntactic relation	SSyntRel	: surface-syntactic relation
DMorphS	: deep-morphological structure	SSyntS	: surface-syntactic structure
DSyntS	: deep-syntactic structure	**1, 2,** …	: semantic actants
L	: a lexical unit	**I, II,** …	: deep-syntactic actants
L(X)	: a lexical unit L that expresses X	« L »	: a fictitious lexeme L
L′	: a copy [= a reduplicate] of L	's'	: a meaning 's'
REDUPL	: operation of syntactic reduplication	⌜L₁ + L₂ + ... + Lₙ⌝	: an idiom $L_1 + L_2 + ... + L_n$
's'	: a communicatively dominant semanteme	[X]	: an actant X
SemS	: semantic structure	{ }	: text added for clarity

Before I go down to business, let it be emphasized that, although the linguistic data analyzed in this paper come from Russian, the formal proposals are universally valid.

2. Reduplication in syntax

2.1. Syntactic reduplication operation

The first thing to do is to introduce the **REDUPL** syntactic operation formally.

REDUPL is the repetition, or doubling, of lexemic expressions in an utterance, that is, the repetition of whole wordforms—as opposed to morphological reduplication, which affects only parts of wordform signifiers, as, for instance, in the Latin perfect: **mord**(-*eo*) 'I.bite' ~ **mo**+**mord**(-*i*) 'I.bit'.

For simplicity, in what follows, only the application of **REDUPL** to single lexemes is considered.[1]

---

[1] The **REDUPL** operation can apply to phrases as well; for instance, see some English examples in Ghomeshi *et al.* 2004: (1g) *Oh, we're not living-together living together* or (59b) *I never talked to him talked to him* and such Russian examples as *Čerez tri dnja tak čerez tri dnja* lit. 'In three days then in three days'. = 'Well, in three days is OK with me' or *Veselit'sja do utra tak veselit'sja*

Definition 1: REDUPL, syntactic reduplication operation

The **REDUPL** syntactic reduplication operation applies to a lexeme L that labels a node in a surface-syntactic structure [SSyntS] and produces an SSynt-subtree that replaces L in the   SSyntS:

REDUPL$(L_{\{G\}}) = L_{\{G\}}$–**r**$_i$→$L'_{\{G'\}}$, where

- $_{\{G\}}$ is the set of morphological deep (= semantically full) grammemes of L, which are, so to speak, inherited from the deep-syntactic structure [DSyntS];
- L′ is a copy of L, exact or with some derivational modification;
- $_{\{G'\}}$ is the set of morphological semantically full grammemes coming to L′ from L;
- **r**$_i$ is, in most cases (but not always), one of the reduplicative SSyntRels.

$L_{\{G\}}$ is called the reduplicand and $L'_{\{G'\}}$, the reduplicate.[2]

Examples

• Consider the Russian phrase (1), which includes a phrase being the result of a **REDUPL** application; the latter phrase implements the syntactic idiom «VERY» (on syntactic idioms, see Section 3).

☛ **1.** «VERY» is a fictitious lexeme, used as the name of this syntactic idiom: see Appendix 1, p. 904.
  2. In the examples, the reduplicate is boldfaced; words in curly brackets are added for better understanding.
  3. The underscoring of a semanteme in a SemS shows its communicatively dominant status (Mel'čuk 2012: Ch. 6, Section 2).

(1)   *{testo dlja} vkusnogo*$_L$-***vkusnogo***$_{L'}$ *torta*
         lit. 'dough for tasty-tasty [= 'very tasty'] cake'

The SemS, DSyntS, SSyntS and the deep-morphological structure [DMorphS] of this phrase are as follows:

SemS   : 'dough←**1**–for–**2**→cake←**1**–tasty←**1**–very'
DSyntS : {DLJA–**II**→TORT$_{(masc)SG}$–**ATTR**→} VKUSNYJ$_{LONG, POSIT}$–**ATTR**→«VERY»
SSyntS   :
{DLJA–**prep**→TORT$_{(masc)SG}$–**modif**→} VKUSNYJ$_{LONG, POSIT}$–**postpos-imm-exact-redupl--**

                               →**VKUSNYJ**$_{LONG, POSIT}$

---

do utra lit. 'To.have.fun till morning then to.have.fun till morning'. = 'Having fun till morning should be done without hesitation and/or intensely'.

  [2] Several publications tend to distinguish different types of reduplication by different terms, calling, for instance, the *vkusnyj-vkusnyj* 'tasty-tasty' type expressions REDUPLICATIONS, and the *vkusnyj, vkusnyj* 'tasty, tasty' type expressions, REPETITIONS. However, such a practice contradicts the general principles of building deductive notional systems, where the classification must start with one most comprehensive class—in our case, the class of syntactic reduplications, which is divided in appropriate subclasses, and so forth.

DMorphS:

{DLJA} VKUSNYJ$_{\text{LONG, POSIT, MASC, SG, GEN}}$ ^{3} **VKUSNYJ**$_{\text{LONG, POSIT,}}$ $\boxed{\text{MASC, SG, GEN}}$ {TORT$_{\text{(masc)SG, GEN}}$}

The **postpos-imm-exact-reduplicative** SSyntRel—see Subsection 4.5—ensures that the reduplicate **VKUSNYJ** receives in the DMorphS all the syntactic grammemes (boxed) of its reduplicand. These grammemes are protected from all possible further modifications.

• The reduplicative phrase in (2), which also is the result of a **REDUPL** application, implements the syntactic idiom «COMPLETELY»:

(2) {*Èx,*} **polnym-***polna korobuška!* [N. A. Nekrasov]
  lit. 'Well, is by.full-full basket!' = 'Well, my basket is completely full!'
  [a song of rural peddlers in the 19$^{\text{th}}$ century Russia].

SemS : 'basket←**1**–<u>full</u>←**1**–completely'

DSyntS :

{KOROBUŠKA$_{\text{(fem)SG}}$←I–BYT'$_{\text{IND, PRES}}$}–II→POLNYJ$_{\text{SHORT}}$–**ATTR**→«COMPLETELY» ^{4}

SSyntS :

{KOROBUŠKA$_{\text{(fem)SG}}$←**subj**–BYT'$_{\text{IND, PRES}}$–**cop-compl**→}POLNYJ$_{\text{SHORT}}$–**antepos-imm-LONG.INSTR-**

redupl→POLNYJ$_{\text{LONG, POSIT}}$

DMorphS : **POLNYJ**$_{\text{LONG, POSIT,}}$ $\boxed{\text{MASC, SG, INSTR}}$ POLNYJ$_{\text{SHORT, SG, FEM}}$ KOROBUŠKA$_{\text{(fem)SG, NOM}}$

The reduplicate gets its grammemes LONG and POSIT already in the SSyntS (in the process of the implementation of the «COMPLETELY» idiom), and the rest of its grammemes—MASC, SG, and INSTR—comes in the DMorphS from the implementation of the **antepos-imm-LONG.INSTR-reduplicative** SSyntRel (for more on the surface implementation of the «COMPLETELY» idiom, see Subsection 4.2.2).

2.2. Reduplicative surface-syntactic relations [SSyntRels]

2.2.1. Introductory remarks

The **postpos-imm-exact-reduplicative** and **antepos-imm-LONG.INSTR-reduplicative** SSyntRels, used in (1) and (2), are called reduplicative, because they appear in the SSyntS exclusively as a result of the application of the **REDUPL** operation. More precisely:

---

^{3} The Russian adjective has, among others, two opposed inflectional forms:

— The LONG form (e.g., ŠIROK+IJ 'broad' or MOLOD+OJ 'young') is used in all possible syntactic roles of the Russian adjective; it expresses number, gender and case.

— The SHORT form (ŠIROK+Ø, MOLOD+Ø) is used only as the copular complement of the verbs BYT' 'be', STAT' 'become', OKAZAT'SJA 'turn out'; it expresses only number and gender and does not have cases.

^{4} Many Russian syntactic idioms feature one of several lexemes of the verb BYT' 'be': BYT'**I.1**—semantically empty copula, BYT'**I.2** 'be identical to', BYT'**I.3** 'be an element of a class', and BYT'**IV** 'be located at'. In what follows the lexicographic numbers with BYT' are omitted as irrelevant for the purposes of this paper.

> A reduplicative SSyntRel can appear in the SSyntS, that is, in a reduplicative phrase, only as a result of an application of the **REDUPL** operation. The converse statement is not true: a reduplicative phrase may contain no reduplicative SSyntRel, because the **REDUPL** operation does not necessarily entail the use of a reduplicative SSyntRel.

The **REDUPL** operation is used exclusively in syntactic idioms and produces reduplicative syntactic idioms. But a syntactic idiom with lexical reduplication can contain no reduplicative SSyntRel: in such an idiom, the reduplicate is the dependent member of a non-reduplicative SSyntRel. For instance, the SSyntS of the syntactic idiom [X Y-u] ⌜BYT′ NE V L′(X)⌝ lit. 'X to.Y is not into L′(X)'. = 'X cannot be enjoyed by Y' contains no reduplicative SSyntRel, cf. (8), p. 892. Here are three more examples of syntactic idioms with a reduplication but without a reduplicative SSyntRel:

(3)   a. [X] ⌜BYT′ L′(X)-*u* ROZN⌝ lit. 'X is to.X difference'. =
'Xs are different'; for instance:

$$\text{\hspace{4em}}\overset{\displaystyle\lceil\text{\textbf{copular-completive}}\longrightarrow\rceil\downarrow}{}$$
Kniga←**subj**–Ø^{BYT′}–**indir-obj**→**knige rozn'** 'Books are different'.

   b. [X] ⌜BYT′ KAK L′(X)⌝ lit. 'X is as X'. = 'X is quite an ordinary X'; for instance:
Kniga←**subj**–*byla*–**copular-completive**→*kak*–**subject-compar-conjunct**→**kniga**
'The book was quite an ordinary book'.

   c. [X] ⌜I L′(X)⌝ lit. 'X and X'. = 'This is X, nothing special'; for instance:
Kniga–**coordinative**→*i*–**coord-conjunctional**→**kniga**
'This is a book, nothing special'.

The two crucial questions to be answered about reduplicative SSyntRels are obvious:
— What kind of SSyntRels are the reduplicative SSyntRels?
— What reduplicative SSyntRels are logically possible?

2.2.2. The nature of reduplicative SSyntRels

The reduplicative SSyntRels are not semantically loaded, or meaningful: a reduplicative SSyntRel does not carry a particular meaning—that is, it is not directly linked to a semanteme or a configuration of semantemes. In this respect, the reduplicative SSyntRels are similar to dozens of "normal," or "meaningless," SSyntRels. As a rule, an SSyntRel, which links two lexemes in an SSyntS, does not carry itself any meaning. (4) gives three examples of such SSyntRels in Russian:

(4) semantic structure [SemS]   surface-syntactic structure [SSyntS]

     a. 'sleep–**1**→boy'   : *Mal'čik*←**subjectival**–*spit*
                        'The.boy is.sleeping'.

     b. 'sleep←**1**–in–**2**→corridor' : *spit*–**circumstantial**→*v*–**preposit**→*koridore*
                        'is.sleeping in the.corridor'

     c. 'intense–**1**→wind'   : *sil'nyj*←**modificative**–*veter* 'strong wind'

☞ The period between words in the glosses is used to unite glossing words in such a way as to obtain one-to-one correspondence between Russian words and their glosses.

A meaningless SSyntRel between two lexemes in an SSyntS expresses the semantic dependency relation between the corresponding semantemes in the SemS, rather than any semantemes as such.

The majority of SSyntRels of a language are exactly like the SSyntRels in (4); they are meaningless, that is, purely syntactic. A tentative inventory of meaningless SSyntRels of world languages is presented in Mel'čuk 2021b: Ch. 2.

However, languages also have meaningful SSyntRels. A meaningful, or semantically loaded, SSyntRel does more than link two lexemes into a phrase; it also expresses a specific chunk of meaning—a semanteme or a configuration of semantemes. In other words, a meaningful SSyntRel carries a meaning of lexical type. A well-known Russian example is the approximate-quantitative phrase:

(5) a. *pjat' tonn* 'five tons'
       vs.
     b. *tonn pjat'* 'maybe five tons'

The phrase in (5b)—with an inverted order of NUM and N—expresses the uncertainty of the Speaker about the indicated quantity, i.e., it expresses the semanteme 'maybe', which appears in the starting semantic structure. In the DSyntS, the semanteme 'maybe' is rendered by the fictitious lexeme «MAYBE», and in the SSyntS, by a meaningful SSyntRel: the **approximate-quantitative** SSyntRel; cf. (6):

(6) SemS   : 'maybe–**1**→five–**1**→tons'
     DSyntS   : «MAYBE»←**ATTR**–PJAT'←**ATTR**–TONNA$_{PL}$
     SSyntS   : PJAT'←**approximate-quantitative**–TONNA$_{PL}$
     DMorphS : TONNA$_{PL, GEN}$ PJAT'$_{NOM}$ (*tonn pjat'*)
                (vs. PJAT'←**quantitative**–TONNA$_{PL}$: *pjat' tonn*)

The reduplicative SSyntRels are special in the following respect: They are, as stated above, meaningless, but they are used exclusively in reduplicative phrases, the latter being the implementations of syntactic idioms (introduced in Section 3 below), which are, of course, meaningful. Thus, these SSyntRels maintain an intimate relationship with syntactic idioms; as a result, they constitute a particular subset of Russian meaningless SSyntRels. It is this subset that is described in Section 4.

2.2.3. The calculus of reduplicative SSyntRels

Now, let us see what reduplicative SSyntRels can in principle exist. The operation of syntactic reduplication can be characterized according to the following three parameters:

• Linear position of the reduplicate:

the reduplicate precedes (**antepos-**) / follows (**postpos-**) the reduplicand.

• Linear contact between the reduplicate and the reduplicand:

the reduplicate is (**-imm[ediate]-**) / not necessarily is (**-non.imm-**) in contact with the reduplicand.

NB The statement "The reduplicate precedes/follows the reduplicand immediately" must be understood *cum grano salis*. Namely, this means that the two cannot be separated by arbitrary lexemes allowed, generally speaking, in this position by standard syntactic rules of the language; but some particular lexemes—mostly, different particles—foreseen by the lexical entry of the corresponding syntactic idiom are possible between the reduplicate and the reduplicand, even if these are said to be in immediate contact.

• Exactness of the reduplicate:

the reduplicate is an exact (**-exact-**) / not an exact (**-inexact-**) copy of the reduplicand.

NB An inexact copy L′ of the lexeme L can be, strictly speaking, inexact in two respects:

— L′ is affixed with a derivational means, which comes from the lexical entry of the syntactic idiom that has the reduplicative SSyntRel under consideration as part of its signifier (for instance, the English «DERISION» syntactic idiom: e.g., *Ah, your theories, **schm**eories!*). This "inexactness" does not concern the corresponding reduplicative SSyntRel. In other words, the reduplicative SSyntRel that links a derived reduplicate to the reduplicand is encoded as "exact," provided no grammemes of the reduplicate are affected.

— $_{\{G'\}}$, that is, the set of syntactic grammemes of the reduplicate L′, contains syntactic grammemes different from syntactic grammemes $_{\{G\}}$ of L: this is the direct and exclusive responsibility of the corresponding reduplicative SSyntRel. The reduplicand L′ is coded as inexact only in this case.

These three parameters specify eight logically possible—that is, language universal—reduplicative SSyntRels.

However, in reality, the set of reduplicative SSyntRels of a particular language does not necessarily contain exactly these eight logically deduced SSyntRels. On the one hand, a language may not have all of the eight logically possible reduplicative SSyntRels: thus, as the reader will immediately see, Russian lacks some of these. On the other hand, an inexact reduplicative SSyntRel specifies the modifications to be performed in the reduplicate L′'s grammemes, and these modifications cannot be foreseen logically. So there may be several different inexact reduplicative SSyntRels, depending on the language. To sum up, the inventory of reduplicative SSyntRels for a particular language must be established empirically, and that is what is done in Section 4 for Russian.

3. The habitat of reduplicative SSyntRels: syntactic idioms

Reduplicative SSyntRels are found, as stated above, only in syntactic idioms, so that they are inextricably linked to the latter. This requires the notion of syntactic idiom to be formally introduced. Let me start with three underlying notions, which concern linguistic signs.

• A sign **s** is complex if and only if its signifier contains more than one linguistic entity.

> **NB** Linguistic entities are of two kinds:
> linguistic expressive means (segmental—segments, i.e. phonemic strings that are signifiers,[5] and non-segmental—operations, prosodies, SSyntRels organized in a subtree, word order, and grammemes)
> and
> signs, whose signifiers are built out of linguistic expressive means.

A particular subtype of complex signs are multilexemic signs. A sign is multilexemic if and only if its signifier:
— either contains the signifiers of two or more lexemes;
— or is a prosodic structure imposed upon two or more lexemes.

• A sign **s** is non-compositional if and only if the components of its signified cannot be distributed between the components of its signifier in a regular (= not-*ad hoc*) way.

> **NB** A non-compositional complex sign is an idiom *tout court*.

• A sign **s** is non-segmental if and only if its signifier includes some non-segmental linguistic expressive means.

> **NB** A non-segmental idiom is a syntactic idiom.

Now the definition of syntactic idiom can be readily formulated.

Definition 2: syntactic idiom

> A linguistic sign **s** is a syntactic idiom if and only if it is
>   (i) multilexemic,
> (ii) non-compositional,
> (iii) non-segmental.

> **NB** On syntactic idioms, see Mel'čuk 1987, 2012: 18–20, 2021a and 2023a: Ch.11.

Examples

> ☞ The top corners ˊ ˋ indicate an idiom; the square brackets [ ] include the actants of the expression under consideration; L(X) means 'lexeme L that expresses X', and L′ is, as stated above, a copy of L; $L_1 + \ldots + L_2$ means 'L_1 precedes L_2 with a possible lexemic gap between L_1 and L_2'.

---

[5] But not phonemes as such: a phoneme is a linguistic means serving to distinguish segmental signifiers.

• The Russian complex sign [X] «WILL.PUNISH» [Z for Y-ing] is a good example of syntactic idiom, illustrated by the sentences in (7):

(7)   a. *Ivan*$_{L(X)}$ *tebe*$_{L(Z)}$ *potancuet*$_{L(Y)}$ ⟨*prygnet*$_{L(Y)}$⟩*!* lit. 'Ivan to.you will.dance ⟨will.jump⟩!' = 'If you dance ⟨jump⟩, Ivan will punish you'.
      b. *Ja*$_{L(X)}$ *emu*$_{L(Z)}$ *budu morožennoe žrat'*$_{L(Y)}$*!* lit. 'I to.him will ice.cream gobble!' = 'If he gobbles down ice cream, I will punish him'.

All lexemes of a sentence that implements this idiom are parts of the idiom's actants; the meaning of the idiom itself—a threat of severe punishment for a reprehensible activity—is expressed by a particular SSynt-structure and a particular prosody. With different prosodies, the sentences in (7) become statements with different meanings (depending on the prosody); thus,

↘   ↘   ↘

Ivan tebe potancuet{*, ne somnevajsja*} 'Ivan will dance for you, don't doubt': a promise;
or

            ↘   ↗ ↘

{*Kak že,*} *Ivan* ↑ *tebe potancuuuet!* 'Don't even hope, Ivan will never dance for you!': a sarcastic negation of a possibility.

Here is the lexical entry of this idiom.

[X] «WILL.PUNISH» [Z for Y-ing], syntactic idiom, clausative.

Signified [= Lexicographic definition]
'X «WILL.PUNISH» Z for Y-ing' = 'If Z does Y, X will punish Z for Y-ing'

Signifier
⎡ 1) L(X)←**subjectival**–L(Y)–**indirect-objectival**→L(Z)
⎢ 2) ↗   ↗   _↗↘;   threat intonation
⎣ 3) L(X) + … + L(Z) + … + L(Y) [6]

Syntactics [= Government Pattern]

'X' ⟺ I	'Y' ⟺ II	'Z' ⟺ III
1. S$_{NOM}$	1. V$_{FUT}$	1. S$_{DAT}$

1) L('Z') is a personal pronoun or (less preferably) a human proper name.

Ivan$_X$ *tebe*$_Z$ *poprygaet*$_Y$*!* lit. 'Ivan$_X$ to.you$_Z$ will.jump$_Y$!' =
'Ivan will punish you for (repeated) jumping'.
Ivan$_X$ *emu*$_Z$ *prygnet*$_Y$*!* lit. 'Ivan$_X$ to.him$_Z$ will.jump$_Y$!' =
'Ivan will punish him for one jump'.

---

[6] The word order indication in the signifier of a syntactic idiom specifies the neutral, most frequent linear arrangement of the lexemes; this arrangement can change under the impact of the communicative structure of the sentence.

The signifier of this sign is complex: it contains a prosodic structure imposed upon a lexemeless syntactic tree (a system of SSyntRels linking the lexemic variables that represent the idiom's actants), plus a word order indication—L(X) must precede L(Z), and L(Z) precedes L(Y). The sign is also non-compositional: on what signifier component can the semantemes 'punish' and 'will' be loaded? And it is obviously non-segmental. So this sign is a syntactic idiom.

• Another example of syntactic idiom is the complex sign [X Y-u] ˊBYTˊ NE V L′(X)ˋ:

(8) *Bez pesen Ivanu*$_{L(Y)}$ *i p'janka*$_{L(X)}$ *ne v p'janku*$_{L′(X)}$
   lit. 'Without singing to.Ivan even a.bender is not into bender'. =
      'If there is no singing, even a bender cannot be enjoyed by Ivan'.

The lexical entry of this idiom is as follows.

[X Y-u] ˊBYTˊ NE V L′(X)ˋ, syntactic idiom, clausative.

Signified [= Lexicographic definition]
'X Y-u ˊbyt′ ne v L′(X)ˋ' {lit. 'X to.Y is not into L′(X)'} 'X cannot be enjoyed by Y'

Signifier

   1) **REDUPL**(L(X)$_{number}$) = L(X)$_{number}$, L′(X)$_{number}$

   ┌─**copular-completive** ─┐↓
   2) L(X)$_{number}$←**subjectival**–BYTˊ NE←**restrictive**–V–**prepositional**→L′(X)$_{number}$

   3) L(Y) + … + L(X)

Syntactics [= Government Pattern]

'X' ⇔ **I**	'Y' ⇔ **II**
1. S$_{NOM}$	1. S$_{DAT}$

Nam$_Y$ *obed*$_X$ *ne v obed*$_{L′(X)}$ lit. 'To.us$_Y$ dinner$_X$ is not into dinner$_{L′(X)}$'. =
'We cannot enjoy the dinner'.

This sign is also complex, since its signifier includes several expressive means: three segments (the signifiers of the lexemes BYTˊ 'be', NE 'not' and V 'into'), and three non-segmental means—the **REDUPL** operation, an SSynt-subtree and a word order indication. It is non-compositional, since it is impossible to attach, in a not-*ad hoc* way, the semantemes 'can' and 'enjoy' in its signified to any component of its signifier. Finally, it is non-segmental, since its signifier includes non-segmental expressive means. Therefore, it is a syntactic idiom.

   NB Note that the signifier of this reduplicative syntactic idiom contains no
      reduplicative SSyntRel: the [X Y-u] ˊBYTˊNE V L′(X)ˋ idiom illustrates the case
      mentioned in Subsection 2.2.1, p. 887.

Syntactic idioms are lexical units—paradoxical ones, but lexical units. They must be stored in the lexicon of the language as all lexical units are and supplied with full-fledged lexical entries.

Now everything is ready to concentrate on the Russian reduplicative SSyntRels.

4. Russian reduplicative SSyntRels

Russian reduplicative constructions have been described several times: for instance, Israeli 1997, Krjučkova 2004 and Sannikov 2008, 2010; there are also numerous studies dedicated to particular cases, which will be indicated when appropriate. However, the question of special reduplicative SSyntRels has not been raised before, as far as I know. The Russian reduplicative idioms are treated in numerous studies by M. Kopotev: see Kopotev 2008 and Janda, Kopotev & Nesset 2020; see also Mel'čuk 2023b.

In the inventory below, each reduplicative SSyntRel is illustrated with Russian syntactic idioms in which it appears as a part of the signifier. (But not all such syntactic idioms are listed.)

Necessary information about the implementation of a syntactic idiom is found in its lexical entry; since the lexical entries of the idioms appearing in the illustrations cannot be supplied here, numerous details concerning the surface form of the corresponding phrase may remain obscure for the reader.

4.1 The antepos-imm-exact-reduplicative SSyntRel

This SSyntRel does not exist in Russian.

4.2 The antepos-imm-INEXACT-reduplicative SSyntRels

The expression "**antepos-imm-INEXACT-reduplicative**" is a cover name for all inexact reduplicative SSyntRels, which are "antepos" and "imm"; it can refer to several particular, i.e. language-specific, inexact SSyntRels. These SSyntRels carry different indications (boxed in the examples below) of the modifications in L′'s syntactic grammemes. Russian has two **antepos-imm-INEXACT-reduplicative** SSyntRels.

4.2.1. The antepos-imm-DAT-reduplicative SSyntRel

$$
\begin{array}{c}
\text{L} \\
\circ \\
| \\
\textbf{antepos-imm-} \\
\textbf{DAT-reduplicative} \\
| \\
\downarrow \\
\circ \\
\textbf{L}'_{\textbf{PL}}
\end{array}
\quad \Leftrightarrow \quad
\textbf{L}'_{\textbf{PL, }\boxed{\textbf{DAT}}} + \text{L}_{\boxed{\text{NOM}}}
$$
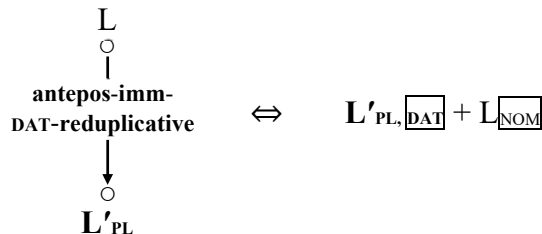
☞ Boldfacing in the rule and in the examples indicates the reduplicate.

This SSyntRel is part of the signifier of the reduplicative syntactic idiom [X] ⌐BYT′ VSEM L′(Y)-am⌐ [Y] {lit. 'L(X) is to.all L′(Y)s L(Y)'} 'X is the most outstanding Y of all Ys'; for instance:

(9) *Užba*$_{L(X)}$ – *vsem **goram***$_{L'(Y)}$← **antepos-imm-DAT-redupl**–*gora*$_{L(Y)}$

   lit. 'Uzhba is to.all mountains mountain'. =
      'Uzhba is the most outstanding mountain of all mountains'.

This idiom appears in the SemS, DSyntS, SSyntS and DMorphS as follows (with 'X' = 'Užba' and 'Y' = 'gora/mountain'):

SemS   : 'Užba←**1**–<u>be</u>–**2**→mountain←**1**–most.outstanding–**2**→mountains←**1**–all'
DSyntS   : UŽBA$_{SG}$←**I**–⌐BYT′ VSEM L′(Y)-am⌐$_{IND, PRES}$–**II**→GORA$_{SG}$
SSyntS   : UŽBA$_{SG}$←**subj**–BYT′$_{IND, PRES}$–**copul-compl**→GORA$_{SG}$–**antepos-imm-DAT-redupl**→**GORA**$_{PL}$–**modif**→VSE
DMorphS : UŽBA$_{SG, NOM}$ Ø$^{BYT′}$ VSE$_{PL, DAT}$ **GORA$_{PL, DAT}$** GORA$_{SG, NOM}$

Comments

1) The SSyntS proposed here for the [X] ⌐BYT′ VSEM L′(Y)-am⌐ [Y] idiom can be questioned: Does the reduplicate (*vsem*) ***goram*** depend on the reduplicand *gora* (as I believe) by the **antepos-imm-DAT-reduplicative** SSyntRel or is it rather an actant (= indirect object) of the verb BYT′ 'be'? One of the BYT′ lexemes does govern a similar construction:

(10) *Ivan byl vsem nam*$_{DAT}$ *drug*$_{NOM}$/*drugom*$_{INSTR}$
      lit. 'Ivan was to.all us friend'. = 'Ivan was a friend to all of us'.

However:
   • The copular complement of BYT′ can be in the nominative or in the instrumental, while the reduplicand in our idiom can only be in the nominative; this is easily ensured by the **antepos-imm-DAT-reduplicative** SSyntRel.
   • The copular complement of BYT′ is linearly quite flexible, while the reduplicate of the idiom under consideration is not:

   Vsem nam Ivan byl drug. vs. **Vsem goram Užba byla gora*.

   • The dative indirect object is possible with BYT′ only if BYT′ has a specific noun as its copular complement; there is a necessary semantic link between DRUG 'friend' and MY 'we': 'friend–**2**→we'. The reduplicate of the idiom can be any noun.
   Therefore, the dependence of the reduplicate on the reduplicand (by a reduplicative SSyntRel) is established.

2) The specificity of the **antepos-imm-DAT-reduplicative** SSyntRel consists in imposing a syntactic grammeme, namely, the NOM(inative) case, on the governing element of the phrase, while the standard situation in Russian is for an SSyntRel—except for subjectival SSyntRels—to impose syntactic grammemes on the dependent member. This is necessary because the **copular-completive** SSyntRel, which subordinates the reduplicand to the verb BYT′, requires the NOM or the INSTR (as function of contextual conditions) for its dependent, while in the implementation of this idiom the reduplicand can be only in the NOM (see Comment 1). As indicated

above (p. 00), a grammeme introduced by a reduplicative SSyntRel (boxed in the rule) is immune from all further possible transformations.

4.2.2. The antepos-imm-LONG.INSTR-reduplicative SSyntRel

$$L_{SHORT}$$
$$\circ$$
$$|$$
antepos-imm-LONG.INSTR-reduplicative $\quad \Leftrightarrow \quad \mathbf{L'}_{LONG, POSIT, \boxed{MASC, SG, INSTR}} + L_{SHORT} \mid L \neq MASC$
$$\downarrow$$
$$\circ$$
$$\mathbf{L'}_{LONG, POSIT}$$
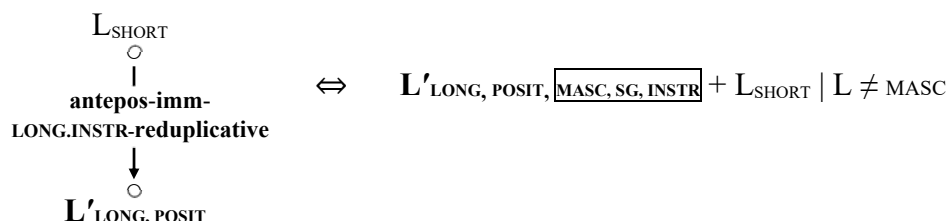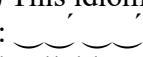
This SSyntRel appears in the signifier of the reduplicative syntactic idiom «COMPLETELY» [X] '[be] completely X':

(11) a. *Vokrug vsë belym*$_{L'(X)}$-←**antepos-imm-LONG.INSTR-redupl**–*belo*$_{L(X)}$
lit. 'Around everything is by.white-white'. =
'The whole landscape around is completely white'.
  b. *Zemlja byla černym-černa ot voronok*
lit. 'Earth was by.black-black from shell.craters'. =
'The earth was completely black because of shell craters'.
  c. *Druz'ja byli p'janym-p'jany* lit. 'Friends were by.drunk-drunk'. =
'The friends were completely drunk'.

The idiom «COMPLETELY» [X] '[be] completely X' on four levels of linguistic representation (with 'X' = 'černyj/black'):

SemS   : '{earth←1–}be.black←1–completely'
DSyntS   : {ZEMLJA$_{(fem)SG}$←**subj**–BYT'$_{IND,PAST}$–**cop-compl**→}ČËRNYJ$_{SHORT}$–**ATTR**→«COMPLETELY»
SSyntS   : {ZEMLJA$_{(fem)SG}$←**subj**–BYT'$_{IND, PAST}$–**copular-compl**→}—
        →ČËRNYJ$_{SHORT}$–**antepos-imm-LONG.INSTR-redupl**→ČËRNYJ$_{LONG, POSIT}$
DMorphS :
{ZEMLJA$_{(fem)SG, NOM}$ BYT'$_{IND, PAST, SG, FEM}$} **ČËRNYJ**$_{LONG, POSIT, MASC, SG, INSTR}$ ČËRNYJ$_{SHORT, SG, FEM}$
      [*Zemlja byla černym-černa.*]

Comments

 1) This idiom is characterized by a particular stress pattern of the implementing phrase: ‿‿́ ‿‿́. The short-form adjective must be bi-syllabic and stressed on the second syllable. As a consequence, this idiom is restricted: far from all adjectives that are semantically fit to serve as its actant 'X' (that is, the adjectives that are compatible with the semanteme 'completely' and have finally-stressed short forms) sound natural when used in it; thus, we do not have *pravym-pravy '[are] completely right' (because the correct short form is *právy*) or *spelym-spely '[are] completely ripe' (*spély*? *spelý*?). This can be related to the fact that the stress in short-form adjectives in contemporary Russian is undergoing a radical shift, so that the speakers are unsure of how to stress such adjectives.

2) The reduplicate—a long-form adjective—receives an "incorrect" stress on the last syllable; outside of this idiom this long-form adjective is stressed always on the first syllable: *černým-černa*, while the "correct" stress in the form under consideration is *čёrnym*.

The "incorrect" stress on the last syllable of the reduplicate comes from the idiom's prosodic structure in its lexical entry.

3) The adverbs DAVNYM-DAVNO 'very long ago' (vs. DAVNO 'long ago') and POLNYM-POLNO 'very many/very much' (vs. POLNO 'many/much') have the same formal structure as the reduplicative phrases implementing the «COMPLETELY» [X] idiom, but they are isolated: there is no other adverb of the same form, and semantically, DAVNYM-DAVNO and POLNYM-POLNO are also different from these adjectival phrases: they mean 'very…' rather than 'completely…'. Therefore, they are separate lexemes that must be stored as such in the lexicon along with DAVNO and POLNO.

4.2.3. The antepos-imm-STRICT.SENSE-reduplicative SSyntRel

This SSyntRel does not exist in Russian, but it is known in English; it seems useful to present it here, first, because it has a detailed and precise description in Ghomeshi *et al.* 2004 (from which all factual data are borrowed), and second, because it serves to implement the English syntactic idiom «IN.THE.STRICT.SENSE», the latter having a curious parallel in the Russian idiom «IN.THE.STRICT.SENSE», where one finds, however, a different SSyntRel: see (19), p. 899.

$$
\begin{array}{c}
\text{L}_{\{G\}} \\
\circ \\
| \\
\textbf{antepos-imm-} \\
\textbf{STRICT.SENSE-reduplicative} \\
| \\
\downarrow \\
\circ \\
\textbf{L}'_{\{G'\}}
\end{array}
\quad \Leftrightarrow \quad
\textbf{L}'_{\{G'\}} + \text{L}_{\{G\}}
\quad \left|
\begin{array}{l}
\textbf{if } \text{a } G' \text{ is expressed by a suffix,} \\
\textbf{then } \text{this } G' \text{ can be deleted; see (12e)}
\end{array}
\right.
$$
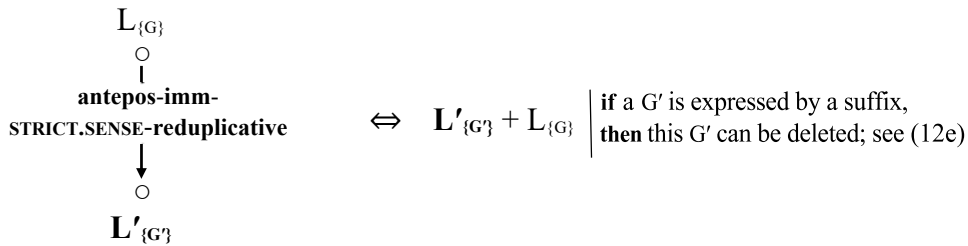
(12) a. *I'll make the tuna salad, and you make the* **salad**$_{L'(X)}$-*salad*$_{L(X)}$.
b. *My car isn't* **mine**-*mine; it's my parents'.*
c. *Are you* **leaving**-*leaving?*
d. *This time, John* **left**-*left.*
e. *I merely talked to him… Not* **talk**-*talked.*

The [X] «IN.THE.STRICT.SENSE» idiom (with 'X' = 'gloves') on four levels of linguistic representation:

SemS   : 'gloves←**1**−in.the.strict.sense'
DSyntS   : GLOVE$_{PL}$−**ATTR**→«IN.THE.STRICT.SENSE»
SSyntS   : **GLOVE$_{PL}$**←**antepos-imm-STRICT.SENSE-reduplicative**−GLOVE$_{PL}$
DMorphS : **GLOVE$_{SG}$** GLOVE$_{PL}$ [*I need* **glove**-*glove*s.]

4.3. The antepos-non.imm-exact-reduplicative SSyntRel

$$L_{\{G\}}$$
$$\circ$$
$$|$$

antepos-non.imm-
exact-reduplicative $\Leftrightarrow$ $\mathbf{L'}_{\{g\}} +...+ L_{\{G\}}$

$$\downarrow$$
$$\circ$$
$$\mathbf{L'}_{\{G'\}}$$
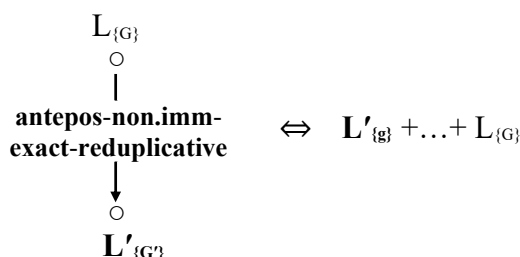
☛ The subscript $_{\{g\}}$ to the reduplicate L′ means 'all the grammemes that the reduplicand L has in the DMorphS'.

This SSyntRel is part of the signifier of the reduplicative syntactic idiom «I.CONFIRMING» [*that* X] 'I confirming that X':

┌─**ant-non.imm-ex-redupl**─┐
↓

(13) a. ***Doma***$_{L'(X)}$-*to Ivan byl doma*$_{L(X)}$
 lit. 'At.home$_{L'(X)}$-as.for Ivan was at.home$_{L(X)}$'. =
 'I confirm that Ivan was at home'.

 b. ***Karlik***-*to, konečno, Ivan karlik{, no nos u nego ogromnyj}*
 lit. 'Dwarf-as.for, of.course, Ivan is dwarf{, but nose at him is enormous}'. =
 'I confirm, of course, that Ivan is a dwarf{, but he has an enormous nose}'.

 c. ***Ivanu***-*to Ivanu my èto poslali*
 lit. 'To.Ivan-as.for to.Ivan we this have.sent'. =
 'I confirm that we have sent this to Ivan'.

 d. ***Perestroila***-*to, ja znaju, ona dom perestroila*
 lit. 'She.has.rebuilt-as.for, I know, she house she.has.rebuilt'. =
 'I confirm that she has rebuilt the house'.

Here is the «I.CONFIRMING» [*that* X] idiom on four levels of linguistic representation (with 'X' = 'doma/at.home'):

SemS : 'was–**2**→at.home←**2**–confirm–**1**→I'
DSyntS : «I.CONFIRMING»←**ATTR**–DOMA←**II**–BYT′$_{IND, PAST}$
SSyntS : TO←**restr**–DOMA←**an-non.imm-exact-redupl**–DOMA←**copul-completive**–BYT′$_{IND,PAST}$
DMorphS : **DOMA**-TO ... BYT′$_{IND, PAST, SG, MASC}$ DOMA

Comments
 1) The «I.CONFIRMING» idiom needs a complex enough description (which cannot be offered here): its L_x cannot be a non-finite form of a verb, and it has a particular communicative structure (the actant 'X' is an Emphasized Sem-Theme (see Mel'čuk 2001: 210–218; Sem-Theme is the fragment of the Sem-structure, i.e. a semantic chunk, about which something is said; Emphasized means 'having emotive importance for the Speaker').
 2) If the actant 'X' is implemented by a finite verb, two additional complications arise: (i) this idiom has a variant, described in Subsection 4.4.1—the finite verb is reduplicated by an infinitive; (ii) all syntactic dependents of L(X) can

be, and often are, transferred to the reduplicate L′(X); this is also true for the case in 4.4.1. For instance (the transferred dependents are boxed):

(14) *Perestroila-to $\boxed{ona\ dom}$, ja znaju, perestroila;*
cf. (13d) above, where ONA and DOM depend on the reduplicand.

4.4. The antepos-non.imm-INEXACT-reduplicative SSyntRel

Here too, as in Subsection 4.2, the name "**antepos-non.imm-INEXACT-reduplicative**" covers various particular inexact SSyntRels. The Russian language uses one of those: the **antepos-non.imm-INF-reduplicative** SSyntRel.

4.4.1. The antepos-non.imm-INF-reduplicative SSyntRel

$$L_{(V)\text{aspect, FIN}}$$
$$\circ$$
$$\mid$$
antepos-non.imm-
INF-reduplicative

$$\Big\Downarrow$$

$$\circ$$
$$\mathbf{L'}_{\text{aspect}}$$

$$\Leftrightarrow \qquad \mathbf{L'}_{\text{aspect, }\boxed{\text{INF}}} + \ldots + L_{(V)\text{aspect, FIN}}$$
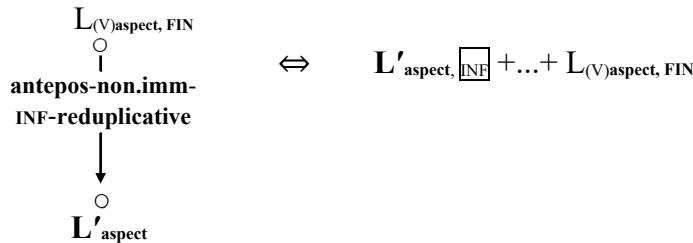
The **antepos-non.imm-INF-reduplicative** SSyntRel serves the same syntactic idiom «I.CONFIRMING» [*that* X] 'I confirm that X', described in Subsection 4.3. It is, so to speak, a contextual variant of this idiom foreseen for the case when the actant 'X' is a finite verb, as illustrated in (15):

```
               ┌──────antepos-non.imm-INF-redupl ──┐
```
(15) ***Perestroit′**_{L′(X)}-to, ja znaju, ona dom perestroila*_{L(X)}
lit. 'To.rebuild-as.for, I know, she house she.has.rebuilt'. =
'I confirm that I know that she has rebuilt the house'.

SemS   : 'I←**1**–confrm–**2**→rebuild–**2**→house'
DSyntS : «I.CONFIRMING»←**ATTR**–PERESTROIT′_{PERF, IND, PAST}–**II**→DOM_{SG}
SSyntS :
-TO←**restr**–PERESTROIT′_{PERF, INF}←**antepos-non.imm-INF-redupl**–PERE-
STROIT′_{PERF, IND, PAST}–**dir-obj**→DOM_{SG}
DMorphS : **PERESTROIT′**_{PERF, INF}-TO … DOM_{SG, ACC} PERESTROIT′_{PERF, IND, PAST, SG, FEM}

4.5. The postpos-imm-exact-reduplicative SSyntRel

$$L_{\{G\}}$$
$$\circ$$
$$\mid$$
postpos-imm-
exact-reduplicative

$$\Big\downarrow$$
$$\circ$$
$$\mathbf{L'}_{\{G'\}}$$

$$\Leftrightarrow \quad L_{\{G\}} + \mathbf{L'}_{\{g\}}$$
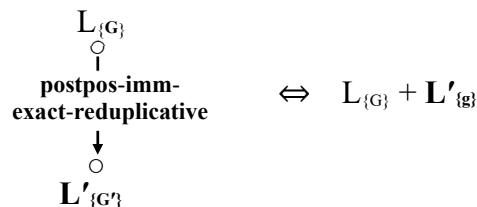
This SSyntRel is part of the signifier of several Russian reduplicative syntactic idioms: «VERY», see (16); «VERY-VERY», see (17); «ABNORMALLY», see (18); «IN.THE.STRICT.SENSE», see (19); [X] ˹TAK L′(X)˺, see (20); ˹ČTO [X], TO L′(X)˺, see (21):

(16)   «VERY»

    a. *Tak žalko ètix glupyx*_{L(X)}-–**postpos-imm-exact-reduplicative**→*glupyx*_{L′(X)}
    detišek! 'One is so sorry for these very stupid kids!'

    b. *Ščenok byl glupyj-**glupyj*** 'The puppy was very stupid'. ~
    *Ščenok kazalsja glupym-**glupym*** 'The puppy seemed very stupid'.

    c. *Pojti tuda bylo glupo-**glupo*** 'To go there was very stupid'.

    d. *Ivan ulybalsja glupo-**glupo*** 'Ivan was smiling in a very stupid way'.

On reduplication of Russian and English adjectives, see Apresjan, V. 2018.

> **NB**   The postposition of the reduplicate in the «VERY» idiom is established by analogy with such cases as *glupo-**preglupo*** in the next idiom.

(17)   «VERY-VERY»

    a. *Tak žalko ètix glupyx*_{L(X)}-**preglupyx**_{L′(X)} *detišek!*
    'One is so sorry for these very-very stupid kids!'

    b. *Ščenok byl glupyj-**preglupyj*** 'The puppy was very-very stupid'. ~
    *Ščenok kazalsja glupym-**preglupym***
    'The puppy seemed very-very stupid'.

    c. *Pojti tuda bylo glupo-**preglupo*** 'To go there was very-very stupid'.

    d. *Ivan ulybalsja glupo-**preglupo***
    'Ivan was smiling in a very-very stupid way'.

(18)   «ABNORMALLY»

*Dožd' lil*_{L(X)}-**lil**_{L′(X)}, *a potom vdrug zasijalo solnce*
lit. 'The.rain was.falling-was.falling, but then suddenly shined sun'. =
'The rain was falling non-stop for too long, but then suddenly the sun shined'.

> **NB**   As indicated in Subsection 2.2.3, p. 889, this syntactic idiom allows the reduplicand and the reduplicate to be separated by a particle, in this case—by NE 'not': e.g., *Ivan ne pisal-*⟦*ne* **pisal**⟧{, *a včera srazu tri pis'ma*} lit. 'Ivan didn't write-didn't write, but yesterday {we got} three letters at.once'.

(19)   «IN.THE.STRICT.SENSE»

*Maša priexala s parnem*_{L(X)}-**parnem**_{L′(X)}, *a ne s parnem-drugom*
lit. 'Masha came with guy-guy, but not with guy-friend'. =
'Masha came with her boyfriend, not with her male friend'.

(20)   [X] ˹TAK L′(X)˺

    a. *V mašinopisnom tak v **mašinopisnom***
    lit. 'In typewritten {form} then in typewritten '. =
    'I agree with the fact that this text [mentioned before] is typewritten'.

b. *Bez tak **bez*** lit. 'Without then without'. =
   'I agree to do without it [something mentioned before].

(21)   ⌜Čто [X], то L′(X)⌝
   Čto dëševo, ja soglasen, to dëševo
   lit. 'That {it is} cheap, I agree, then {it is} cheap'. =
   'I agree: I confirm that this is really cheap'.

4.6. The postpos-imm-INEXACT-reduplicative SSyntRel

As before, "**postpos-imm-INEXACT-reduplicative**" is a cover name. In Russian, we find the following particular **postpos-imm-INEXACT-reduplicative** SSyntRel.

4.6.1. The postpos-imm-INSTR-reduplicative SSyntRel

$$L_{(N)number}$$
○
│
postpos-imm-
INSTR-reduplicative   ⇔ $L_{(N)number, \boxed{NOM}}$ + $\mathbf{L'}_{(N)number, \boxed{INSTR}}$
↓
○
$\mathbf{L'}_{(N)number}$
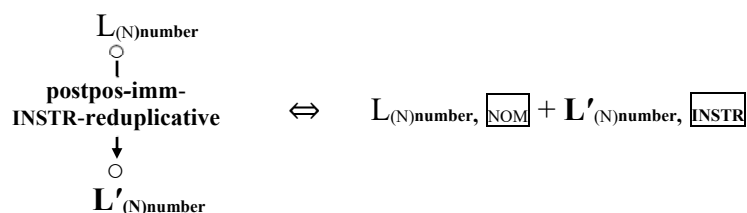
This SSyntRel serves the reduplicative syntactic idiom «REAL» [X], see (22):

(22)   a. *Ivan byl durak*–**postpos-imm-INSTR-reduplicative**→***durakom***
      'Ivan was a real fool'.
   b. *Ivan s vidu – durak **durakom*** 'In appearance Ivan is a real fool'.
   c. *Ivan sidel mračnyj, rasterjannyj, durak **durakom***
      'Ivan was sitting somber, confused, as a real fool'.
   d. *Ivan vël sebja durak **durakom*** 'Ivan was behaving as a real fool'.
   e. *Iz ètix škol deti vyxodjat duraki **durakami***
      'Kids graduate from these schools real fools'.

This idiom on four levels of linguistic representation, with 'X' = 'durak/fool':

SemS   : 'Ivan←**1**–<u>was</u>–**2**→fool←**1**–real'
DSyntS   : BYT′_{IND, PAST}—**II**→DURAK_{SG}—**ATTR**→«REAL»
SSyntS   :
BYT′_{IND, PAST}—**copular-completive**→DURAK_{SG}—**postpos-imm-INSTR-reduplicative**→DURAK_{SG}
DMorphS   : BYT′_{IND, PAST, SG, MASC} … DURAK_{SG, NOM} **DURAK_{SG, INSTR}**

Comment

This idiom has a complex syntactics: roughly, the reduplicand L(X) can depend only on a verb of the copula or quasi-copula type from a limited set (for instance, not on one of the standard Russian copulas—the verb JAVLJAT′SJA 'be'!) or on the fictitious lexeme «KAK» 'as' (for details, see Mel'čuk 2023b). This is,

however, not relevant for the description of the **postpos-imm-INSTR-reduplicative** SSyntRel.

4.7. The postpos-non.imm-exact-reduplicative SSyntRel

$$\text{L}_{\{G\}}$$
○
|
postpos-non.imm-
exact-reduplicative   ⇔ $\text{L}_{\{G\}} + \ldots + \mathbf{L'}_{\{g\}}$
↓
○
$\mathbf{L'}_{\{G\}}$
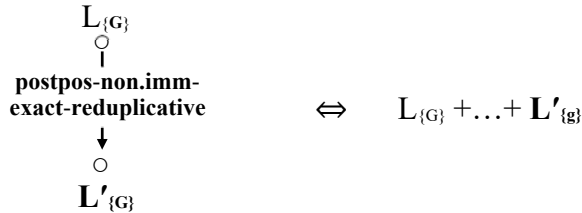
This SSyntRel is part of the signifier of the reduplicative syntactic idiom «I.INSISTING» [*on* X], see (23):

┌**post-non.imm-ex-redupl**┐
(23)   a. *Da prišla*$_{L(X)}$ *Maša,*   **prišla**$_{L'(X)}$*!*
        lit. 'But she.arrived Masha, she.arrived!' =
        'But Masha arrived, she did!'
      b. *Ivana*$_{L(X)}$ *ja vstretil,* **Ivana**$_{L'(X)}$
        lit. 'Ivan$_{ACC}$ I met, Ivan$_{ACC}$'. = 'It is Ivan whom I met, Ivan'.
      c. *Sup vegetarianskij*$_{L(X)}$, **vegetarianskij**$_{L'(X)}$
        lit. 'Soup is vegetarian, vegetarian'. =
        'The soup is vegetarian, don't doubt'.

The «I.INSISTING» idiom on four levels of linguistic representation, with 'X' = 'prišla/arrived':

SemS   : 'Masha←**1**–arrived←**2**–insist–**1**→I'
DSyntS   : MAŠA←**I**–PRIJTI$_{PERF, IND, PAST}$–**ATTR**→«I.INSISTING»
SSyntS   :
MAŠA←**subjectival**–PRIJTI$_{PERF, IND, PAST}$–**postpos-non.imm-exact-reduplicative**→PRIJTI$_{PERF, IND, PAST}$
DMorphS +
DMorph-ProsS: PRIJTI$_{IND, PAST, SG, FEM}$ MAŠA$_{NOM}$ | **PRIJTI$_{IND, PAST, SG, FEM}$**
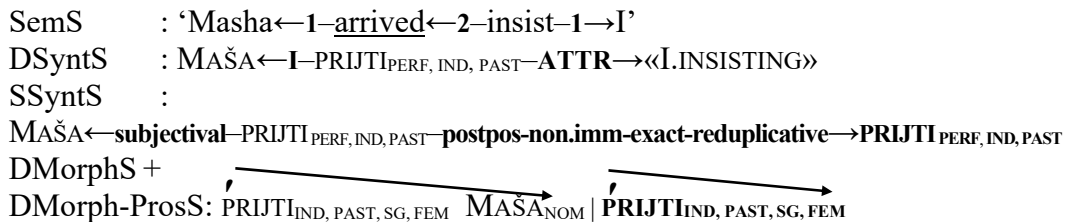
4.8. The postpos-non.imm-inexact-reduplicative SSyntRel

This SSyntRel does not exist in Russian.

5. Conclusions

Russian has seven reduplicative SSyntRels:

The **antepos-imm-DAT-reduplicative** SSyntRel (*vsem **goram** gora*)
The **antepos-imm-LONG.INSTR-reduplicative** SSyntRel (***belym**-belo*)
The **antepos-non.imm-exact-reduplicative** SSyntRel (***Karlik**-to Ivan karlik.*)

The **antepos-non.imm-INF-reduplicative** SSyntRel (***Perestroit'**-to ètot dom ona perestroila.*)

The **postpos-imm-exact-reduplicative** SSyntRel (*glupyj-**glupyj***)

The **postpos-imm-INSTR-reduplicative** SSyntRel (*durak **durakom***)

The **postpos-non.imm-exact-reduplicative** SSyntRel (*Prišla Maša, **prišla**.*)

These SSyntRels belong to the domain that L. Iomdin aptly baptized "microsyntax": 'syntactic phenomena intimately related to phraseology'; he has convincingly demonstrated its prime importance for linguistic theory (Iomdin 2008, 2010, 2020 and Avgustinova & Iomdin 2019). However, as of today, this domain still does not receive sufficient attention of researchers. The proposed set of Russian reduplicative SSyntRels is intended as a modest contribution to the project "Syntactic Typology: Surface-Syntactic Relations in the World Languages," which, hopefully, will be launched one day.

Acknowledgments

REFERENCES

Apresjan, Valentina. 2018. Russian constructions with syntactic reduplication of colour terms: A corpus study. *Russian Journal of Linguistics* 22 (3). 653–674. https://doi.org/10.22363/2312-9182-2018-22-3-653-674

Avgustinova, Tania & Leonid Iomdin. 2019. Towards a typology of microsyntactic constructions. In Gloria Corpas-Pastor & Ruslan Mitkov (eds.), *Computational and Corpus-Based phraseology*, 15–30. Cham: Springer.

Bolshakov, Igor. 2002. Surface syntactic relations in Spanish. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing. Third International Conference, CICLing 2002, Mexico City, Mexico, February 2002. Proceedings. Lecture Notes in Computer scienc*e, vol. 2276, 210–219. See also: https://link.springer.com/content/pdf/bfm%3A978-3-540-45715-2%2F1 (accessed 07 October 2022).

Gerdes, Kim, Bruno Guillaume, Sylvain Kahane & Guy Perrier. 2018. *SUD or Surface-Syntactic Universal Dependencies: An Annotation Scheme Near-Isomorphic to UD*. Universal Dependencies Workshop, Brussels. See: <hal-01930614>

Ghomeshi, Jila, Nicole Rosen, Ray Jackendoff & Kevin Russel. 2004. Contrastive focus reduplication in English (The Salad-Salad paper). *Natural Language and Linguistic Theory* 22 (2). 307–357.

Giljarova, Ksenija. 2010. Takaja devočka-devočka. Semantika reduplikacii suščestvitel'nyx v russkoj razgovornoj reči i jazyke interneta [Such a girl-girl. Semantics of noun reduplication in colloquial Russian and the Internet language]. *Kompjuternaja lingvistika i intellektual'nye texnologii* 9 (16). 90–96. (In Russ.)

Iomdin, Leonid. 2008. V glubinax mikrosintaksisa: Odin leksičeskij klass sintaksičeskix frazem [In the depths of microsyntax: A lexical class of syntactic idioms]. *Kompjuternaja lingvistika i intellektual'nye texnologii* 7 (14). 178–184. (In Russ.)

Iomdin, Leonid. 2010. Sintaksičeskie otnošenija [Syntactic relations]. In Jurij Apresjan, Igor Boguslavskij, Leonid Iomdin & Vladimir Sannikov, *Teoretičeskie problemy russkogo sintaksisa. Vzaimodeistvie grammatiki i slovarja* [*Theoretical problems of Russian syntax. Interaction between grammar and lexicon*], 21–43. Moskva: Jazyki slavjanskix kul'tur. (In Russ.)

Iomdin, Leonid. 2020. Ni pri čem: Ob odnoj mikrosintaksicheskoj konstrukcii s otricaniem [Ni pri čem: A microsyntactic construction with negation]. In Artëmij Kejdan (ed.), *Druzja, slova, tablicy. Sbornik statej v čest' 75-letija A.K. Polivanovoj*, 160–168. Moskva – Ekaterinburg: Kabinetnyj uchenyj. (In Russ.)

Israeli, Alina. 1997. Syntactic reduplication in Russian: A cooperative principle device in dialogues. *Journal of Pragmatics* 27 (5). 587–609.

Janda, Laura, Mihail Kopotev & Tore Nesset. 2020. Constructions, their families and their neighborhoods: The case of *durak durakom* 'a fool times two'. *Russian Linguistics* 44 (1). 109–127.

Kopotev, Mixail. 2008. *Principy sintaksicheskoj idiomatizacii* [*Principles of Syntactic Idiomatization*]. Helsinki: Helsinki University Press. (In Russ.)

Krjučkova, Ol′ga. 2004. Voprosy lingvističeskoj traktovki leksičeskoj reduplikacii v russkom jazyke [Questions of the linguistic description of lexical reduplication in Russian]. *Russkij jazyk v naučnom osveshchenii* 2 (8). 63–83. (In Russ.)

Marneffe, Marie-Catherine de, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford Dependencies: A cross-linguistic typology. *Proceedings of the Ninth International Conference on Language resources and Evaluation*, European Languages Resources Association (ELRA), 4585–4592.

Marneffe, Marie-Catherine de & Christopher Manning. 2008. The Stanford typed dependencies representation. In Johan Bos (ed.), *Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, Manchester: Coling 2008 Organizing Committee. See also: https://worksheets.codalab.org/rest/bundles/0x953afe5537074b4b9cd3c57e08e2d865/contents/blob/StanfordDependenciesManual.pdf

Mel'čuk, Igor. 1962. Ob algoritme sintaksičeskogo analiza jazykovyx tekstov (obščie principy i nekotorye itogi) [An algorithm for syntactic analysis of texts (general principles and some conclusions). *Mašinnyj perevod i prikladnaja lingvistika* 7. 45–87.

Mel'čuk, Igor. 1963. Avtomatičeskii analiz tekstov (na materiale russkogo jazyka) [Automatic text analysis (as applied to Russian)]. *Slavjanskoe jazykoznanie*. Moskva: Akademija Nauk SSSR. 477–509.

Mel'čuk, Igor. 1974. *Opyt teorii lingvističeskix modelej "Smysl ⇔ Tekst". Semantika, sintaksis* [*Outline of a "Meaning ⇔ Text" Linguistic Model Theory. Semantics, Syntax*]. Moskva: Nauka.

Mel'čuk, Igor. 1987. Un affixe dérivationnel et un phrasème syntaxique du russe moderne. Essai de description formelle. *Revue des études slaves* 59 (3). 631–648.

Mel'čuk, Igor. 2001. *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, Igor. 2012. Mestoimennye vyraženija s imenem čertyxatel′nym tipa [*Ona uexala*] *čёrt znaet kuda* i im podobnye v russkom jazyke [Pronominal expressions with a «blasphemous» noun of the [*Ona uexala*] *čёrt znaet kuda* type and similar ones in Russian]. *Russkij jazyk v naučnom osveščenii* 2 (24). 5–22.

Mel'čuk, Igor. 2012–2015. *Semantics: From Meaning to Text*. Vols 1–3. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, Igor. 2015–2016. A general inventory of surface-syntactic relations in world languages. Part one/Part two. *Moscow Linguistic Journal*, 2015, 17 (2). 75–103; 2016, 18 (1). 94–120.

Mel'čuk, Igor. 2016. *Language: From Meaning to Text.* Moskva/Boston: Academic Studies Press.

Mel'čuk, Igor. 2018a. Genitive adnominal dependents in Russian: Surface-syntactic relations in the N→N$_{GEN}$ phrase. *Voprosy jazykoznanija* 4. 25–46.

Mel'čuk, Igor. 2018b. "Wordlets": One of Zholkovsky's major contributions to the notion of deep-syntactic structure. In Dennis Ioffe, Marcus Levitt, Joe Peschio & Igor Pilshchikov (eds.), *A/Z: Essays in Honor of Alexander Zholkovsky*, 350–360. Boston: Academic Studies Press.

Mel'čuk, Igor. 2020. *Stat'ja kak stat'ja: pisat' tak pisat'!* O sintaksičeskix idiomax [*Stat'ja kak stat'ja: pisat' tak pisat'!* On syntactic idioms]. In Artëmij Keidan (ed.), *Druz'ja, slova, tablicy. Sbornik k 75-letiju A. K. Polivanovoj*, 259–278. Moskva – Ekaterinburg: Kabinetnyj učënyj. See also: http://olst.ling.umontreal.ca/pdf/Mel'čuk.2020.Stat´ja%20kak%20stat´ja.pdf

Mel'čuk, Igor. 2021a. Morphemic and syntactic phrasemes. *Yearbook of Phraseology* 12. 33–74.

Mel'čuk, Igor. 2021b. *Ten Studies in Dependency Syntax.* Berlin: De Gruyter Mouton.

Mel'čuk, Igor. 2023a. *General Phraseology: Theory and Practice.* Amsterdam/Philadelphia: John Benjamins. *To appear.*

Mel'čuk, Igor. 2023b. The Russian reduplicative construction of the *gudit-gudit 'It rumbles-rumbles'* type: A syntactic idiom. *To appear.*

Mel'čuk, Igor & Jasmina Milićević. 2020. *An Advanced Introduction to Semantics. A Meaning-Text Approach.* Cambridge: Cambridge University Press.

Poiret, Rafaël & Haitao Liu. 2019. Les dépendants adnominaux prépositionnels en français: Relations syntaxiques de surface dans le syntagme N→SP. *Français Moderne* 2. 259–280.

Sannikov, Vladimir. 2008. *Russkij sintaksis v semantiko-pragmatičeskom prostranstve* [*Russian Syntax in Semantic-Pragmatic Space*]. Moskva: Jazyki slavjanskix kul'tur.

Sannikov, Vladimir. 2010. Konstrukcii s tožestvennymi slovoformami [Constructions with identical wordforms]. In Jurij Apresjan, Igor Boguslavskij, Leonid Iomdin & Vladimir Sannikov (eds.). *Teoretičeskie problemy russkogo sintaksisa. Vzaimodeistvie grammatiki i slovarja* [*Theoretical Problems of Russian Syntax. Interaction between Grammar and Lexicon*], 191–208. Moskva: Jazyki slavjanskix kul'tur.

Appendix 1: Some important linguistic notions mentioned in this paper

Clausative

The clausative is a part of speech whose elements are syntactically full clauses; e.g.: *Yes! | Down* [*with the virus*]*! | Plop! | For Heaven's sake!*

Fictitious lexeme

A fictitious lexeme is an artificial lexeme introduced by the researcher in order to represent—in the lexicon and in the DSyntS—either a meaningful SSynt-relation or a syntactic (= non-segmental) idiom (see Mel'čuk 2018b). Fictitious lexemes are enclosed in angular brackets «…». For instance, the fictitious lexeme «VERY» encodes the Russian syntactic idiom implemented by adjectival reduplicative phrases, such as *bol'šoj-bol'šoj* lit. 'big-big' = 'very big'.

As any lexeme, a fictitious lexeme has its lexical entry with a lexicographic definition, a government pattern, etc.: see the lexical entry for the fictitious lexeme [X] «WILL.PUNISH» [Z for Y-ing], Section 3, p. 891. It is, of course, the lexical entry of the corresponding idiom.

Grammeme

A grammeme is a value of an inflectional category; for instance, in English, the category of number has two grammemes: SG ~ PL.

Deep(-syntactic) grammemes are all semantically full grammemes characterizing a given lexeme in a DSyntS; surface(-syntactic) grammemes are only those semantically full grammemes that are expressed synthetically, or morphologically, i.e. inside a wordform, rather than analytically, by grammatical lexemes. Thus, to represent the phrase *had been working* in a DSyntS the verbal lexeme WORK$_{(V)}$ has the set of deep grammemes IND, PERF, PROGR, PAST:

$$\text{WORK}_{(V)\text{IND, PERF, PROGR, PAST}}$$

Its surface-syntactic correspondence is WORK$_{(V)\text{PPRES}}$ (*working*), the grammemes IND, PERF, PROGR and PAST being expressed by the forms of the auxiliary verbal lexemes HAVE and BE.

Surface-syntactic relation [SSyntRel]

A surface-syntactic relation **r** is a direct syntactic dependency link between two lexemes L$_1$ and L$_2$ in an SSynt-structure: L$_1$–**r**→L$_2$, such that **r** fully specifies L$_1$'s and L$_2$'s mutual linear position in the deep-morphological structure and their surface-syntactic grammemes, if any. (See, e.g., Mel'čuk 2021b: Ch. 2, Section 3.) SSyntRels are language-specific.

Surface-syntactic structure [SSyntS]

The surface-syntactic structure of an utterance is a tree whose nodes are labeled with the lexemes of the utterance (each lexeme being supplied with all its SSynt-grammemes) and the branches, with the corresponding SSyntRels. For instance, the Russian sentence (24a) and its SSyntS:

(24) a. *Sup kipel-kipel i vykipel*
       lit. 'Soup was.boiling-was.boiling and boiled.away'. =
       'The soup was boiling for too long and finally boiled away'.
     b.
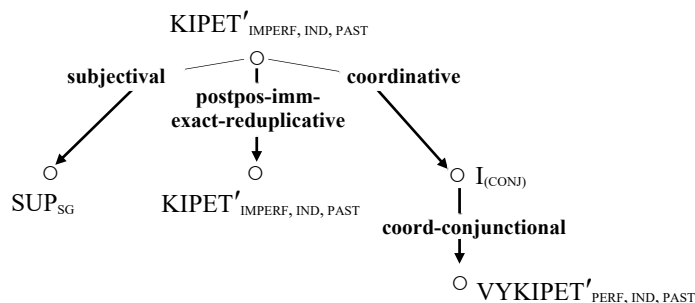
Appendix 2: Russian surface-syntactic relations mentioned in this paper

The reduplicative SSyntRels are not included in this list.

approx(imate)-quant(itative)	: *tonn*–**approx-quant**→*desjat'* 'maybe 10 tons'
circum(stantial)	: *spal*–**circum**→*spokojno* 'slept quietly'
	prijti–**circum**→*v pjatnicu* 'come on Friday'
subj(ecive)-compar(ative)-	
conj(unctional)	: *sil'nee, čem*–**subj-compar-conj**→*Ivan*
	'stronger than Ivan is'
coord(inative)	: *Ivan*–**coord**→*i Maša* 'Ivan and Masha'
coord-conj(unctional)	: *Ivan i*–**coord-conj**→*Maša* 'Ivan and Masha'
cop(ular)-compl(etive)	: *Oni byli*–**cop-compl**→*bol'ny* 'They were ill'.
	*Ivan Ø*ᴮʸᵀ′–**cop-compl**→*karlik* 'Ivan is a.dwarf'.
	byt'–**cop-compl**→*drugom* 'be a.friend'
dir(ect)-obj(ectival)	: *čitat'*–**dir-obj**→*romany* 'read novels'
indir(ect)-obj(ectival)	: *byt'*–**indir-obj**→*Ivanu drugom* lit. 'be to.Ivan a.friend'
modif(icative)	: *dobryj*←**modif**–*drug* 'good friend'
prepos(itional)	: *prijti v*–**prepos**→*pjatnicu* 'come on Friday'
	daleko ot–**prepos**→*goroda* 'far from the.city'
quant(itative)	: *desjat'*←**quant**–*tonn* '10 tons'
restr(ictive)	: *tol'ko*←**restr**–*pil* 'only drank'
	ne←**restr**–*pil* lit. 'not drank'
	Ivan–**restr**→*že* lit. 'Ivan as.for'
subj(ectival)	: *Oni*←**subj**–*byli bol'ny* 'They were ill'.
	Ivan←**subj**–*Ø*ᴮʸᵀ′ *karlik* 'Ivan is a.dwarf'.

Appendix 3: Syntactic idioms mentioned in this paper

«ABNORMALLY» [X] (*Sup kipel-**kipel** i vykipel*
               lit. 'Soup was.boiling-was.boiling and boiled.away'.)

[X Y-*u*] ˹ʙʏᴛ′ ɴᴇ ᴠ L′(X)˺ (*Nam prazdnik ne v **prazdnik***
                  lit. 'To.us feast is not into feast'.)

[X] ˹ʙʏᴛ′ ᴠsᴇᴍ L′(Y)-ам˺ [Y] (*Èto vsem **goram** gora*
                    lit. 'This is to.all mountains mountain'.)

«COMPLETELY» [X] (***černym**-černy* lit. '{are} by.black-black')

˹Čᴛᴏ [X], ᴛᴏ L′(X)˺ (*Čto Ivan umnyj, to umnyj*
               lit. 'That Ivan is smart, then {he} is smart'.)

«I.CONFIRMING» [*that* X] (***Karlik**-to Ivan karlik* lit. 'Dwarf-as.for he is dwarf'.)
                 ***Spat'**-to Ivan spal* lit. 'To.sleep-as.for Ivan was.sleeping'.)

«I.INSISTING» [*on* X] (*Mebel' vynosite, **mebel'***! lit. 'Furniture take.out, furniture!')

«IN.THE.STRICT.SENSE» [X]
   (English: *Not **talk**-talked.*
   Russian: *Moskva-**Moskva**, a ne Moskva Tovarnaja* lit. 'Moscow-Moscow, and not
        Moscow Tovarnaya'.**)**

«REAL» **[X]** (*durak **durakom*** lit. 'fool by.fool')

[X] ˊтак L′(X)ˋ (*Piva tak piva!* lit. 'Of.beer then of.beer!')
«VERY» [X] (*xolodnyj-xolodnyj* lit. 'cold-cold')
«VERY-VERY» [X] (*xolodnyj-prexolodnyj* lit. 'cold-overcold')
[X] «WILL.PUNISH» [Z for Y-ing] (*Ivan tebe potancuet!* lit. 'Ivan to.you will.dance!')

Bionote:
Igor MEL'ČUK, Professor Emeritus at the University of Montreal, Canada, is the author of 49 books and 305 scientific papers. His research interests include general linguistics (with special attention to its conceptual apparatus and terminology), semantics (with special attention to lexicology and lexicography), syntax (with special attention to dependency structures), and morphology; his work is based mostly on Russian and French, but concerns also English and several other languages (Spanish, Hungarian, Lezghian, Alutor, Dyirbal, Bafia, Kirundi, Korean). He is the proponent of the Meaning-Text approach to natural language and one of the pioneers of Machine Translation—since 1954.
e-mail: igor.melcuk@umontreal.ca
https://orcid.org/0000-0002-4520-0554

Сведения об авторе:
Игорь Александрович МЕЛЬЧУК – заслуженный профессор Монреальского университета (Канада) и член Королевского общества Канады, автор 49 книг и 305 статей. Его научные интересы включают общую лингвистику (в особенности ее понятийный аппарат и терминологию), семантику (в особенности лексикологию и лексикографию), синтаксис (в особенности структуры зависимостей) и морфологию. Работы И.А. Мельчука основываются, в первую очередь, на данных русского и французского языков, но затрагивают также английский и ряд других языков (испанский, венгерский, лезгинский, алюторский, дьирбал, бафия, рунди, корейский). И.А. Мельчук является автором лингвистического подхода «Смысл–Текст» и одним из пионеров машинного перевода – с 1954 года.
e-mail: igor.melcuk@umontreal.ca
https://orcid.org/0000-0002-4520-0554

Research article / Научная статья

I and Thou: Universal human concepts present as words in all human languages

Anna WIERZBICKA ⓘD ✉

Australian National University, Canberra, Australia
✉ anna.wierzbicka@anu.edu.au

Abstract

This paper argues that "YOU" and "I" ("I" and "THOU") are fundamental elements of human thought, present as distinct words (or signs) in all human languages. I first developed this thesis in my 1976 article "In defense of YOU and ME" (and before that, introduced it in my 1972 book *Semantic Primitives;* cf. also my 2021 article "'Semantic Primitives', fifty years later"). Since then, it has been confirmed by wide-ranging cross-linguistic investigations conducted in the Natural Semantic Metalanguage (NSM) framework. But neither the truth of this thesis nor its importance have become widely recognised in linguistics or anthropology. Influential scholars in both these fields continue to undermine the notion of the fundamental unity of humankind and to put total emphasis, instead, on the diversity of languages and cultures. As cross-linguistic investigations of the last fifty years show, however, despite the phenomenal diversity of human languages a shared "alphabet of human thoughts" was not just a figment of Leibniz's imagination but a fitting metaphor for something real and immeasurably important. As the present article aims to show, "YOU" and "I" ("I" and "THOU") are two twin cornerstones of this reality. To quote the entry on "Psychic unity of humankind" in the *Encyclopedia of Anthropology*, "Ineluctably, the idea [of a deep psychological unity of humankind] has ethical significance. For attempting to inform humans about what they are and what they have in common is not a neutral act" (Prono 2006). As the present article seeks to demonstrate (and as Martin Buber compellingly affirmed a century ago), "I" and "THOU" are an ineluctable part of who we are: how we think, how we speak and how we relate to others.

Keywords: *semantic universals, universal semantic primes, Natural Semantic Metalanguage, the concepts of "I" and "THOU", sign languages, psychic unity of humankind*

Я и ты: универсальные концепты, существующие во всех языках мира

Анна ВЕЖБИЦКАЯ ⬤ ✉

Австралийский национальный университет, Канберра, Австралия
✉anna.wierzbicka@anu.edu.au

Аннотация

В статье высказывается идея о том, что «Вы» и «Я» («Я» и «Ты») – это базовые элементы человеческого мышления, которые присутствуют как отдельные слова (или знаки) во всех языках мира. Впервые это положение прозвучало в (Wierzbicka 1972, 1976, см. также Wierzbicka 2022). С тех пор оно было неоднократно подтверждено разноплановыми кросс-лингвистическими исследованиями, проведенными в рамках Естественного Семантического Метаязыка (ЕСМ). Но справедливость этого положения и его значимость не были признаны ни в лингвистике, ни в антропологии. Влиятельные ученые, представляющие данные направления, продолжают подвергать сомнению фундаментальное единство человечества и вместо этого абсолютизируют разнообразие языков и культур. Однако, как показывают кросс-линг-вистические исследования, проведенные в последние пятьдесят лет, несмотря на феноменальное разнообразие человеческих языков общий для них «алфавит человеческой мысли» – это не просто плод воображения Лейбница, а метафора, подходящая для чего-то реального и чрезвычайно важного. Цель настоящей статьи – показать, что «Вы» и «Я» («Я» и «Ты») – два краеугольных камня этой реальности. Как говорится в статье "Psychic unity of humankind" в «Энциклопедии антропологии» (*Encyclopedia of Anthropology*), "идея [глубинного психологического единства человечества] безусловно имеет этическую значимость. Поскольку показать людям их сущность и то, что их объединяет, – это не нейтральный акт" (Prono 2006). Как демонстрирует данная статья (и что убедительно доказано Мартином Бубером сто лет назад), «Я» и «ТЫ» – неотъемлемая часть того, кто мы есть: как мы мыслим, как мы говорим и как взаимодействуем с другими.

Ключевые слова: *семантические универсалии, универсальные семантические примитивы, Естественный Семантический Метаязык, понятия «Я» и «ТЫ», жестовые языки, духовное единство человечества*

Dedication

I am dedicating this study to my dear friend Igor Mel'cuk, whom I first met at the International Congress of Slavists in Sofia exactly sixty years ago. It was then that I first heard the term "yazyk posrednik" ("language mediator"). The way Mel'cuk and his colleagues thought of what a semantic "language mediator" should look like was different from how my colleagues and I think of it, but in essence this is what NSM is: a semantic language which can act as a mediator between different languages.

A key difference is that our language mediator is carved out of natural languages and therefore looks like a natural language (in miniature), whereas Igor

and his colleagues saw their "yazyk posrednik" as an artificial, or semi-artificial language. All the same, it is a joy to acknowledge the historical connection between the idea of a "yazyk posrednik" ("language mediator") and the idea of a Natural Semantic Metalanguage (NSM).

It is also a joy to acknowledge that my life-long interest in semantics first developed during my extended stay in Moscow in 1964–1965, where I was fortunate to be able to work with, and learn from, Igor and his closest colleagues, Aleksandr Zholkovski and Jurij Apresjan. As discussed by Apresjan in his essay "On the Language of Explications and Semantic Primitives" (2000), in the early 1960s the Moscow Semantic School, that is Igor Mel'cuk and his colleagues, were also pursuing the notion of "semantic primes" in their work. Their approach was different from that of my Polish mentor Andrzej Bogusławski (see Bogusławski 2003[1963]) and of what Apresjan called the Polish Semantic School, but the two had a great deal in common. I feel immeasurably indebted to Igor and his colleagues in the Moscow Semantic School.

Most importantly, perhaps, I caught from them a fascination with words and meanings, a double focus reflected fifty years later in the title of a book co-authored with my closest Australian collaborator and partner in NSM Cliff Goddard, *Words and Meanings* (2014). In this paper, I want to honour Igor at the beginning of the tenth decade of his life, a life dedicated largely to words and meanings, and also, to *chelovechnost'*.

Dear Igor, this paper is for YOU from ME, with love.

1. Introduction: What is at stake?

Nearly fifty years ago, in a paper entitled "In defense of YOU and ME", I wrote:

> What is the status of the notions "you" and "I" in human thinking? Can they be reduced to, explained in terms of, certain other notions or are they among those notions which are so basic and so clear of themselves that any attempt at explaining or further analysing them must be judged as futile and absurd (see **Descartes 1642**, Pascal 1958, **Arnauld & Nicole 1662**)? Are they or are they not semantic primitives, i.e. essential, irreducible elements of our 'mental language'? (Wierzbicka 1976: 4)

As the title of that paper indicates, I was defending the thesis that YOU and ME (or I and THOU) are essential and irreducible elements of human thought and language. In mounting that defense, I noted that "you" and "I" as semantic primitives had found a splendid advocate in the person of the Danish structuralist Holger Steen Sorensen, who more than a decade earlier firmly declared "'I' and 'you' are indefinable", adding: "They are consequently semantic primitives of the English language." (Sorensen 1963: 96). As I noted at the time, however, Sorensen, the pioneer in the search for semantic primitives, true to the spirit of structural linguistics did not venture to conceive that the set of "semantic primitives of

English" could be identical with that of every other human language. Going beyond Sørensen, I argued just that.

At that time, in the mid-1970s, writing in a new historical context, I was seeking to rebut, above all, two ideas widespread in linguistics and philosophy of that time: first, that the meanings of "I" and "thou" (you and me) should be interpreted as "the speaker" and "the addressee"; and second, that the meaning of these elements should be represented by means of referential indices (promulgated by "generative semanticists" such as James McCawley and George Lakoff). I concluded:

> "I" and "you" are semantic primitives. They cannot be defined away in terms of other primitives and they cannot be dispensed with in favour of some arbitrary indices. The idea advanced by McCawley, George Lakoff, and others, that semantic primitives can be thought of as "atomic predicates" because everything in the semantic representation which is not a predicate is a referential index, is mistaken. <…> [I]t has not been arrived at by empirical semantic research but simply assumed a priori, on the model of the artificial language of symbolic logic. Empirical semantic studies suggest that there are semantic primitives which function as arguments, not as predicates. "You" and "I" are among their number (Wierzbicka 1976: 19).

Today, after almost fifty years of extensive cross-linguistic (as well as intra-linguistic) investigations undertaken by many scholars working within the NSM framework, the evidence in support of this conclusion seems to me overwhelming[1].

It seems also overwhelmingly clear that this is not a technical issue, relevant to linguistics and philosophy but without any broader implications. On the contrary, the place of the concepts 'I' and 'THOU' ("you" and "I") in human thought and in human languages is an essential aspect of who we are.

The goal of this paper is to demonstrate both the truth and the importance of the thesis that, as evidence indicates, "I" and "THOU" are present, as words (or signs) in all human languages.

---

[1] The acronym NSM stands for the Natural Semantic Metalanguage. Scholars working in the NSM framework believe that through decades of sustained cross-linguistic and intra-linguistic investigations, they have identified the complete inventory of simple and universal concepts that are embedded in the lexicon of all human languages. As Cliff Goddard and I put it in our 2014 book *Words and Meanings* (pp. 11–12)*,* we claim that "a plausible, stable, and well-evidenced set of 'universal words' have been identified and that this can provide the seemingly solid foundation for the prospect of decoding meanings across languages, <…> These putatively indefinable word-meanings are known as semantic primes and they are 65 in number" (See Wierzbicka 1972, 1980, 1996, 2021, Goddard 2011, Goddard (ed.) 2018, 2021, Goddard & Wierzbicka (ed.) 2002, Gladkova & Larina 2018).

As for the status of "YOU" (sg) and "I" as universal semantic primes, they are well attested, as distinct lexical elements, in every known human language. For a rebuttal of the claims that languages like Thai and Japanese lack personal pronouns, see Diller 1994 and Onishi 1994, in (Goddard & Wierzbicka 1994).

2. The concept of 'I' and moral reflection

According to the primatologist Frans de Waal (2006), a key difference between humans and (other) animals lies in people's capacity for self-reflective reasoning and judgement. Discussing de Waal's view in his 2013 book *The Gap*, evolutionary scientist Thomas Suddendorf strongly agrees, adding: "we can consider the long-term consequences of our actions. We are the only species on that planet with the foresight capable of plotting a path towards a desirable future" (Suddendorf 2013: 283). Suddendorf also cites Charles Darwin's assessment:

> A moral being is one who is capable of reflecting on his past actions and their motives – of approving of some and disapproving of others; and the fact that man is the one being who certainly deserves this designation, is the greatest of all distinctions between him and the lower animals (Darwin 2003[1871]: 610).

Trying to unpack Darwin's, and de Waal's, conclusions about "the gap" in simple words, we can identify the following clear ideas:

> people can think like this about something:
>   when I did this, I did something bad,
>   when I did this I did something good

If we want to similarly unpack Suddendorf's ideas about "the gap", we can say:

> people can think like this about something:
>   if I do this, some time after this something bad can happen because of this
>   if I do this, some time after this something good can happen because of this

To be able to conceive such "self-reflective judgments" and to consider possible long-term consequences of our actions we need to have at our disposal a handful of basic concepts to think with, including "do", "happen", "good", "bad", "if", "because", "before", after", and – crucially – "I". But do all languages have conceptual and lexical resources which would allow their speakers to conceive and express such thoughts? In particular, do they all have a concept expressed in English in the word *I* (or *me*)? And do they all have a word (equivalent to the English *I*), which would enable a person to keep this concept in focus for some time, while reflecting on his or her past actions (or planning for future ones)?

If Darwin, de Waal and Suddendorf are right, the matter is of "unspeakable importance" (to use David Hume's words), as clearly recognised by a number of thinkers who thought about language deeply and considered languages both in their diversity and in their fundamental unity – such as Wilhelm Humboldt, Franz Boas, and Emile Benveniste. To quote just one of them, Benveniste (1971: 224):

> Consciousness of self is only possible if it is experienced by contrast. I use *I* only when I am speaking to someone who will be a *you* in my address. It is this condition of dialogue that is constitutive of *person*, for it implies that reciprocally *I* becomes *you* in the address of the one who in his turn designates himself as *I*. Here we see a principle whose consequences are to spread out in

all directions. Language is possible only because each speaker sets himself up as a *subject* by referring to himself as *I* in his discourse. Because of this, *I* posits another person, the one who, being, as he is, completely exterior to "me", becomes my echo to whom I say *you* and who says *you* to me. This polarity of persons is the fundamental condition in language, of which the process of communication, in which we share, is only a mere pragmatic consequence.

Given the importance of the relationship between 'I' and 'you' as a condition of language, it is hardly surprising that thinkers such as Humboldt, Boas and Benveniste were all keenly interested in the question of whether 'you' and 'I' were clearly recognised in all languages. Again, I will only quote Benveniste (1971: 225):

> It is a remarkable fact—but who would notice it, since it is so familiar?—that the "personal pronouns" are never missing from among the signs of a language, no matter what its type, epoch, or region may be. A language without the expression of person cannot be imagined. It can only happen that in certain languages, under certain circumstances, these "pronouns" are deliberately omitted; this is the case in most of the Far Eastern societies, in which a convention of politeness imposes the use of periphrases or of special forms between certain groups of individuals in order to replace the direct personal references. But these usages only serve to underline the value of the avoided forms; it is the implicit existence of these pronouns that gives social and cultural value to the substitutes imposed by class relationships.

Broad typological surveys of many languages such as Ingram's (1978) study "Personal pronouns" confirm the view that personal pronouns "are never missing from among the signs of a language, no matter what its type, epoch, or region may be". In particular, such surveys confirm that while there can be a good deal of variation concerning the conceptualisation of human *groups* (reflected in words comparable to *we*, see Goddard 1995, Goddard & Wierzbicka 2021); and of persons spoken of (reflected in words comparable to *he* and *she*), words matching 'I' and 'you' ('thou') can be found in all studied languages.

This doesn't stop some linguists, however, from raising doubts about it in a surprisingly cavalier manner (as if it were a matter of no great significance). For example, in their wholesale attack on "the myth of language universals" Evans and Levinson (2009) write: "There are languages without tense, without aspect, without numerals, or without third-person pronouns (or even without pronouns at all, in the case of most sign languages)." (Evans & Levinson 2009: 435)

3. I and you (thou) in sign languages

The fact that many languages don't have words like "he" and "she" (so-called third person pronouns) is well known and is no obstacle to what a language can express, because one can always say "this someone" instead of "he" or "she". But the availability of words for "I" and "you" (thou) is of fundamental importance,

because there are no phrases which could be substituted for these two words without a change in meaning. To begin with I (the so-called first-person pronoun), there is no expression which could be substituted for I in sentences like "when I did this, I did something bad" (cf. Sorensen 1958, Lyons 1977). Accordingly, if a language didn't have a word meaning "I", the speakers could not express the basic self-reflective judgment "when I did this, I did something bad".

The point is dramatised by Evans & Levinson (2009) reference to sign languages. They provide a table titled "Every language has X, doesn't it?". Proposed substantive universals (from Pinker & Bloom 1990) supposedly common to all languages" which list eight points (one of them relating to pronouns). Evans & Levinson (2009) reject all the universals included in the table: "There are clear counterexamples to each of these claims" (Evans & Levinson 2009: 431) Commenting on the point relating to pronouns they write: "Some Southeast Asian languages lack clear personal pronouns, using titles (of the kind "honorable sir") instead, and many languages lack third-person pronouns <...>) Sign Languages like ASL (American Sign Language) also lack pronouns, using pointing instead." (Evans & Levinson 2009: 431).

Apart from the Southeast Asian languages (which are mentioned in the quote from Benveniste and to which I will return shortly), the rejection of first- and second-person pronouns ("I" and "you") as lexico-semantic universals hinges, above all, on sign languages. The argument goes as follows: even if all spoken languages had words like "you" and "I", these words could not be regarded as language universals because many sign languages don't have such words.

The logic of this argument seems odd. Speaking for the moment about "I" alone, if "I" is necessary for reflective self-judgment, then either sign languages do have "I" (like all spoken languages do) and can express such judgments or they don't have the same expressive power as spoken languages. The authors of "The myth of language universals", however, appear to want to have it both ways: they claim that many sign languages don't have a sign for "I", and at the same time insist that they have the same expressive power as spoken languages. Thus, in a section entitled "The challenge of sign languages" they write:

> Many proposed universals of language ignore the existence of sign languages – the languages of the deaf, now recognized to be full-blown languages of independent origin. <...> When due allowance is made for the manual-visual interface, sign languages seem to be handled by the same specialized brain structures as spoken ones, with parallel aphasias, similar developmental trajectories (e.g. infants "babble" in sign), and similar processing strategies as spoken languages <...> The neurocognition of sign does not look, for example, like the neurocognition of gesture, but instead recruits, for example, auditory cortex. (...) These results show that our biological endowment for language is by no means restricted to the input/output systems of a particular modality (Evans & Levinson 2009: 438).

But how could sign languages be "full-blown languages" (in a conceptual sense) if they didn't have signs for "you" and "I" and if the closest they could get

to the idea of "I" would be to use the manual equivalent of the expression "this someone"? Replacing "I" and "you" with "this someone" (accompanied by pointing) could do at the level of communication epitomised by the statement "Me Tarzan you Jane", but it would hardly do for the purposes of ethical reflection that Darwin was, so tellingly, concerned about.

Translated into a conceptual language in which "I" would be rendered as "this someone", Darwin's statement about people would read like this:

> people can think like this about someone: "this someone did something bad"
> (animals cannot think like this)

This, however, would be tantamount to saying that people can pass moral judgment on (other) people, not that they can pass moral judgment on themselves: for this latter, more demanding, act of reflection they would need a word (or sign) meaning "I".

The point is really quite simple and can be illustrated with mundane examples of reported speech. For example, if Mary says (speaking of John): "he said to me: I did it", the *referents* of "me" and "I" are different (the referent of "me" is Mary, and that of "I", John), but the *meaning* is the same in both cases: "I" ("me"). The sentence cannot be paraphrased as "he said to this someone: this someone did it". To convey the intended meaning, the speaker needs a word (or sign) meaning "I".

Unfortunately, dictionaries of sign languages are often misleading in their entries for "I". For example, the otherwise impressive dictionary of Auslan (Australian Sign Language) edited by Trevor Johnston (1998) defines one of its signs (pointing with the index of the right hand to the middle of one's chest) as follows: "Reference. Used to refer to the signer by the signer. (...) English = I, me." But clearly, when a signer says "she said: it is me" using the sign in question, he or she is not using it to refer to the signer. Rather, the gesture is a conventional rendering of the meaning 'I', and it cannot paraphrased as "this someone" or "this person". Evidently, the dictionary's description of the meaning of the sign pointing to one's chest as referring to the signer does not take into account the use of this sign in reported speech. The only possible description which would fit both direct speech and reported speech is one which interprets this sign as meaning 'I', not 'this person'.

Or consider the basic Jewish prayer" "Shema Israel", which religious Jews recite twice daily, and which Jewish children are expected to say before going to sleep. The prayer ends with the words: "And the Lord spoke to Moses, saying: I am the Lord, your God, who led you from the land of Egypt to be a God to you". How could a deaf person recite such a prayer in a sign language if they only had a sign meaning 'this person' and didn't have one meaning "I"? It seems hardly necessary to add that not even God can point at God. The only way a signer can convey the meaning of the words attributed to God in "Shema Israel" is first to establish that it is God who is speaking and then to use a sign meaning 'I'.

In other words, a sign with which the signer points to his or her own chest means, conventionally, "I", not "this person", and this is why it can be used in

quoting any statements in which God (or anybody else) says "I". So when, for example, God says to Moses (out of the burning bush): "Moses, Moses", and Moses replies: "Here I am", God is not saying to Moses: "I want to say something to someone called Moses", and Moses is not saying to God "this someone is here". Furthermore, when God continues by saying "I am the God of Abraham, the God of Isaac, and the God of Jacob", he (being invisible) is not pointing to himself and meaning "this someone". The notion of 'I' is essential to the exchange, both in relation to Moses and to God.

To cite one more example from a classic text (Matthew's Gospel, Mt 14:27), when Jesus walks on the water and the disciples take him for a ghost and are frightened, Jesus tells them: "Be of good cheer; it is I, be not afraid" (KJV). To paraphrase these words as "be not afraid, it is this someone" would make no sense: they can only be understood through a word, or a sign, meaning "I".

To reiterate the main point: either sign languages have the same expressive power as spoken languages or they don't. If they didn't have a sign meaning "I", they wouldn't be able to convey meanings such as those expressed in reported speech. Nor would they be able to carry "self-reflective judgment" ("she thinks like this: 'I did something bad'") singled out by Darwin as a key difference between humans and animals. In fact, evidence suggests that sign languages are fully equipped to express such meanings – but only because they do have signs meaning "I" and because they do allow for what is known in the literature on sign languages as "role shift" (cf. e.g. Goswell 2014).

In his *Penseés*, Pascal wrote: "Le moi est haïssable", "the I is repugnant" ("self is hateful", Pascal 1958 online). Can this idea be expressed in sign languages? For example, could it be explained through signing to deaf students in a school for the deaf? Presumably, not in one sentence. Once we recognise, however, that Pascal's "thought" is highly compressed and that for its sense to be fully grasped it needs to be "translated" into several sentences, it becomes clear that it is not beyond the power of sign languages to express it. As the passage in which this pithy saying is embedded suggests, the idea can be unpacked along the following lines:

> people often think like this about something: "I want this, this can be good for me"
> often, when they think like this, they don't think about other people
> this is very bad

Recent literature on sign languages and pronouns makes it abundantly clear that certain signs function in sign languages as "shifters", and that a signer can easily attribute thinking in terms of "me (I)" to someone other than him- or herself. For example, Cormier et al. (2013) cite the following two sentences, one in British Sign Language (BSL) and the other in English:

a. BOY SAY (LOOK-FOR ME)$_{NS:body}$ (BSL)
b. The boy said, "Are you looking for me?" (English)

Given the secure finding that sign languages do allow reported speech and "role shifting", it is in fact puzzling that the availability of "I"-based thinking in

these languages can still be questioned or denied, and that the Evans & Levinson (2009) claims on the subject have not been clearly and vigorously opposed in the relevant literature. This is not to say that these claims have not been challenged at all, but unfortunately the discussion which they gave rise to has usually been framed in technical specialist language rather than in plain "human" terms. Thus, more often than not the main question addressed is this: are the counterparts of "I" and "you" in sign languages pronouns (in the full sense of the word 'pronouns') or are they not? Since there is no God-given definition of the word 'pronoun', the issuing discussion can slide into points of grammatical terminology and grammatical theory (where linguists can be bitterly divided), rather than addressing the conceptual and semantic aspects of Evans's & Levinson's (2009) claims.

For example, Cormier et al. (2010: 2665) write in their abstract: "we discuss E&L's claim that sign languages lack pronouns", and later argue that "there is evidence for considering these pointing signs to be pronouns in sign languages, not only pointing gestures". In a later paper, Cormier et al. (2013) discuss Evans's & Levinson's (2009) claims about sign languages and personal pronouns again, and again focus on the technical aspects of the question. They conclude that "pronominal signs clearly share canonical properties of both pronouns and pointing gestures" (ibid: 243) and that these signs "cannot be characterised exclusively either as pronouns or as pointing gestures" (ibid: 244). But although they profess a keen interest in the question about "real universals of human language" (ibid: 244), they do not take a clear stand on what I see as the central issue: do sign languages have signs meaning "you" and "I" or not?

It seems clear that the authors don't want to accept that the sign pointing to one's chest means simply "this someone" and would prefer to allow that it also means "I". But since their major frame of reference lies in syntax and morphology and not in semantics, they do not formulate their position in semantic terms, and as a result, they equivocate: from a semantic point of view, a word, or a sign, cannot mean both 'this someone' and 'I' – it is either one or the other.

On this point, Evans & Levinson (2009) seem to be more consistent: they explicitly reject semantic universals along with syntactic and morphological ones, and, accordingly, they reject "I". It is worth mentioning in this context that in a seminar at the Australian National University, Evans (2014) appeared to be retreating from the strong anti-universalist position taken by E&L (2009), when he said that "I is I, in all languages". When I questioned him subsequently on this apparent retreat from the Evans's & Levinson's (2009) position, he replied (in a personal email) that what he said was an overstatement, and later elaborated:

> …certainly this is a place where I would be close to seeing universality, but I'm not convinced this is the case for all sign languages, where the argumentation turning on whether pointing to oneself is conventionalised 'I' vs just pointing at someone who happens in this case to be the speaker is a subtle one so I would not take it as read (27/7/2014, quoted by permission)

I agree with Evans that this is indeed the choice we have to make: either we accept that sign languages have a sign meaning 'I', or we argue that they don't and that they have to make do with a sign meaning 'this someone'. As I see it, however, this second option does not take into account situations where the signer is reporting someone else's words. For example, in a signed version of the sentence "Be not afraid: it is I" the conventional sign of pointing at one's chest performed by the signer evidently refers to the person whose words are being reported, not to the one who happens in this case to be the signer.

4. I and you (thou) in spoken language

Assuming then that the issue of sign languages as lacking personal pronouns in general and a word for 'I' in particular is really a red herring, let us return to the more obvious question concerning spoken languages. Do they all have words for "you" and "I"?

In his celebrated paper "Cultural constraints on Pirahã grammar", which presented Pirahã as lacking words like ONE, TWO and ALL, Daniel Everett (2005) did not go as far as claiming that Pirahã had no words for "you" and "I", but he suggested that that was the case at an earlier stage (no longer open to empirical investigations). The footnote accompanying this suggestion reads:

> It is possible that tones were used rather than free-form pronouns (...) One reader of this paper found it "inconceivable" that there would have been no first-versus-second person distinction in the language at any point in its history. In fact, however, Wari' (Everett n.d.) is a language that currently lacks any first-versus-second-person distinction.

Thus, not only is Everett suggesting that at an earlier stage in its history (presumably, before contact with Portuguese) Pirahã had no distinction between "I" and "you", but is also asserting that another Amazonian language, Wari', lacks such a distinction now.

Yet the data presented in the comprehensive grammar of Wari' of which Everett was a co-author (Everett & Kern 1997) clearly contradict this assertion, as the *Grammar* provides the words *wata'* and *wum*, glossed, respectively, as "first singular" (i.e. 'I') and "second singular" (i.e. 'you (sg)', 'thou').

As if to make it crystal-clear that Wari' in fact does have words for 'I' and 'you', the *Grammar* offers example sentences such as the following ones (ibid: 305):

(572) a. A: Ma'   wari'   ma'   quem?
             that:prox:hearer   person   that:prox:hearer   ref
             'Who is it?"

         B: Wata'.
             emph:1s
             '(It is) I.'

> B: Warut.
>   emph:1pexcl
>   '(It is) we.'

(572) b. Wum   ra?
           emph:2s   2s:rf
           '(That is) you, isn't it?'

The contradiction between Everett's assertion about Wari' and the data included in his own *Grammar* is so startling that it requires an explanation. Evidently, such an explanation can be found Everett and Kern's (ibid 1997: 305) description of *wata'* (i) and *wum* as "emphatic pronouns". Thus, the *Grammar* says that "First and second person emphatic pronouns are frequently used in answer to questions of the type 'Who is it?' or to ask 'Is that you?'" (ibid: 305). In many other contexts, one gathers, the words *wata'* and *wum* are not used, and the person is marked on the verb.

It is a truism, however, that in many languages (including, for example, Spanish and Polish), the words for 'I' and 'you' are used only in contexts where they are clearly needed (for emphasis, contrast, enumeration, and so on). This doesn't mean that such languages have no words for 'you' and 'I', and clearly, the same applies in Wari'. In fact, the *Grammar* itself provides an example where the word *wata'* 'I' is used not for emphasis but for enumeration and clarification (ibid: 304):

(571) a. Ji'am   xi'   jowin   pain   ca'   ma'
           hunt   1pincl:rf   monkey:species prep:3n   this:n   that:prox:hearer

    'urut,   Jimain Hwara' Waji,   Wem Xao,   wata'.
    1pexcl:rp/p   m:name   m:name   emph:1s
    '"We will hunt for *jowin* monkey", we (said), Jimain Hwara' Waji,
      Wem Xao and I.'

The same claim about Wari' not having words for 'I' and 'you' is repeated in Everett's other publications on Wari', clashing with the data presented in these very publications. Thus, in Everett's chapter on Wari' in *The Handbook of Morphology* (2001, online) one reads: "There are two classes of pronouns: demonstrative and emphatic (there are no personal pronouns). <...> Demonstrative pronouns occur only in the first person". Yet this claim about the absence of personal pronouns is followed by the statement that "Emphatic pronouns may occur in any person and number", and by the same examples cited in Everett (2005a) and glossed as "(It is) I" [said in answer to the question "Who is it"?] and "we, Jimain Hwara' Waji, Wam Xao and I".

Similarly, in Everett (2005: 305) one reads: "There are no first and second personal pronouns in Wari'". Yet the so-called "emphatic pronouns" *wata'* ('I') and *wum* ('you') are cited here too. The self-contradiction seems evident.

5. YOU (THOU) in reported speech and in the language of religion

Essentially, what applies to the word for 'I', applies also to the word for 'you' (thou). To begin with, 'you' too is needed for reported speech. For example, if someone wanted to report Everett and Kern's (1997: 305) sentence 'Wum ra?', glossed as "That's you, isn't it?", they would obviously need a word (or, in a sign language, a sign) meaning 'you': "She said: that's you, isn't it?" In direct discourse and at the level of "me Tarzan you Jane" a word for 'you' may seem to be dispensable (because Tarzan could simply point, first at himself and then, at Jane). But in reported speech pointing would clearly not suffice and a conventional sign meaning 'you' would be needed.

Furthermore, apart from purely factual reports on who said what to whom, a word for 'you' is essential for moral discourse and for social order. To cite a biblical example again, in the book of Genesis (4: 9–10), after Cain killed his brother Abel, the following famous conversation takes place between Cain and God:

> And the Lord said unto Cain, Where is Abel thy brother? And he said, I know
> not: Am I my brother's keeper?
> And he [God] said: What hast thou done? the voice of thy brother's blood cries
> unto me from the ground.

In all human groups, and at all times, it has been essential to be able to identify the breakers of moral, and social, law, and to require a person's personal testimony (e.g. "Did you do it?").

Speaking more generally, if 'I' is necessary for self-reflection, 'you' is necessary for the transmission of culture to children – especially moral culture. The reason for this is that moral instruction often needs to rely on scenarios involving 'you' and 'someone else'. Such instruction may be given to a group of people, but each member of this group needs to understand that he or she is being addressed as an individual (thou). For example, to understand the Gospel precept "Give to him that begs from you, and do not refuse him that would borrow from you" (Mt. 5:42), every reader needs to imagine himself or herself as being individually addressed as 'thou'.

If transmission of moral culture requires the concept of 'you' (thou), so does religion. To illustrate again from the Hebrew Bible, Genesis opens with a dialogue between God and Adam: "And the Lord God called unto Adam and said unto him, Where art thou?"', to which Adam replies: "I heard thy voice in the garden, and I was afraid, because I was naked; and I hid myself" (Genesis 3:9–10, KJV).[2] There is no way this foundational myth could be conveyed to speakers of sign languages without a word for you (thou).

In the Judeo-Christian tradition, religion –linked with the Latin root – *lig* 'to bind, to tie' and the cognate verb *religare* – is widely interpreted as being essentially

---

[2] Sentences from the Bible often provide excellent material for cross-linguistic semantic investigations because they are familiar to many readers, easy to access online, and available in hundreds of languages.

about "connection", dialogical connection, between human beings and God (cf. Nongbri 2013) While most speakers of English are not aware of these etymological links, they are aware of the connection between religion and prayer, and of the "I-thou" relation which the concept of 'prayer' presupposes. To illustrate from Psalm 23, which is perhaps more than any other prayer (except the "Our Father") part of the European 'cultural literacy': "(...) though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me; thy rod and thy staff, they comfort me." (KJV)

The great Jewish philosopher and theologian Martin Buber articulated the fundamental importance of the "I-Thou" relationship to God in Judaism in his classic book *I—Thou* (1966 [1923]) to which I will return shortly. In Christianity, too, the "I-thou" relationship between God and "man" (human person) is seen as constitutive of religion. To quote from one representative book, *Credo for Today: What Christians Believe* (Ratzinger 2006: 47):

> The first "thou' that—however stammeringly—was said by human lips to God marks the moment in which spirit arose in the world. (...) The theory of evolution does not invalidate faith, nor does it corroborate it. But it does challenge faith to understand itself more profoundly and thus to help man to understand himself and to become increasingly what he is: the being who is supposed to say "thou" to God in eternity.

The fact that in most languages the meanings 'I' and 'you' (thou) are grammaticalised and expressed in different forms of the verb (e.g. *amo* 'I love', *amas* 'you love' in Latin) underscores the supreme importance of these meanings in human thinking. As Wari' illustrates, however, the marking of these meaning in the verb is not enough: in some contexts, above all, in the context of self-identification, separate words meaning 'I' and 'you' (thou) are also needed. To adduce yet another example from the Bible, in Matthew's Gospel, John the Baptist sends two of his disciples to Jesus to ask: "Art thou he that should come, or do we look for another?" (KJV); Jesus replies quoting from the prophet Isaiah: "Go and show John again those things which you do hear and see: The blind receive their sight, and the lame walk, the lepers are cleansed and the deaf hear (...)". (Matthew 11: 3–4, RSV) To be able to convey John's question to Jesus, and to seek his own testimony on whether or not he is indeed the awaited Messiah, the word 'thou' is indispensable. A translation of Matthew's Gospel into a sign language requires a sign meaning 'you' (thou).

6. YOU in English: 'you' vs. 'thou'

At first sight, English may seem to be a glaring counterexample for the generalisation that all languages have a word meaning 'thou', that is, 'you' singular, and indeed in various typological surveys English has been presented as a language which doesn't distinguish between 'you' as a plural (addressed to two or more people) and 'you' as a singular (addressed to one person, that is, 'thou'). For

example, Ingram's (1978) survey of pronoun systems cites (building on Forchheimer 1953), inter alia, the following three:

Four-person system		Languages
I	we	KOREAN (65)
thou		KAMANUGU (66)
he		

Five-person system

A. I	we	BURMESE (43)
thou	you	
he		

B. I	we	ENGLISH
thou		
he	they	

Six-person system

A. I	we	CHINESE (42)	TURKISH (54)
thou	you	JAPANESE (43)	EAST SUKETI (55)
he	they	KOTTISH (48)	CHUKCHEE (56)
		SUMERIAN (49)	KHASI (61)
		SHILH (50)	MASAI (67)
		TSHIMSHIAN (75)	AKKADIAN (68)
		LATIN (76)	AZTEC (74)
		HAUSA (71)	TLINGIT (78)
		HOPI (73)	WIYOT (105)
		FINNISH (53)	

From the point of view of the theme of this paper, the most striking aspect of Ingram's(and Forchheimer's data is the invariable presence of words for 'I' and 'thou'. But the portrayal of English as a language which doesn't have a word for 'you plural' is misleading. In order to recognise (as Ingram does) that present-day English does have a word meaning 'thou' we need to show that the word *you* has now two meanings: 'you singular' (meaning 'thou') and 'you plural'. It is therefore a "six-person system", just like, for example, Latin, which has two phonologically distinct words *tu* (you.Sg) and *vos* (you.Pl.).

The problem of the seemingly missing 'thou' in modern English has been discussed many times in NSM literature. Rather than discuss it here *de novo*, I will simply quote the relevant passage from my 1992 book *Semantics, Culture and Cognition* (Wierzbicka 1992: 13−14), where I highlighted the problem of polysemy as an obstacle in any search for lexical universals (for further discussion, see Goddard 1995):

> I have postulated 'you' and 'I' as universal semantic primitives, but what I mean by 'you' is 'you SG' ('thou') rather than 'you PL' or 'you' SG/PL'. Yet one doesn't have to look further than modern English to find a language which doesn't seem to have a word for 'thou'. To maintain the claim that 'thou' is a lexical universal we would have to posit polysemy for the word *you*: (1) 'you SG', (2) 'you PL'. Initially, this seems an unattractive solution,

but I think there are good reasons for accepting it. Polysemy is a fact of life, and basic, everyday words are particularly likely to be polysemous (cf. **Zipf** 1949). <…>

It goes without saying that polysemy must never be postulated lightly, and that it has always to be justified on language-internal grounds, but to reject polysemy in a dogmatic and a priori fashion is just as foolish as to postulate it without justification. In the case of the English word *you*, I think its polysemy can be justified on the basis of the distinction between the forms *yourself* and *yourselves*; the choice between *yourself* and *yourselves* is determined by the choice between *you* SG and *you* PL ("you must defend yourself" vs. "you must defend yourselves") (Wierzbicka 1976).

7. Pronouns and people

In linguistic literature, pronouns they are usually discussed in terms of forms and "systems" rather than meanings and people. Unfortunately, Peter Müllhausler and Rom Harré's wonderfully named book *Pronouns and People* (1990) is no exception in this regard. Thus, the authors write:

> That third persons are either absent or merged with second persons is also characteristic of the development of Papuan languages as discussed by Laycock (1977) <…>. The systems identified by Laycock run from two-pronoun systems dividing the universe 'solipsistically, into the speaker and everybody else', this type being exemplified by the Trans-New Guinea Phylum language Morwap, to sixteen-pronoun systems with complex number and generder distinctions exemplified by Vanimo language (Sko Phylum) of Papua New Guinea.

This acknowledges that even Morwap – allegedly the language with "a minimum number of contrasts" (ibid: 80) – has a word for 'I', but at the same time it denies the presence of a word meaning 'thou' (you.sg.) in this language.

In making this claim about Morwap, the authors of *Pronouns and People* are relying on the data from Don Laycock's fieldwork, which Laycock himself described as unreliable: "I should perhaps say that my data on the language is not of high reliability, in that I was working through Malay, a language I do not control well" (ibid: 36) In addition, Laycock comments: "I am not quite sure whether (...) *ka* is first singular only, and *sa* all the others, or whether (...) *ka* is first person, all numbers, and *so* is all other persons and numbers" (Ibid: 38)

Given Laycock's uncertainty about his own data it is understandable that in his account of pronouns in Papuan languages presented in his book *The Papuan Languages of New Guinea*, William Foley (1986) decided to ignore Morwap. Nonetheless, Foley's own account of Papuan pronouns also raises serious questions. To quote:

> The simplest pronominal systems attested for any Papuan language are those of certain languages of the Chimbu family, such as Golin (Bunn 1974) and Salt-Yui <…>. To take Golin as an example, it has only two true pronouns: *na*, first person, undifferentiated for number, hence 'I/we', and *i*, second person, again indistinct for number, hence 'you'. (Foley 1986)

When we consult Bunn's *Golin Grammar*, however, we find that this statement is inaccurate. What Bunn (1974: 55) actually says is that "there are only two personal pronouns which are not compounds and they are *ná* 'I/We' and *í* 'you'". At the same time, Bunn provides a much fuller list of personal pronouns, which clearly do differentiate between 'I' and 'we' (as well as 'we two'), and also, between 'thou' and 'you-plural' (as well as 'you two'):

ná yasu	– 'we two'
í yasu	– 'you two'
yasu	– 'they two'
ná ibál kobe	– 'we'
í ibál kobe	– 'you (plural)'
ibál kobe	– 'they'

(As Bunn notes, *yasu* is formed from *yál* 'man' and *su* 'two'.) Importantly, Bunn also records that the distinction between 'I' and 'we', and 'thou' and 'you (Plural)' is also made overtly in the reflexive pronouns, for example:

ná inán 'myself'
ná ibál kobe inán 'we ourselves'

Presumably, when Foley says that Golin has only two "true pronouns" he means that it has only two forms which are morphologically simple. But it is unclear why compounds should not be regarded as true pronouns, and in any case, the conceptual distinctions between 'I' and 'we' (as well as 'we two') and 'thou' and 'you (plural)' (as well as 'you two') are evidently there. This means that Golin does have words for 'I' and 'thou'.

Accordingly, Foley's description of the Golin word *ná*, as "first person, undifferentiated for number, hence 'I/we'", and of the Golin word *i* as "second person, again indistinct for number, hence 'you'", cannot be accepted at face value. The Golin word *í* cannot be "indistinct for number" any more than the English word *you* is. The Golin word *ná* does not mean 'I/we', but simply 'I', and the Golin word *í* does not mean *you* (singular or plural) but simply 'thou' (you singular). As for the corresponding duals and plurals, their core meanings can be stated as follows (Bunn doesn't make it clear whether 'thou' is excluded from the meanings glossed as "we" or not):

ná yasu (glossed by Bunn as 'we two') –
two people, I am one of these people (thou are not one of these people)

í yasu (glossed by Bunn as "you two")
two people, thou are one of these people (I am not one of these people)
ná ibál kobe (glossed by Bunn as "we")
some people, I am one of these people (thou are not one of these people)

í ibál kobe (glossed by Bunn as "you (plural)")
some people, thou are one of these people (I am not one of these people)

The situation in Golin is in fact quite similar to that in the Melanesian creole Tok Pisin (cf. Foley 1986: 67) and in other English-based creoles, where the semantic relationship between 'I' and 'thou' and the corresponding duals and plurals is also morphologically transparent. (The suffix *–pela* derives, etymologically, from the English word *fellow,* and *mi, yu* and *tu*, from the English words *me, you,* and *two*):

> *mi* 'I', *yu* 'thou'
> *mitupela*
> 'two people, I am one of these people, thou are one of these people'
>
> *mipela*
> 'some people, I am one of these people, thou are not one of these people'
>
> *yutupela*
> 'two people, thou are one of these people, I am not one of these people'

Discussing Papuan languages in general, Foley (1986: 67) writes: "Papuan languages are especially interesting in their pronoun systems because many of them exhibit restricted, abbreviated systems not commonly found elsewhere." A close examination of the Papuan data, however, confirms, rather than undermines, the universality of 'I' and 'thou'. The pronominal systems found in these languages may be unusual from a formal (morphological) point of view, but from a semantic point of view, they conform to the generalisation that 'I' and 'thou' are essential elements of "Basic Human".

The title of the book *Pronouns and People* is inspired, because pronouns can tell us a lot about people, and especially, about how people think about themselves and others. Arguably, the most important lesson that pronouns can teach us is that in thinking about people, speakers of all languages make two fundamental distinctions: One distinction separates the "conversational pair" (I and you) from those people who (in a given situation) are thought of as "other people" (i.e., people other than) the "conversational pair". The other distinction is that between 'me' and 'you' (thou) – the two poles defining the basic human act of interpersonal communication: "I want to say something to you, you can say something to me".

8. "You and I" as a "conversational pair"

In his paper "The second person is rightly so called" Joseph Greenberg (1993: 15) coined the phrase "the conversational pair" (in my view both clearer and more elegant than the common technical phrase "speech act participants", let alone "SPA"). In drawing attention to the special status of the "conversational pair" among pronouns, Greenberg was clearly following in the footsteps of Martin Buber's classic *I and Thou*. Unfortunately, instead of drawing explicitly on Buber and building on his insights Greenberg fell under the sway of Bertrand Russell (whom he quotes with admiration), and didn't seem to fully appreciate the

uniqueness of the relationship between 'I' and 'you (sg)' as members of that primordial conversational pair.

> I shall first consider the opposition between the first person, on the one hand, and the second and third person, on the other. This might be characterized as the distinction between the ego and the non-ego. In fact the various typological properties which help to confirm this opposition fall quite easily into two types: those having to do with the uniqueness of the ego and those having to do with the common properties of the two non-first persons. (Greenberg 1993: 15)

Greenberg's emphasis on the "uniqueness of the ego" is of course welcome. The same cannot be said, however, of his distinction between "the ego and the non-ego", which subsumes 'thou' under one umbrella with 'he' and 'she' and ignores the uniqueness of 'thou' and its intimate relationship with 'I'. Thus, of 'I' ("the ego") Greenberg wrote:

> The ego has two linguistically relevant peculiarities. It is unique and unlike the second or third person it has no true plural. Even the 'chorus we' is not really a plural of the first person. Each person uttering it, whether the utterance is preconcerted or not, is referring to himself or herself plus others.

One can only agree with what Greenberg says here about 'ego', but not with what he says of the "second person": as a little focused reflection must show, 'thou' has no "true plural" any more than 'I' does. If one person says "we" when speaking on behalf of a group, then the core meaning conveyed is indeed "some people, I am one of these people" (cf. Goddard 1995). If someone addresses one person as a representative of a group, then the core meaning of *you* is, roughly, 'you plus others', and more precisely, "some people, you are one of these people". In the case of a "chorus *we*", each person uttering it is referring to himself or herself as 'I'. In the case of a group being addressed, each member of the group is being addressed as "thou".

The precise semantics of 'you plural' and of comparable words in other languages require further investigation but the main point seems clear: if 'I' is, inherently, individual ('singular'), so is "thou". This makes the primordial conversational pair fundamentally different from the so-called "third person" – a point to which Martin Buber devoted a whole book, and to which I will return below. Before doing so, however, I want to discuss briefly the second claim made by Greenberg – in my view, wrongly –about 'I' ('the ego'):

> The second relevant characteristic of the ego is that it is the primary reference point for deixis, in the broad sense in which certain terms such as the pronouns *here, there, then,* and *now* shift their denotation depending on the utterance. The existence of such terms has been recognized, probably independently, both by linguists and philosophers. They are often called shifters, following Jesperson and Jakobson. The philosopher Russell, to my mind, showed real insight in employing the term 'egocentric particulars'. He states (Russell 1948: 69) that "there are a number of words of the sort that I call 'egocentric'

which differ in meaning according to the speaker and his position in time and space. Among those the simple ones are learned ostensively, for instance 'I', 'you', 'here', 'now'."

Greenberg appears to accept Jespersen's and Jakobson's idea that the words 'I' and 'you' are "shifters", and also, Bertrand Russell's idea that they are learned by ostension, that is, by pointing. But the two ideas are profoundly at odds with one another: a concept whose reference shifts from one person to another cannot be learned by pointing with one's finger at one particular person. A child can learn by ostension who the words "Mummy" and "Daddy" stand for (because they are used in a predictable way) but not who the words 'I' and 'you' stand for (because they can be spoken by different people and to different people).

As Martin Buber puts it, "he" and "she" can be grounded in space, but 'I' and 'you' can't. They can only be grounded in eye contact (or, in exceptional cases, touch contact) between two people who are both using 'I' and 'you' in the same encounter. When the child grasps that 'you' addressed to her by her mummy can be "shifted" and used in the reverse direction, and that her mummy's "I" can also be used in the reverse direction (to refer to the child herself), then, and only then, the true meanings of 'you' and 'I' have been acquired. Before that, a child may learn the **words** 'I' and 'you', by ostension, treating them as labels for particular people (perhaps 'I' for Mummy and 'you' for the child), but this is before the **concepts** 'I' and 'you' are acquired.

When one reflects on the notion of a "conversational pair" and on the idea that the concepts 'I' and 'thou (you)' are acquired as a pair, one can appreciate the importance of eye contact between the two members of the pair for a reversible "I – you" relation to be built. An interlocking gaze is possible only between two people, not three or more, and it seems clear that this gaze is an important factor in the child's grasping the concept of 'thou'.

9. The emergence of the concept 'YOU'[THOU] in young children

How exactly can mutual gaze lead to the emergence of the concept 'you'? By itself, it could not – if this concept were not already there, in the initial stock of innate, unlearned concepts that, according to the developmental psychologist Susan Carey (2009: 12) 'any theory of conceptual development must specify'. On this point, one can only agree with Buber: that the concept of 'thou' is, in a sense, inborn, as is also the concept of 'I'. But to bring these concepts to the surface of one's consciousness one must be able to 'operationalise' them, as it were, in interpersonal encounters, in which the use of the words *you* and *I* can be coordinated with mutual eye contact.

In his path-breaking little book *Acts of Meaning* (1990: 72) psychologist Jerome Bruner wrote: "How we 'enter language' must rest upon a selective set of pre-linguistic 'readiness for meaning'. That is to say, there are certain classes of meaning to which human beings are innately tuned and for which they actively search". When one observes how avidly very young babies seek eye contact, one

can hardly doubt that this "pre-linguistic readiness of meaning" applies pre-eminently to the concepts 'thou' and 'I'.

It is worth noting, though, that babies can enter communicative interaction with other persons at a very early stage. Especially, of course, with their mothers, but also with other babies, as these two photos of my twin grandsons John and Benedict, taken when they were three months old, clearly show.

I am not suggesting that three-months old babies have a full-fledged concept of 'I' (and 'thou') at their disposal, but rather that they have what Bruner calls 'pre-linguistic readiness for meaning' in relation to 'you' and 'I'.

There is of course plenty of evidence showing that some meanings seemingly including the concept of 'I'- such as 'I want', and 'I don't' want ' – appear very

early in the child's cognitive and communicative repertoire, long before the child can say, and understand, words as tricky as 'you' and 'I'. There can be little doubt, however, that those early meanings ("I-want", "I-don't-want") seemingly including 'I' are in fact holophrastic, that is, function as unanalysable units. To be able to conceptually separate 'I' from 'want' in the pre-linguistic 'I want', the child needs to grasp that somebody else, and in particular, one's partner in the communicative encounter ('you'), may also want something, and that what 'you' want may be different from what 'I' want.

Concept like 'ball' and 'cat' can be grounded in ostention, but before a child learns to apply them to many different balls and cats, and not to apply them, for example, to oranges or rabbits, she has to master not only the ideas of "this something" and "something like this" but also that of "something of one kind". 'You' and 'I', however, do not build on ostention ('this someone', 'someone like this'), and, needless to say, they do not generalise to "kinds of people". They can only build on the experience of being in direct contact with one particular person – by touch, and by what Buber calls visual touch, that is eye contact.

Buber's thinking, often expressed in a poetic and figurative language, is profound and powerful. To quote:

> If *Thou* is said, the *I* of the combination *I-Thou* is said along with it. (Buber: 3)
> The primary word *I-Thou* establishes the world of relation. (Ibid: 6)
> I become through my relation to the *Thou*; or I become *I*, I say *Thou*. (ibid: 11)
> ... the *I-Thou* relationship requires a mutual action which in fact embraces both the *I* and the *Thou*... (ibid: 125).

As philosopher Andrea Lailach-Hennrich (2011) puts it, following in Buber's footsteps, "The communicative relationship with another person, made possible by the capacity to take over this other person's perspective, can be seen as the ground for an interpersonal relation, without which there can be no awareness of oneself". (ibid: 234). Accordingly, a full understanding of the concept of 'I' must be anchored in earlier exchanges with someone else, a 'you' (thou). And a few more highly pertinent quotes from Lailach-Hennrich:

> ...to have a concept of one's own self (...) one must be able to take the perspective of other persons and to enter into communicative interaction with them […] a person can only fully acquire a concept of oneself ('Selbstbegriff') in communication. (ibid: 235–237)

> …psychological predicates would have no meaning if they could not be attributed to different persons ("Subjekten") regardless of the perspective". (Ibid: 233)

> A ... an individual can only then apply psychological predicates to him or herself, when he or she can competently apply them to other persons (Ibid: 235). The ability to take ["hinnehmen"] the perspective of other persons is an essential prerequisite for the acquisition of the concept of oneself ["von sich"]. (Ibid: 235)

> Creatures which neither have the ability to take on the perspective of others
> nor communicate with one another, cannot be creatures aware of themselves.
> (Ibid: 235)

This means that some acts of communication with someone else must occur before the concept of 'I' can be fully acquired, and also, that 'I' and 'THOU', must be, in some sense, acquired together. Echoes of Buber's thinking can be clearly heard here. They can also be heard in the passage from Benveniste adduced at the outset of this paper, but not in the writings of Bertrand Russell or in the "Myth of language universals".

One cannot help being struck by the absence of any dialogue with the humanistic thought of the past which at times could bring together scholars as different as Humboldt, Boas, Benveniste, Buber, and even Darwin, in those strands of modern linguistics which want to align themselves, above all, with the science of the brain, and seem prepared to sacrifice 'you' and 'me' at the altar of "Cognitive Science".

10. Concluding Remarks

As I discussed in my 2021 paper 'Semantic Primitives Fifty Years Later' (Wierzbicka 2021), according to the encyclical of Pope Francis *Fratelli Tutti*, "In today's world the sense of belonging to a single family is fading" (section 30). From this point of view, it seems particularly important to recognise that the principle of the psychological unity of all people on earth is not just a "pious" slogan (cf. Shweder & Sullivan 1990: 400) or a well-meaning declaration not based on evidence, but a truth supported by empirical findings; and that these findings can enhance our sense of belonging to a single family and a universal community of communication.

The current one-sided emphasis on the diversity of languages without acknowledgment of their fundamental unity undermines this truth about the unity of the human mind and of the "human race". The emphasis that many influential linguists place today on linguistic diversity is such that the underlying conceptual unity of all languages tends not to be mentioned at all. When it is mentioned (which is very rare) it is done only in general terms, without any concrete examples. Typically, both in scholarly linguistic works and in publications for the general reader, numerous examples of astounding diversity are offered, without a single example of something that all languages share.

The message implicitly (if not explicitly) conveyed is that "the unity of the human mind" is only a pious slogan. There are no shared human concepts, there can be no "universal human community of communication". The thing to do is to celebrate the diversity of languages, and not to seek what we humans share.

By contrast, the NSM approach, which was initiated by the publication of *Semantic Primitives* fifty years ago, has always seen the diversity of human languages as combined with, and undergirded by, a shared conceptual core, and has

sought to determine what that shared core was, regarding this search as a task of utmost importance (see Wierzbicka 2021).

As cross-linguistic investigations of the last fifty years show, despite the phenomenal diversity of human languages and cultures, a shared "alphabet of human thoughts" was not just a figment of Leibniz' imagination. This is what makes as human: shared simple concepts, like PEOPLE, KNOW and THINK, SAY and WORDS, GOOD and BAD, TWO and MANY, and just as importantly, the complex but indispensable concept 'we'. But there could be no 'we' without the prior concepts YOU and I.

CODA

As I was finalizing this paper in the third week of March 2022 I received two emails which went to the heart of what this paper and my fifty-year pursuit of universal semantic primes are all about. One of these emails, which arrived on March 18, brought with it an article from the Russian newspaper *Novaya Gazeta*, the first sentence of which reads: "The current catastrophe of humanity (*chelovechnost'*) has become a challenge for religious consciousness". For my purposes, the key word here is "humanity", *chelovechnost'*.

The other email, which came on March 16 (from a British pastor, Andrew Rowell), I will quote at some length. It starts with a question about the Polish linguist Andrzej Bogusławski who in the early 1960s posited the existence of what he called "indefinibilia", that is, universal, non-arbitrary semantic primes. Referring to my 1980 book *Lingua Mentalis*, Rowell wrote:

> I'm reading the Introduction in your *Lingua Mentalis* and I feel like I have found a buried treasure. Do you know how Professor Bogusławski became interested in the semantic primes? Did he think of the idea himself and then rediscover the tradition or did he discover the tradition first and take it up again? I feel as though this is an important story that should be told.

An hour later, responding to my questions, and referring to the whole NSM program and research community (in particular, the work of David Bullock), Rowell wrote:

> My academic background was originally molecular biology... I know this does not seem very related but it was actually the information in DNA and how it works that fascinated me. I taught biology for some years then became interested in how information can be stored in relational databases.
> Soon after this I felt called into the Christian pastoral ministry and during my training became fascinated by how language can be stored in databases. This triggered an interest in what is at the roots of language...what are the starting points....I became very frustrated with the circles in all the dictionaries...it seemed like they were hiding something very important!
> I thought that if I set up a database where I disambiguated the important verb definitions in Wordnet I could use the high frequency of recurring small

circles of definitions to identify the primes (I did not call them that...I just called them the root verbs).
I discovered your work and NSM fairly recently when I was trying to find out if anyone had already done what I was trying to do.
I am interested in Christian apologetics and I am fascinated by the Apostle Paul's arguments in Romans 2 about how God has "written" his law on our hearts. I am interested in the connection between this and the foundations for language and logic.

The sentence from Apostle Paul's *Letter to the Romans* to which Rowell refers reads: "When the Gentiles who have not the law [Moses' law] do by nature what the law requires <...> they show that what the law requires is written on their hearts"[3] (Romans, 2: 14–15, RSV). The question to which Andrew Rowell draws our attention is this: in what language is the law that is "written on human hearts" formulated? My own answer to this question would be: it is written in Basic Human, humanity's shared conceptual language, the core of which is constituted by universal semantic primes.

But this is only the first question. The second, which we must also ask and which is closely related to the first, is this: *what* exactly is "written on human hearts"? The Apostle Paul might have given the answer: "the Ten Commandments". But, the Ten Commandments were formulated in Hebrew, not in Basic Human. So *what* "law", written on human hearts and shared by all humanity, are we able to formulate in Basic Human? Here is my tentative and incomplete proposal, with an emphasis on *chelovechnost'*, that is, "humanity":

---

[3] I am not suggesting that "law written on human hearts" as interpreted here will be generally regarded as "natural" or will be universally accepted. I think that St Paul appeals here to our common humanity and to our best moral intuitions, available to individuals in all cultures and societies. These intuitions may go against the grain of the dominant culture, and yet be available to every person living within that culture if they deeply and sincerely search their heart.

In an expanded version of the "law written on human hearts" (as I interpret it) I have included two extra lines:

1. It is bad if people want to kill some other people.
2. It is good if people want to do good things for their mother, for their father.

It is of course impossible to know for sure whether St Paul saw the content of these two extra lines as moral intuitions available to all people (**even** without Revelation). I can only say that they are both supported by the Ten Commandments. My friend James Franklin, a philosopher and a mathematician, has commented on the expanded version as follows: "I guess St Paul doesn't say 'Thou shalt not kill' is part of the content of law written in the heart, but I suppose it is reasonable to say that would be central to it." (personal email, March 31, 2022)

I would add that a Christian commentator might want to go further than what the expanded version says, and to say, like my friend Mark Durie, a linguist and a theologian (personal email, March 23, 2022):

• Is not 'basic human' a gift from the creator to human beings?
• Is there not a universal ethical sensibility, expressible in all languages, and known by all people, which can be expressed simply in 'basic human'?
• Do not these two truths mean that the calling to fully embrace our humanity leads us to acknowledge our creator? (…)

But one doesn't need to be a Christian to accept the middle one of the three points above.

> We can know that in his Letter to the Romans (2:14–15, RSV) the Apostle Paul says this:
> "When Gentiles who have not the law, do by nature what the law requires, they are a law to themselves, even though they do not have the law.
> They show that what the law requires is written on their hearts, while their conscience also bears the witness, and their conflicting thoughts accuse or perhaps excuse them."
>
> We can think about it like this:
> when Paul said these words, he wanted to say something like this:
>
> "something in me says: don't do bad things, do good things
> at the same time, something in me says:
> 'it is bad if people want to do bad things to some other people
> it is good if people want to do good things for some other people'"

In a way, this is a distillation of what I proposed in my 2018 paper "Charter of global ethics" (in Goddard ed. *Minimal English for a Global World*). Assuming that this is on the right track, the key semantic primes on which the "law written on people's hearts" (as interpreted here) relies include GOOD and BAD, DO and NOT, PEOPLE and OTHER; they also include ME and YOU – indispensable building blocks of human ethics, understanding and communication.[4]

REFERENCES

Apresjan, Jurij. 2000. On the Language of Explications and Semantic Primitives. In his *Systematic Lexicography* (Trans. K. Windle). New York: Oxford University Press: 215–230.

Arnauld, Antoine & Pierre Nicole. 2011 [1662]. La Logique ou l'art de penser (édition critique par Dominique Descartes). Paris: Champion.

Benveniste, Emile. 1971. *Problems in General Linguistics*. Translated by Mary Elizabeth Meek. Coral Gables, Florida: University of Miami Press.

Bogusławski, Andrzej. 2003 [1963]. A note on Apresjan's concept of "Polish school of semantics". With an Appendix: K voprosu o semanticheskom analize (On the issue of semantic analysis) [1963]. *Lingua Posnaniensis* XLV. 7–18.

Buber, Martin. 1966 [1923]. *I and Thou.* 2nd ed. translated by R. G. Smith. Edinburgh: T. & T. Clark.

Bunn, Gordon. 1974. *Golin Grammar.* Workpapers in Papua New Guinea Languages 5.

Carey, Susan. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.

Cormier, Kearsy, Adam Schembri & Bencie Woll. 2010. Diversity across sign languages and spoken languages: Implications for language universals. *Lingua* 120 (12). 2664–2667.

Cormier, Kearsy, Adam Schembri & Bencie Woll. 2013. Pronouns and pointing in sign languages. *Lingua* 137. 230–247.

Darwin, Charles. 2003 [1871]. *The Descent of Man.* London: Gibson Square Books.

---

[4] It is worth noting in this connection that autistic children are known to tend to start using 'you' late, and also to have persistent difficulties with sustained eye contact (Tanaka & Sung 2016, Hadjikhani et al. 2017, Moriuchi & Jones 2017, Shield & Tager-Flusberg 2015, Lee & Chiat 1994, Naigles 2016). Blind people can no doubt develop the concept of 'you' through touch alone. But observations of young autistic children point to both difficulties with eye contact and with the use of the pronouns "I" and "you" and suggest an interconnection between the two.

de Waal, Frans. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton, N.J.: Princeton University Press.

Descartes, Rene. 2009 [1642]. *La recherche de la vérité par la lumière naturelle*. Paris: Presses universitaires de France.

Diller, Anthony. 1994. *Thai*. In Cliff Goddard & Anna Wierzbicka (eds.). 149–170.

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32 (5). 429–448.

Evans, Nicholas. 2014. The wellsprings of linguistic diversity: Project presentation. *Seminar presented at Australian National University*, 23/07/14.

Everett, Daniel L. & Barbara Kern. 1997. *Wari': The Pacaas Novos Language of Western Brazil.* London: Routledge Descriptive Series.

Everett, Daniel L. 2001. Wari' (Amazonia). In Andrew Spencer & Arnold M. Zwicky, *The Handbook of Morphology.* Blackwell Publishing. Blackwell Reference Online. 20 August 2014. http://www.blackwellreference.com/subscriber/tocnode.html?id=g9780631226949_chunk_g978063122694934 (accessed 26 September 2022)

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Piraha: Another look at the design features of human language. *Current Anthropology* 46 (4). 621–634.

Foley, William A. 1986. *The Papuan Languages of New Guinea.* Cambridge: Cambridge University Press.

Forchheimer, Paul. 1953. *The Category of Person in Language*. Berlin: Walter de Gruyter & Co.

Gladkova, Anna & Tatiana Larina. 2018. Anna Wierzbicka, language, culture and communication. *Russian Journal of Linguistics* 22 (4). 717–748.

Goddard, Cliff & Anna Wierzbicka (eds.). 1994. *Semantic and Lexical Universals: Theory and Empirical Findings*. Amsterdam: John Benjamins.

Goddard, Cliff & Anna Wierzbicka (eds.). 2002. *Meaning and Universal Grammar: Theory and Empirical Findings*, vols 1 and 2. Amsterdam: John Benjamins.

Goddard, Cliff & Anna Wierzbicka. 2014. *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. Oxford: Oxford University Press.

Goddard, Cliff & Anna Wierzbicka. 2021. "We": Conceptual semantics, linguistic typology and social cognition. *Language Sciences* 83.

Goddard, Cliff (ed.). 2018. *Minimal English for a Global World*. Cham: Palgrave Macmillan.

Goddard, Cliff. 1995. Who are "we"? The natural semantics of pronouns. *Language Sciences* 17 (1). 99–121.

Goddard, Cliff. 2011. *Semantic Analysis: A Practical Introduction*. Oxford/New York: Oxford University Press.

Goswell, Della. 2014. Being there: Role shift in English to Auslan interpreting. *Signed Language Interpreting*. 67–92. Routledge.

Greenberg, Joseph H. 1993. The second person is rightly so called. In Mushira Eid & Gregory Iverson (eds.), *Principles and Prediction: The analysis of natural language*, 9–23. Amsterdam: John Benjamins.

Hadjikhani, Nouchine, Jakob Åsberg Johnels, Nicole R. Zürcher, Amandine Lassalle, Quentin Guillon, Loyse Hippolyte, Eva Billstedt, Noreen Ward, Eric Lemonnier & Christopher Gillberg. 2017. Look me in the eyes: Constraining gaze in the eye-region provokes abnormally high subcortical activation in autism. *Scientific Reports* 7 (1).

Ingram, David. 1978. Typology and universals of personal pronouns. In Joseph Harold Greenberg, Charles Albert Ferguson & Edith A. Moravcsik (eds.), *Universals of Human Language: Word structure*, 213–247, vol. 3. Stanford, California: University of Stanford Press.

Johnston, Trevor. 1998. *Signs of Australia: A New Dictionary of Auslan (The Sign Language of the Australian Deaf Community)*. North Rocks, N.S.W.: North Rocks Press.

KJV (King James Version). 1998. The Bible: Authorized King James Version. Oxford: Oxford University Press.

Lailach-Hennrich, Andrea. 2011. *Ich und die anderen: Zu den intersubjektiven Bedingungen von Selbstbewusstsein.* Berlin: de Gruyter.

Laycock, Don. 1977. Me and you versus the rest: Abbreviated pronoun systems in Irianese/Papuan languages. *Irian* 7. 33–41.

Lee, Anthony, Peter R Hobson & Shulamuth Chiat. 1994. I, you, me, and autism: An experimental study. *Journal of Autism and Developmental Disorders* 24 (2). 155–176.

Lyons, John. 1977. *Semantics.* Cambridge: Cambridge University Press.

Moriuchi, Jennifer, Ami Klin & Warren Jones. 2017. Mechanisms of Diminished Attention to Eyes in Autism. *The American Journal of Psychiatry* 174 (1). 26–35.

Muhlhausler, Peter & Rom Harre. 1990. *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: B. Blackwell.

Naigles, Letitia R, Michelle Cheng, Nan Xu Rattansone, Saime Tek, Neha Khetrapal, Deborah Fein & Katherine Demuth. 2016. "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in Autism Spectrum Disorders* 27. 11–20.

Nongbri, Brent. 2013. *Before Religion: A History of a Modern Concept.* New Haven and London: Yale University Press.

Onishi, Masayuki. 1994. Semantic Primitives in Japanese. In Cliff Goddard & Anna Wierzbicka (eds.). 361–385.

Pascal, Blaise. 1958. *Pascal's Pensées*. New York: E.P. Dutton.

Pinker, Steven & Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13 (4). 707–727.

Prono, Luca. 2007. Psychic unity of humankind. In H. James Birx (ed.), *Encyclopedia of Anthropology*, (Sage Publications, online): https://sk.sagepub.com/reference/anthropology (accessed 26 September 2022)

Ratzinger, Joseph. 2006. *Credo for Today: What Christians Believe.* Vatican City: Libreria Editrice Vaticana.

RSV (Revised Standard Version). 1952. The Bible. Washington, DC: National Council of the Churches of Christ.

Shield, Aaron Shield, Richard P Meier & Helen Tager-Flusberg. 2015. The Use of Sign Language Pronouns by Native-Signing Children with Autism. *Journal of Autism and Developmental Disorders* 45 (7). 2128–2145.

Shweder, Richard A. & Maria A Sullivan. (1990). The semiotic subject of cultural psychology. In Lawrence A. Pervin (ed.), *Handbook of personality: Theory and research*, 399–416. New York: The Guilford Press.

Sorensen, Holger Steen. 1958. Word-classes in modern English. Copenhagen: G.E.C. Gad.

Suddendorf, Thomas. 2013. *The Gap: The Science of what Separates us from Other Animals.* New York: Basic Books.

Tanaka, James W. & Andrew Sung. 2016. The "Eye Avoidance" Hypothesis of Autism Face Processing. *Journal of Autism and Developmental Disorders* 46 (5). 1538–1552.

Wierzbicka, Anna. 1972. *Semantic Primitives*. Frankfurt: Athenäum.

Wierzbicka, Anna. 1976. In defense of YOU and ME. In W. Girke & H. Jachnow (ed.), *Theoretische Linguistik in Osteuropa: Originalbeiträge und Erstübersetzungen*, 1–21. Berlin, Boston: De Gruyter.

Wierzbicka, Anna. 1980. *Lingua Mentalis: The Semantics of Natural Language*. Sydney: Academic Press.

Wierzbicka, Anna. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York: Oxford University Press.

Wierzbicka, Anna. 1996. *Semantics, Primes and Universals*. Oxford: Oxford University Press.

Wierzbicka, Anna. 2021. Semantic Primitives, fifty years later. *Russian Journal of Linguistics* 25 (2). 317–342. https://doi.org/10.22363/2687-0088-2021-25-2-317-342

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

Bionote:
Anna WIERZBICKA is Professor of Linguistics (Emerita) in the School of Literature, Languages, and Linguistics at Australian National University. Her work spans a number of disciplines, including cognitive science, anthropology, psychology, philosophy and religious studies as well as linguistics, and has been published in many journals across all these disciplines. She is an author of over twenty books. Her latest books are *What Christians Believe: The Story of God and People in Minimal English* (Oxford University Press, New York, 2019), and *Vo Chto Veryat Khristiane: Istoriya Boga i Lyudei* [What Christians Believe: The Story of God and People]. Moscow, YaSK Publishers, 2021). Professor Wierzbicka is a Fellow of the Australian Academy of the Humanities, the Australian Academy of Social Sciences, the Russian Academy of Sciences and the Polish Academy of Arts and Sciences. She is the winner of the International Dobrushin Prize for 2010 and of the Polish Science Foundation's 2010 prize for the humanities and social sciences.
e-mail: anna.wierzbicka@anu.edu.au
https://orcid.org/0000-0002-6074-7865

Сведения об авторе:
Анна ВЕЖБИЦКАЯ – известный лингвист, профессор (эмерит) Института литературы, языков и лингвистики Австралийского национального университета. Помимо лингвистики ее научные интересы охватывают целый ряд дисциплин, включая когнитологию, антропологию, психологию, философию и религиоведение. Ею опубликовано большое количество статей в журналах по данным направлениям, а также более двадцати книг. Среди последних – "What Christians Believe: The Story of God and People in Minimal English" (New York: Oxford University Press, 2019) и «Во что верят христиане: История бога и людей» (Москва: ЯСК. 2021). Профессор Вежбицкая является членом Австралийской академии гуманитарных наук, Австралийской академии общественных наук, Российской академии наук, Польской академии знаний, лауреатом Международной премии им. Добрушина (2010) и премии Польского научного фонда (2010).
e-mail: anna.wierzbicka@anu.edu.au
https://orcid.org/0000-0002-6074-7865

Research article / Научная статья

Semantic decomposition of four Quranic words

Mark DURIE ⬤ ✉

Australian National University, Melbourne School of Theology,
Canberra / Melbourne, Australia
✉mdurie@jefferycentre.mst.edu.au

Abstract
In this paper, the author proposes Natural Semantic Metalanguage (NSM) semantic decompositions of four difficult-to-translate quranic Arabic words using Natural Semantic Metalanguage (Goddard & Wierzbicka 2014, Wierzbicka 2021). This is the first study to propose an explicit semantic explication of these core Islamic lexical items, which are foundational for the spiritual worldview of the almost two billion followers of Islam in the world today. The first word considered is *rasūl*, which refers to intermediaries sent by Allah to humans and is used in the Quran alongside *nabī*, which has almost the same meaning. An NSM semantic explication of *rasūl* is contrasted with explications of biblical Hebrew *nābā'* 'prophesy' and *nabī'* 'prophet'. In English translations of the Quran, *rasūl* is usually rendered as 'messenger' and *nabī* as 'prophet', yet these translations are misleadingly inadequate.
Three further quranic concepts are examined, which have received the most diverse and unsatisfactory renderings in English translations of the Quran: *shirk* 'association' and *kāfir* 'disbeliever' refer to two dimensions of disbelief, and *ittaqā*, a difficult-to-translate verb, refers to cautious piety. The use of Natural Semantic Metalanguage overcomes the resistance of these terms to translation into English, by means of fine-grained semantic explications using semantic primes. These explications are designed to be readily accessible to speakers of languages other than English.
Keywords: *Quran, prophet, Natural Semantic Metalanguage, Islam, semantic prime*

Семантическая декомпозиция четырех слов из Корана

Марк ДЬЮРИ ⬤ ✉

Австралийский национальный университет, Мельбурнская школа теологии,
Канберра / Мельбурн, Австралия
✉mdurie@jefferycentre.mst.edu.au

Аннотация
В статье предлагается использовать Естественный Семантический Метаязык (ЕСМ) (Goddard & Wierzbicka 2014, Wierzbicka 2021) для семантической декомпозиции четырех трудных для перевода арабских слов из Корана. Это первое исследование, в котором предлагается

эксплицитное семантическое толкование данных ключевых лексем, основополагающих для духовного мировосприятия почти двух миллиардов последователей ислама в современном мире. Первое рассматриваемое слово – *rasūl,* называющее посредников, посланных к людям Аллахом, используется в Коране наряду с *nabī*, имеющим почти то же самое значение. Семантическое толкование лексемы *rasūl* противопоставляется толкованию библейского *nābā'* 'пророчество' и *nabī'* 'пророк'. В английском переводе Корана *rasūl* обычно передается словом 'messenger', а *nabī* – 'prophet', однако оба этих перевода не являются адекватными и вводят в заблуждение. Далее анализируются три концепта из Корана, которые передаются различными способами, при этом ни один из них не может считаться удовлетворительным: *shirk* 'association' и *kāfir* 'disbeliever' относятся к двум аспектам неверия, а трудный для перевода глагол *ittaqā* обозначает 'осторожное благочестие'.

Использование Естественного Семантического Метаязыка позволяет преодолеть сложности перевода данных лексем на английский язык с помощью их детального толкования с использованием семантических примитивов. Предназначение этих толкований – быть легкодоступными для носителей иных языков, кроме английского.

Ключевые слова: *Коран, пророк, Естественный Семантический Метаязык, ислам, семантический примитив*

1. Introduction

1.1. In Honour of Igor Mel'čuk

When I was invited to contribute to this special edition of the *Russian Journal of Linguistics* to honour Igor Mel'čuk's 90th birthday, my affection for Igor made me very keen to share in the privilege of honouring him. Although I had been out of professional academic linguistics for more than two decades, heartfelt affection prevailed, so here is my offering.

I first had the pleasure of meeting Igor in the early 1980s, when he was visiting the Australian National University to give a workshop on Meaning–Text Theory (MTT). To this day I have a vivid memory of Igor's impassioned appeal at the end of his presentation, calling for co-workers to join him in his linguistic mission. I was a graduate student at the time, and although I did not feel ready to sign up on the spot to devote my life to MTT, I was deeply impressed by Igor's humanity, creativity and joy, and inspired by his love for words. He won my heart as well as my head, and a friendship formed which has endured beyond many others.

Twenty years after meeting Igor, I retired from academia to serve as an Anglican priest. If anything, our friendship grew stronger after this. He would address me as *Saint Mark*, and I would address him as *Jesus*, in honour of his father's thwarted desire to name him *Yehoshua*.[1]

The last time I checked, Igor Mel'čuk was an atheist. Yet when he quoted Anna Wierzbicka's (2001: 21) semantic explication of God, he reminded us that God is good (Mel'čuk 2018: 536). Given Igor's admiration for Wierzbicka, and his

---

[1] English *Jesus* is from Latin *Iesus*, from Greek *Yesous*, from Aramaic *Yeshua*, from Hebrew *Yehoshua*.

celebration of her "gift to the world" of semantic decomposition (Mel'čuk 2018: 522), he will appreciate why, although my height may be enough for most purposes, I have chosen for the best of reasons to stand on Anna Wierzbicka's excellent shoulders to deliver this homage to 'Jesus of Montréal'.

1.2. The Challenge of Translating Distant Texts

In my work as an Anglican priest I developed an interest in comparative theology and especially in the relationship between Islam and Christianity. I have long been fascinated by the ways in which differences in the meanings of key words frame how people of different faith traditions understand – and misunderstand – each other. Ludwig Wittgenstein (Wittgenstein 1958: §114) once wrote: "One thinks that one is tracing the outline of the thing's nature over and over again, and one is merely tracing round the frame through which we look at it." His point was that, far from merely describing the nature of reality, logical propositions tell us as much about the language they are couched in as about the thing they purport to describe. So, too, do the words we use: they tell us as much about our thought-culture as they do about the world they point us to.

We are imprisoned by the frames we peer through, above all by those familiar friends, the words we use. This has certainly been evident in the long history of European engagement with Islam, in which Christian scholars have viewed Islam and its scripture, the Quran, through the lens of biblical concepts. Over the centuries the biblical framing of Islam has embedded itself deeply into Western scholarship on Islam, including translation practices. This paper seeks to shine light on this framing, and challenge it by exploring the meanings of a handful of key Arabic terms using Natural Semantic Metalanguage (Goddard & Wierzbicka 2014). These terms have been chosen because they all present significant obstacles to translation and they are theologically central concepts for the belief and practice of Islam.

The challenges that these words present are not all the same. In some cases, translators have no alternative but to shatter a word into pieces, deploying seemingly disconnected terms in English to render a single Arabic word. In other cases, there is an obvious choice to translate a term, but that choice misrepresents the original meaning to a considerable extent.

In the first case, an English-speaking student of the Quran might struggle to grasp a concept because of the seeming lack of coherence of its many translational equivalents. In the second case, the same student might innocently yet falsely assume they have understood a word because one consistent translational equivalent has been deployed by the translators.

The task of translation is contradictory. One's initial intention, elucidating the text in another language, is limited by unsatisfactory approximations which say both more and less than the original text, the translation both subtracting and adding meaning. This difficulty is amplified when dealing with more distant texts. To translate a piece of French journalism into English is to plunge into a veritable sea of exuberance and deficiency (Ortega y Gasset 1959), but this pales into

insignificance compared with the difficulties that arise when the Arabic Quran is translated into languages that have been formed by a biblical tradition. Despite much talk over the past 60 years of 'Abrahamic religions' (Hughes 2012), the biblical and quranic traditions are more different than they might at first appear, and key concepts often do not translate readily from one faith language into another.

One of the great conceptual divides in the world today is the gulf that exists between the cultures and languages that have been shaped by the Quran, and those that have been shaped by the Bible. Viewed through the eyes of a modern Western reader, the faith of Islam is replete with foundational texts that are conceptually distant, being set far apart in time, place and conceptual worldview from Europeans, whose native languages have been shaped by a biblical tradition. For someone who comes to the study of Islam from the vantage point of a European cultural background, to understand the core concepts of Islamic texts requires patience and skilful vigilance.

1.3. Islam, Europe and Words

In 2001, when the September 11 atrocity took place, the eminent scholar of Islam Bernard Lewis was working on the page proofs of his latest book. In *What Went Wrong: Western Impact and Middle Eastern Response*, Lewis was tracing the crisis of the Muslim world's engagement with the West and with modernity. This engagement had made some Muslims aware that all was not well in the House of Islam. In *What Went Wrong*, Lewis makes many wise observations about differences in the conceptual worlds of Islam and of Europe.

The differences Lewis points out in this handy book are embodied in the meanings of words. For example, he observed that "in Middle-Eastern usages, liberty or freedom was a legal not a political term. It means one who was not a slave, and unlike the West, Muslims did not use slavery and freedom as political metaphors" (2002: 54). In another example, Lewis commented that "secularism in the modern political meaning … is, in a profound sense, Christian" (2002: 96). In support of this statement he observed that "in the course of the centuries, Christian jurists and theologians devised or adapted pairs of terms to denote this dichotomy of jurisdiction: sacred and profane, spiritual and temporal, religious and secular, ecclesiastical and lay" (2002: 98). (One could add church and state to this list). In contrast, as Lewis reports, the Arabic of medieval Islam had no comparable terminologies: it simply lacked the lexicon to make any clear separation between the religious and the secular.

Lewis can be faulted over the way he makes his points. He says that "in Middle-Eastern usages, liberty or freedom was a legal not a political term", but it will not do to treat *freedom* and *liberty* as meaning the same thing (Wierzbicka 1997: 25ff). It is an even worse error to describe *freedom*-or-*liberty* as a "legal term" in "Middle-Eastern usages", for this affords a universal status to English which it does not merit. The English words *freedom* and *liberty* are not universal concepts to be realised in different languages. As Anna Wierzbicka (1997: 138ff) has shown,

even the closest translational equivalents in other European languages of the English word *freedom* have distinct meanings, which reflect the differing cultural histories of the speakers of those languages. It would have been more precise if Lewis had said that the usual suspects for translating *freedom* into the major Middle Eastern languages of Islam – Arabic *ḥuriyah*, Persian *āzādi* and Turkish *özgürlük* – are legal and not political terms.[2]

Setting such niceties aside, Lewis's point is nevertheless both insightful and sound, that the lexicalised concepts of Europe and of the Islamic world have been deeply shaped by the respective religious traditions of Christianity and Islam, and differ considerably as a result. Wierzbicka (2019: 295–299) has argued that English *love*, German *lieben*, French *aimer* and Russian *ljubit'* have all derived their meanings from Greek *agapao*, as used in the New Testament, and ultimately from the Hebrew verb *'āhēb*. This concept of love, she observes, is peculiar to cultures influenced by Christianity. In contrast, in recent years I have been serving as a pastor to Iranian converts from Islam to Christianity, and have indeed found that there is no single Farsi word that can be used to translate *agapao* of the Bible. This means that to preach to Iranians about the biblical concept of the love of God requires careful cross-cultural semantic reflection.

Before we explore the meaning of our handful of quranic Arabic terms, we first need to lay a foundation of some fundamental Islamic theological concepts.

2. Foundations of Islamic Theology

The message of the Quran is grounded in certain beliefs about humanity, Allah, and the relationship between them. According to the Quran, Allah created the world and as creator he stands in relation to human beings as a master to slaves. This type of relationship was familiar to the community within which the Quran was first recited.

In exercising his mastery, Allah commands people to perform certain actions and to refrain from others. However, the human capacity to follow Allah's directives is imperfect, for, as the Quran explains, "the human was created weak" (*wa-khuliqa l-insānu ḍa'īfan*, Q4:28;[3] Q30:54) and can easily "go astray" (*ḍalla* and *aḍalla*).

To address this fundamental human deficit, Allah provides means of correction. These means are referred to as *hudā*, a nominal from the verb *hadā* «guide» (root *h-d-y*).[4] *Hudā* is normally translated in English as 'guidance'. Because of the centrality of the concept of *hudā* in Islam and in the Quran, we will briefly clarify its meaning before moving on to discuss our key terms.

---

[2] Today the meanings of these terms are expanding under the influence of global culture, so I am not sure that Lewis's observation still holds true of these terms as they are currently used.

[3] Q4:28 refers to Sura 4, verse 28 of the Quran.

[4] Arabic verbs are cited in the third person singular perfect, which is the standard citation form in Arabic linguistics. Most Arabic words are formed around roots of three consonants.

The English verb *guide* is polysemous. It has certain distinct but related meanings in reference to physical movement. One is that someone accompanies another on a journey. As the OED puts it, to guide is "to go with or before for the purpose of leading the way."[5] The one who does not know the way can be assisted by a *guide*, who accompanies and shows the way. A guide is not in a position of superiority or command over the person guided, and neither do they take responsibility for transporting the person: the person being guided moves freely of their own volition, assisted by the presence of the guide.

A second, closely related meaning of English *guide* is when God, providence or some other higher power or point of reference is the agent. In this case, the higher power is not said to be moving with the person, but it is as if this higher power were present with the person. Thus someone can speak about being "guided by God" or "guided by the stars".

The Quran's concept of *hudā* is different from English *guidance* because *hudā* does not imply the presence of an accompanying guide.[6] Rather, directions are given by someone who tells another which way to go, but without accompanying them. Thus the concept of *hudā* is about giving directions, not accompanying guidance. In a rock inscription written in Safaitic, an Arabic dialect which predated the Quran, the word *hdy* appears in reference to a military commander, who is someone who gives directions to others (Al-Jallad 2015: 317).

In the worldview of the Quran, the alternative to being rightly directed is the ignorance of not knowing the way. According to the Quran, someone in this situation will "go astray". It is bad to go astray: someone can perish in the wilderness when they wander off the track.

Consistent with the concept of the walk of faith as a journey, the Quran includes repeated references to *al-ṣirāṭ al-mustaqīm* 'the straight road'. Although *ṣirāt* was borrowed from Roman *strata* 'paved road' via Greek and Aramaic (Jeffery 1938: 195–196), the metaphor of finding and staying on the *ṣirāṭ al-mustaqīm* is not about following a highway, a made road, or even a well-beaten track. The way of Islam is not so obvious that people cannot easily stray off it. Indeed, a great deal of effort is devoted in the Quran to preventing people from straying.

According to the Quran, in order to help people to avoid straying, the right way is pointed out to human beings by *ʾāyāt bayyināt* 'clear signs' (Q2:99). These are provided by Allah, but they need to be recognised for what they are and acted upon. These signs include natural features such as the sun and the rain (Q10:5, 24), as well as stories and lessons learned from people who have lived in the past, general observations about life, and even verses of the Quran itself (Q2:185; Q10:15).

---

[5] "guide, v." *OED Online*. Oxford University Press, December 2021 (accessed 15 February 2022).

[6] From a different Arabic root, *r-sh-d*, the word *murshid* is formed, which can be an accompanying guide.

3. Data and Methodology

Here we explore the meanings of four quranic Arabic words: *rasūl*, *shirk*, *kāfir* and *ittaqā*.

These terms play a central role both in the Quran and in Islamic thought, but none of them has a straightforward English translation.

For each Arabic word considered here there is an established tradition of English translation. I will call words or phrases conventionally, but not necessarily accurately, used in such translations 'Conventional Translational Equivalents' or CTEs. For example, the Arabic word *mushrik* has an actual meaning (roughly speaking) of 'someone who wrongly attributes shared power over something to Allah[7] and another'. 'Polytheist' is often used to translate *mushrik*, as well as 'idolater', 'pagan', 'associator' and 'unbeliever', but in reality, there is no English word that even approximates this meaning.

Here I use double angled brackets to signal a CTE; for example, *mushrik* «polytheist».

The Arabic words we will consider here, together with the number of occurrences in the Quran, are:

1. *rasūl* «messenger» (322 occurrences) and *nabī* «prophet» (75 occurrences)
2. *shirk* «polytheism» (5 occurrences) and related terms based on the root *sh-r-k*: *ashraka* «to associate» (71 occurrences) and *mushrik* «polytheist» (33 occurrences)
3. *kāfir* «disbeliever» (156 occurrences) and related terms based on the root *k-f-r*: *kufr* «disbelief» (37 occurrences) and *kafara* «to disbelieve» (289 occurrences)
4. *ittaqā* «to guard oneself» (166 occurrences)

In researching this study I have considered all instances of these forms found in the Quran.

The method of semantic analysis used here is Natural Semantic Metalanguage or NSM (Goddard & Wierzbicka 2014, Wierzbicka 2021). This method of semantic analysis uses semantic decomposition, deploying 65 semantic primes, which are postulated to exist in the lexicon of every human language, and to be sufficient for the semantic explication of all linguistic meanings. In addition, NSM postulates a universal syntax for the primes. Tables of NSM primes have been drawn up for many languages (Goddard & Wierzbicka 2014: 12).

NSM has proven extremely useful for explicating culture-specific meanings that are difficult to translate (Wierzbicka 1992, 1997, 1999, 2014, Goddard & Wierzbicka 2014, Levisen & Waters 2017, Bromhead & Ye 2020), including theological meanings (Wierzbicka 2001, 2019, 2020, Habib 2011, 2017). My purpose in presenting this research is to further test NSM by using it to explicate a key group of quranic Arabic words which are core concepts of Islam and notoriously difficult to translate.

[7.] I will use *Allah* to refer to God in the Quran and *God* for biblical references.

The need to use a diversity of English words to translate certain core quranic Arabic concepts has prompted some scholars of the Quran to discern complex polysemies at every turn. As Hughes (2022: 39) puts it, "these terms can adopt a rather large range of meanings in different contexts". This is the approach taken by Juan Cole (2020) in a recent study of quranic Arabic *kafara*. However, the fact that a word in one language requires a bewildering variety of contextual translations in another language can by no means be relied upon as evidence of polysemy. On the contrary, I shall argue that the difficult-to-translate quranic words considered here can be given unitary NSM explications, and the need to deploy a variety of English translations reflects the fact that parts of a complex concept, which can be easier to translate than the whole concept, come into focus in different contexts.

A key point is that it will not do to attempt to provide semantic explications of complex concepts using other equally complex concepts. Complex meanings must be decomposed using words with simpler meanings. As we shall see, the use of NSM primes allows one to cut the Gordian knot of the supposed "large range of meanings" of these theological terms, to produce compelling semantic analyses which account well for the textual data.

4. Results

4.1. rasūl

This now brings us to our first key quranic word, *rasūl* (pl. *rusul*), conventionally translated «messenger», but sometimes «apostle».

In classical Arabic, the root *r-s-l* is used with various derivatives related to sending a message, including the form IV verb *arsala* 'send someone with a message' and the form III verb *rāsala* 'to correspond or exchange messages'. The word *rsl* is also attested in Sabaean (Old South Arabian) inscriptions referring to royal emissaries (Biella 1982: 490), and *mursal*, a nominal based on the same root, is used in the Quran with the meaning 'ambassador' (Q27:35), that is, someone sent by a ruler to deliver a message.

In Islamic theology, *rasūl* has a precise religious meaning. The Quran recites multiple stories of messengers from the past, whose biographies follow a standard pattern, summarised by David Marshall as follows:

> …the messenger will typically criticize his people for not worshipping God alone, and perhaps for certain moral failings as well. However, he is rejected by most of his contemporaries, although he does have some obedient followers. The messenger also warns his people that, if they do not repent, they will suffer a great punishment from God. The story ends with a dramatic act of divine intervention: the unbelievers, as warned, are destroyed by God in a variety of ways… The completeness of the destruction of the unbelievers is often emphasized. The messenger and his followers are saved and vindicated. (Marshall 2014: viii-ix)

An example of a *rasūl* in this mould is Mūsā (Moses), who was sent, according to the Quran, to the Egyptians (Q7:103–105) to warn them of impeding destruction. When they refused to heed the warnings, they were destroyed and Moses was rescued.

As Marshall explains, the Quran repeatedly insists on the uniformity of the biographies of the *rusul*.[8] Although there are several recurring elements in the accounts of every *rasūl* (for example, that they are mocked, that the people reject their message and that Allah rescues his *rasūl*), not all of these are defining characteristics of the office.[9] This can be seen from the Quran's account of the *rasūl* Yūnus (Jonah), who the Quran states was the only *rasūl* whose people heeded his warning and repented (Q10:98). This exceptional outcome does not make Yūnus any less a *rasūl* in the Quran's eyes, which is evidence that the usual negative response of the people is not part of the essential meaning of *rasūl*.

Here is a proposed semantic explication of the meaning of *rasūl*:

rasūl
a. someone,[10] not like many other people
b. people can say what this kind of someone is with the word *rasūl*
c. people can think like this about this someone:
d.   "this someone is very good
e.   this someone does what Allah wants
f.   this someone wants other people to do what Allah wants
g.   Allah says something to this someone because he wants this someone
      to say it to this someone's people
h.   Allah always says something like this to someone of this kind
i.   Allah says to this someone:
j.   "it is good if you say this something to the people"
k.   this someone says this something to the people after this
l.   this someone says to the people:
m.   "Allah says this to you because he wants you to do something
n.   it is good if you do what Allah wants
o.   if you do not do what Allah wants, Allah will do something very
        bad to you"
p.   this someone wants the people to think: "Allah says this to us" "

Each of the elements in this explication is emphasised repeatedly in the Quran, including that a *rasūl* is sent to his own people, that *rusul* are perfect examples of

---

[8.] I have called this the doctrine of Messenger Uniformitarianism (Durie 2018).

[9.] The Quran attributes this uniformity to the *sunna* 'way' of Allah, which it says never changes (Q33:62).

[10.] Although all the quranic messengers are male, it is far from clear that maleness should be part of the lexical explication: messengers, the Quran states, are chosen from among humans (*al-nāsi*), not from among men (Q22:75). The Quran also refers to angelic *rusul*: for example, Q35:1 calls angels "messengers with wings". However, it seems that angels are considered *rusul* in the more general sense of 'someone who delivers a message', for elements of the typical *rasūl* biography are never attributed to angels.

faithful obedience,[11] and that past *rusul* all brought the same message. This last point is stressed, for example, in the following verse:

> (1) This was our way with the messengers we sent before you: you will find no change in our ways. (Q17:77)

To call someone a *rasūl* is to attribute all the elements of this explication to this person. Thus if all someone did was bring messages from Allah to other individuals as personal, private revelations, this would not justify calling this messenger from Allah a *rasūl* in the quranic sense. To merit being called a quranic *rasūl*, someone must be sent by Allah to a people with a warning of future punishment and a call to repent.

The function of the elaborate religious meaning of *rasūl* is that the stories of past messengers are used in the Quran to establish a template which validates the mission of the quranic Messenger, named four times in the Quran as Muhammad. The semantic elements included in the meaning of *rasūl* function to validate the structure of Muhammad's own mission.

The Quran also uses a related term, *nabī* «prophet» (root *n-b-y*). Scholars have pondered how and whether *nabī* is different from *rasūl*. Jeffery (1950: 115) concluded that Muhammad "made no special distinction between the two names *rasūl* and *nabī*". Wansbrough (1977: 54) came to the same view: "rigorous and consistent distinction between the designations *nabī* and *rasūl* is not justified by quranic usage".

In essence, a quranic *nabī* is a particular flavour of *rasūl*. In the Quran, the two characteristic features of a *nabī*, in contrast to the *rasūl*, are first, that the *nabī* can be a hereditary office – "some of them are descendants of others" (Q3:34) – and second, that this hereditary office was granted to the "sons of Israel" (Q5:20; Q29:27). However, the Quran reports that this office has been taken away from the Jews, who "disbelieve in it", and given to the Arabs, because they are "a people who do not disbelieve in it" (Q6:89).

It is the orthodox belief of most Muslims that Muhammad was not only the *rasūl* of the Quran; he was also the last *nabī*, the final holder of the office of *nabī*. The Quran discounts any further continuation of a hereditary prophetic office when it declares that "Muhammad is not the father of any of your men but he is Messenger of Allah and the seal of the prophets" (Q33:40).

Let us now contrast the explication of the quranic *rasūl* with that of the biblical Hebrew *nābī'*, bearing in mind that the quranic *nabī* differs only slightly in meaning from *rasūl*.

Wierzbicka has given a description of what a biblical prophet is in *What Christians Believe*. For our purposes, the key component in Wierzbicka's explication is the following:

[11.] The doctrine of the moral perfection of messengers is known in Islam as ʿ*ismah*.

> When God wanted to say something to the people of Israel,
> God said it to someone not like many other people,
> after this, this someone said it to the people of Israel.
> Someone like this was called a prophet.
> Many prophets spoke to the people of Israel[12] at many times (some were women).
> They spoke not like other people. When they said something, they wanted people to think: "God says this to us."[13] (Wierzbicka 2019: 80)

This semantic explication needs further refinement.

In the Hebrew Bible, the verb *nābā'* 'prophesy' (root *n-b-'*) is the basic concept upon which *nābī'* 'prophet' is built. This can be seen, for example, in the fact that people can prophesy without being prophets. King Saul, who was not a prophet, prophesied on two occasions (1 Samuel 10:11, 19:23–24); hence the proverb, "Is Saul also among the prophets?", which is given as commentary on both of these incidents. The intended answer is "No, Saul is not a prophet." On the second occasion, three groups of messengers, who had been sent by Saul to take David, also prophesied, likewise without being considered prophets (1 Samuel 19:20–21). In an earlier incident, seventy elders of Israel prophesied once only at their commissioning under Moses, inspired by the Holy Spirit (Numbers 11:25–29), yet this did not make them prophets either, for "they did so no more" (Numbers 11:25). Another example is when David, who is not referred to in the Bible as a prophet, sang a prophetic song (2 Samuel 23:2–7).

Here is a proposed NSM explication of Hebrew *nābā'* 'prophesy', which develops Wierzbicka's Minimal English account:

> *X nābā'* 'X prophesied'
> a. Someone (X) said something to someone else at that time
> b. this someone (X) said it like someone can say something like this to someone else when someone thinks like this:
> c.   "God said something to me
> d.   God wants me to say it to someone else after this
> e.   God wants that someone else to think like this:
> f.   "God is saying this to me" "

Components (b–f) of this explication capture the idea that this speech act has a conventional character in which God speaks to someone wanting them to pass it on to another person so that the other person can know God is saying this to them.

In Exodus 7:1, Aaron is described as Moses' "prophet" to Pharoah: "See, I have made you like God to Pharoah, and your brother Aaron will be your prophet." Earlier, in Exodus 4:15–16, it is said that Aaron was to function as Moses' "mouth",

---

[12.] The mission to Israel is characteristic of many prophets in the Hebrew scriptures, but not the post-Pentecost prophets of the early church.

[13.] This explication is in Minimal English, which is a "highly reduced version of English" (Wierzbicka 2019: 45) built upon a core of the 65 NSM primes.

and Moses was to "serve as God" to Aaron. This means that when Aaron was acting as Moses' mouthpiece to Pharoah, people should think this:

> Moses said something to Aaron
> Moses wanted Aaron to say it to Pharoah after this
> Moses wanted Pharoah to think like this:
> "Moses is saying this to me"

Many biblical passages depict prophesying as a two-stage process, as described in the semantic explication. The first stage, when God speaks to the one prophesying, is often described as "the word of the LORD came to X", where X is the prophet's name (e.g. 1 Samuel 15:10). In Ezekiel 3:1–4, the reception of the message is enacted graphically through a vision in which the prophet eats a divine scroll, after which he is instructed to take these consumed words to heart and then go and repeat them to Israel: "Mortal, all my words that I shall speak to you receive in your heart and hear with your ears; then go to the exiles, to your people, and speak to them." (Ezekiel 3:10–11).

The semantic explication proposed here of *nābā'* implies that a false prophecy can still be called an act of prophesying, as indeed happens in Jeremiah 23:21: "I did not send the prophets, yet they ran; I did not speak to them, yet they prophesied." It is consistent with the explication of *nābā'* that someone could prophesy lies, without actually thinking that "God said something to me". This appears to be the case in Zechariah 13:3: "If anyone continues to prophesy, his own father and mother will tell him, 'You must die, for you have prophesied lies in the name of the Lord.'" The way the explication is worded, it suffices that the one prophesying is doing this in the **manner** of someone who thinks God has said something to them.

Note also the comparative simplicity of the meaning of biblical Hebrew *nābā'*, which does not require that the prophesying be directed to the nation of Israel: it could be directed to individuals, as often happens in the Bible (e.g. the prophecy of the old prophet of Bethel in 1 Kings 13:20–22). Moreover, the prophetic message need not include a warning: there is no restriction on its contents.[14]

Here is a proposed explication of the biblical Hebrew *nābī'* 'prophet':

nābī'
a. someone, not like many other people
b. people can say what this kind of someone is with the word *nābī'*
c. people can think like this about this someone:
d.   "God can say something to this someone
e.   God does it because he wants this someone to say it to someone else after this
f.   when this someone says it to that someone else God wants that someone to think like this:
g.   "God is saying this to me" "

[14.] In the Bible, prophesying is not even limited to the messages from the God of Israel (see e.g. 1 Kings 18:19; Jeremiah 23:13): people can also be said to prophesy in the name of other gods. However, for the sake of comparison with the quranic *rasūl* I will here limit my focus to prophesying in the name of God.

As we have seen, the meaning of *prophesy* allows for the possibility that someone could prophesy without being considered a prophet, as well as for an act of prophesying to be insincere or false. What qualifies someone as a prophet is whether people think that the person can prophesy: this is a socially recognised role. If someone like Saul, who was not thought of in that way, did prophesy on occasion, this was insufficient in itself to justify speaking of him as someone who could prophesy: an isolated act of prophesying does not make someone a prophet.

The explications I have proposed here differ from Wierzbicka's characterisation in that she has the prophets addressing the people of Israel, and sets this in the context of Israel's overall salvation history. While this is a valid attribute of the New Testament concept of *hoi prophetai* 'the prophets' (e.g. Matthew 2:23; Acts 3:25; 1 Peter 1:10), it is not a necessary part of the meaning of the Hebrew *nābī'* as it is used in much of the Hebrew Bible.

It should be apparent that there are major differences between the biblical *nābī'* and the quranic *rasūl* / *nabī*. Whereas in biblical Hebrew there is a distinct speech act of prophecy, there is no such speech act for the quranic *rasūl*: there is no Arabic verb meaning 'to prophesy'. The phrase *al-balāgh* 'the reaching, the attaining' can refer to the delivery of a message from Allah by a *rasūl*; the verb *talā* 'recite' can refer to delivery of verses from the Quran; the verb *arsala* 'send' can refer to Allah's act of sending a messenger or signs to a people; and the verb *awḥa* 'suggest, inspire' can refer to the process in which verses are send down to the *rasūl*. However, no Arabic verb is attested in the Quran that describes the whole prophetic process of a *rasūl* receiving and then delivering a message from Allah to others.

There are other differences. Biblical prophecy knows no limitations on the kind of message brought, but the quranic prophet brings an unchanging message that warns of imminent punishment. The biblical prophet can bring a message for an individual, group or nation, but the quranic prophet is always sent to a community, such as a tribe or a town. Biblical prophecy takes place in the context of a long history of God's communications with the people of Israel, in which each individual prophet contributes to a conversation spanning centuries – this is reflected in Wierzbicka's explication of the role of the prophet – but the quranic *rasūl* is sent to their own people to give them a unique, one-time opportunity to repent before they are destroyed. In the Bible, there are no examples of a hereditary biblical prophetic office, unlike the other two anointed offices of priest and king which are hereditary, but the quranic *nabī* can be a hereditary office. The quranic *rasūl* is righteous by definition, but in the Hebrew Bible a *nābī'* is not necessarily attributed with righteousness.

These many differences point to difficulties for translation. The question arises: Is it even appropriate to use a biblical term, such as *prophet* or *apostle,* for the quite distinct quranic office of *rasūl* (or its variant, the *nabī*)? In the case of *rasūl*, the majority solution, adopted both by scholars and by English-speaking Muslims, is to use 'messenger' as a calque for *rasūl* and 'prophet' as a calque for *nabī*. Thus Muhammad's title, *al-rasūl Allah*, is customarily rendered in English as

"the Messenger of God". A less frequent translation is "the Apostle of God", no doubt chosen because the Greek *apostolos* is derived from *apostello* 'send', which is similar to the core meaning of the root *r-s-l*. However, the New Testament Greek *apostolos* differs even more in meaning from *rasūl* than does Greek *prophetes*.[15]

In the discussion of *rasūl*, we have considered a concept that has some degree of similarity to a biblical concept, and a stable CTE of «messenger», but this translation is inaccurate, because a quranic *rasūl* is much more than just a messenger. At the same time, the quranic *nabī*, although translated as «prophet», is very different from the biblical prophet, having a meaning similar to and based on that of *rasūl*.

These translation difficulties are compounded by the fact that the Quran considers the faith it preaches to be the true Judaism and the true Christianity, stating that Abraham was neither a Christian nor a Jew but a Muslim (Q3:67), so its concept of a *rasūl* is believed to be original and authentic to both Christianity and Judaism.

Now we shall consider two terms that present a different kind of challenge for translation.

4.2. shirk

The Quran uses two primary lexical concepts to refer to someone who is not a Muslim. For each concept there is an abstract noun, a verb, and an agent nominal:

Root	Abstract Noun	Verb	Agent Nominal
sh-r-k	shirk	ashraka	mushrik (pl. mushrikūn)
k-f-r	kufr	kafara	kāfir (pl. kuffār / kāfirūn)

We will be proposing NSM analyses of the abstract noun *shirk* and the plural of *kāfir*.

The concepts of *shirk* and *kufr* are distinct but complementary, referring to different dimensions of disbelief. In a nutshell, a person who is a *kāfir* rejects Allah, his signs and his messengers while concealing or denying the truth, while a person who is a *mushrik* transgressively claims that another being shares in Allah's unique powers and prerogatives.

Derivatives of the Arabic root *sh-r-k* are grounded in a core meaning of shared ownership. A co-owner of an animal or a slave is a *sharīk*. The form III verb *shārika*, which does not occur in the Quran, means 'he is a co-owner of something with someone' and the causative form IV verb *ashraka* means 'he makes or treats someone as a co-owner'. As is the case for some other form IV verbs, for *ashraka* the causative function includes declaring someone a co-owner.[16]

---

[15] The complexity of inter-religious cross-cultural communication is exacerbated because Arabic-speaking Christians use the word *rasūl* to translate *apostolos*.

[16] For example, compare the form I verb *nakira* 'not know' with form IV *ankara* 'disavow, declare someone to be unknown'.

In the Quran, the noun *shirk* refers either to a share or part ownership of something (e.g. Q35:40, "Do they have a **share** (*shirk*) in the heavens?") or to the sin of attributing "partners" to Allah (e.g. Q31:13, "O my son, do not **attribute partners** (*ashraka*) to Allah, for *shirk* is the worst of wickedness").

Shirk is the "foremost religious crime in Islam" (Böwering 2002: 329), and is considered a gross, uniquely unforgivable sin. In the understanding of the Quran, it is a terrible transgression to attribute "partners" to Allah. Q4:48 (repeated word-for-word in Q4:116) states that Allah can forgive any sin except attributing partners [*ashraka*] to him.

The term *mushrik* is used frequently in the Quran for someone who calls someone an associate or 'co-owner' with Allah. Although the root meaning of *sh-r-k* is co-ownership, this is broadened to include the attribution of the powers and prerogatives of Allah to other beings, in particular as pertains to Allah's benevolence. Thus a *mushrik* is someone who "in his behavior and attitudes … proceeds as if other beings, supernatural or perhaps sometimes human, have powers which a true monotheist would recognize as belonging to God alone" (Hawting 2002: 477). An example is someone who calls upon someone other than Allah for help of a kind that only Allah can provide.

Underlying the concept of *shirk* is the conviction that, as the sole creator, Allah is the only 'owner' and disposer of human beings. That it is a logical impossibility for Allah to enter into any kind of partnership over his creation is explained by means of an analogy with the plight of a slave owned by two masters, which compares unfavourably to the state of a slave owned by just one master:

> (2) Allah presents a parable: a man was owned by several quarrelling **partners** (*shurakā'u*) and a man belongs exclusively to one man. Are the two equal? (Q39:29)

This analogy takes it to be obvious that the condition of a co-owned slave is greatly to be pitied. Such a slave will be pulled between two masters quarrelling over the slave. The slave, unable to meet the conflicting commands of the two masters, can please neither of them, for "he would be confused as to whom of them he should serve" (al-Maḥallī & al-Suyūṭī 2007, commentary on Q39:29). For humans to imagine that they serve several gods when there is in fact only one God, Allah, would be calamitous indeed for them. For Allah's part, the Quran goes on to declare that those who make such a claim are "telling a lie about Allah" and for these a special place in hell has been preserved (Q39:32). In another parable, the Quran states that for Allah to have a *sharīk* would be like a slave owner co-owning wealth with his slave (Q30:28). Just as no slave owner would tolerate this, neither will Allah.

Aversion to co-ownership of creation is spelled out in another passage, where it is claimed that if there were more than one creator god, each would have tried to dominate the other, fighting the other over creation (Q23:91). For this reason, Q21:22 concludes that the creation would have been wrecked if there had been more than one creator god (Mir 2004: 161).

This argument against polytheism appears to be original to Islam: such arguments are not found in the Bible. When Jesus tells a parable of a slave with two masters (Luke 16:13, Matthew 6:24), he uses the analogy to warn that a slave with two masters will have divided loyalties, and a person must choose whom they will serve. This is an argument for the exercise of freedom of choice by human beings: as Joshua said to Israel, "choose for yourselves this day whom you will serve" (Joshua 24:15). In contrast, the Quran uses the analogy of a slave with two masters to make a strikingly different point, namely that, just as it is a terrible thing to have two masters because they would compete with each other and make life a torment for the slave, so it would be a disaster for humanity if Allah had co-owners of creation.

The Quran's polemic against *shirk* focuses on the impossibility that another being could exert the power of Allah. It asserts that no-one but Allah could raise the dead (Q21:21); no-one has the right to question Allah about anything he does (Q21:23); those whom some call 'sons' of Allah are but Allah's slaves, who can only speak or act by his command (Q21:27); and no other 'gods' can defend their servants from Allah, who can do what he wants with such 'gods' (Q21:43), including sending them to hell (Q21:29). The command to serve or worship Allah alone (Q21:25) derives logically from these considerations, since someone should only serve another if they have power to do with you what they want.

It seems also that the concept of *shirk* is invoked in the Quran whenever a beneficial power is attributed to another, of a kind which only Allah can exert. To seek help from Allah alone protects against committing *shirk*:

(3) Say: "I call only upon my Lord, and I do not **attribute partners** (*ushriku*) to him." (Q72:20)

Shirk is thus an error of attribution, saying that another has a beneficent power to command creation which in reality only Allah has (Q18:26; Q30:40). Various verses describe *shirk* in terms of saying falsehoods about Allah:

(4) Say: "My Lord has only forbidden … that you **attributed partners** (*tushrikū*) to Allah without his authority, and you said things about Allah of which you had no knowledge." (Q7:33)

Shirk can manifest in a variety of different ways, such as misdirected prayer or worship, verbal confession, or reliance on some other 'god' for aid. The Quran calls it *shirk* when someone looks to anyone, whether human or a spirit, to do something for them which only Allah can do. An example is taking someone other than Allah as a patron or protector:

(5) Shall I take for my patron (*walī*) any other than Allah, the maker of the heavens and the earth? … Say, "No! I am commanded to be the first to bow to Allah. Do not be one of the **associators** (*mushrikīna*)." (Q6:14)

Another example of *shirk* is a confession of faith that conflicts with Allah's prerogatives; for example, saying that Jesus Christ is divine (Q5:72) or talking about Allah as having sons or daughters. On judgement day, Allah will say to all who do this, "Where are the partners you talked about?" (Q6:22) The point of this statement is that these alleged "partners" will be powerless to aid their devotees at the Last Day, and thus prove not to be partners of Allah in his rule. Another example of *shirk* is giving thanks to someone else besides Allah for the birth of a child (Q7:190), which is wrong because only Allah can give life.

Although some translators render *shirk* as «idolatry», the word *ṭāghūt* 'idol' is never mentioned in the Quran in the same verse as an instance of the root *sh-r-k*. Where the concept of *sh-r-k* is invoked is in contexts when the Quran emphasises that there is only one creator (Q35:40) ruling over the creation, and particularly when the focus is on Allah's beneficial mastery over everything. In such contexts, the Quran repeatedly reminds us that those who look to others besides Allah for aid will find their alleged protectors to be powerless before Allah (Q10:28; Q28:64), for "he has no partners in his rule" (Q18:26). Thus the focus in contexts where *shirk* is invoked is not on idolatry per se, but on the exclusivity of Allah's rule and the imperative for human beings to rely only on the benevolence of Allah.

The being to whom Allah's attributes are falsely attributed in an act of *shirk* can be an angel or an (imagined) god, but it can also be human, as reflected in the following verse naming priests, monks and Jesus (understood by the Quran to be no more than a human being) as objects of 'association':

(6) They take their priests and their monks as their lords to the exclusion of Allah, and Christ the son of Mary; yet they were commanded to worship but one god: there is no god but he. Praise and glory to him. He is glorified above what they **associate** (*yushrikūna*). (Q9:31)

Finally, we note that in the Quran, *shirk* is often said to lead to divine punishment:

(7) We will cast terror into the hearts of those who disbelieved (*al-ladhīna kafarū*), because they **attribute partners** (*ashrakū*) to Allah, for which he has sent down no authority. Their refuge is the Fire: evil is the lodging of evildoers. (Q3:151)

In the light of all these considerations, it seems unnecessary to define *shirk* in terms of ownership: it is enough to focus on the unique power of Allah as benevolent master over his creation. The offense of attributing 'associates' to Allah consists in looking to other beings for help that only Allah can provide. This is bad for the person, because such help is illusory, and it denies Allah his due.

Here is a proposed semantic explication of *shirk*, which gathers these insights together. In this explication the container for the explication (lines a–d) follows the semantic analysis of abstract nouns in Goddard and Wierzbicka (2014: 205–237).

shirk

a. something
b. people can say what this something is with the word *shirk*
c. someone can say something about something with this word when
    someone thinks like this:
d.   "it can be like this:
e.   someone says about someone else:
f.   "this someone can do something very good for me"
g.   at the same time, people can know that it is like this:
h.   Allah can do something like this for people, no-one else can
i.   this someone is not Allah
j.   it is very bad if it is like this
k.   people can know that if someone says something like this, after some
    time Allah will do something very, very bad to them because of it"

4.3. *kuffār / kāfirūn*

We will now consider the concept of *kufr*, proposing an explication of the plural of *kāfir* «disbeliever». This noun has two main plural forms, *kuffār* and *kāfirūn*, which are identical in meaning.[17]

The concept of *kufr* is even more frequently invoked in the Quran than *shirk*. As Toshihiko Izutsu has explained in *Ethico-Religious Concepts in the Qurʾān*:

> "Even a cursory reading of the Scripture [i.e. the Quran] will convince one that the role played by the concept of *kufr* is so peculiarly influential that it makes its presence felt well-nigh everywhere in sentences about human conduct or character." (Izutsu 1966: 119)

In its non-religious sense, the verb *kafara* means 'cover, conceal' (Adang 2001: 220); for example, clouds covering the sky or the earth covering seed.[18] While we are interested here in derivatives with *k-f-r* that refer to disbelief and rejection of faith, there are *k-f-r* forms in the Quran that do not have this meaning. The form II verb *kaffara* means 'to grant a pardon' (i.e. to cover over someone's transgression); in Q57:20, *kuffār* is used with the sense 'sower' (someone who covers over seed with earth); the noun *kaffārat* is used with the meaning 'atonement' in the fifth Sura (Q5:45, 89, 95); and in the following verse the noun *kufrān*,[19] a *hapax legomenon* in the Quran, is used in the context of affirming that a person's good deeds will not be omitted from their record:

(8)   Whoever does righteous deeds, as a believer, (there will be) no **denying** (*kufrān*) his striving (by Allah). We shall record it to his account. (Q21:94)

[17.] Ambros & Procházka (2004: 239) report two other forms, *kafara(t)* and *kāfira(t)*, as well as the feminine plural *kawāfir*, but these are all rare.

[18.] This root is cognate with the Hebrew *kāpar* 'to cover, atone'.

[19] Some translators render *kufrān* here as 'ingratitude', but it seems inconsistent with the Quran's theology to attribute gratitude or ingratitude to Allah.

Kāfir as a non-religious term can also mean 'ingrate'. In the following example, Pharoah rebukes Moses, presumably for killing an Egyptian (Exodus 2:11–12). Moses' crime is not specified in the Quran. The point of the rebuke is that Moses should have been grateful to the Egyptians for raising him:

(9) He (Pharoah) said (to Moses): "Didn't we bring you up among us, and weren't you among us for some years of your life? Yet you did what you did, and are one of **the ungrateful** (*al-kāfirīna*[20])." (Q26:18–19)

Setting aside these non-religious meanings, we will focus on *kafara* «disbelieve», *kāfir* «disbeliever» (pl. *kuffār / kāfirūn*), and *kufr* «disbelief», as used to characterise disbelief in and rejection of the message of Islam. For the purposes of discussion here, and in the absence of any convincing evidence to the contrary, I will treat these three forms as semantically equivalent. Thus a *kāfir* is someone who does *kafara*, and *kufr* refers to the act of committing *kafara* or being a *kāfir*.

In the Quran, the verb *kafara* is often used intransitively, but it can also take a second argument introduced by the preposition *bi-* (e.g. 'they **rejected** (*kafarū bi-*) Allah and the Messenger' (Q9:54)). As a religious term used for rejection of the Quran's message, *kafara* receives a diversity of renderings in the English translations, including 'disbelieve', 'be an unbeliever', 'misbelieve', 'be an infidel', 'be ungrateful', 'deny', 'knowingly reject the truth', and 'be without faith'.

One of the most salient aspects of the religious meaning of *k-f-r* is ingratitude: indeed, it was this meaning that the classical Muslim lexicographers considered to be *kufr*'s fundamental meaning (Adang 2001: 221). For example, in this next verse the people of the town are criticised for being ungrateful for Allah's good deeds which have benefited them:[21]

(10) Allah tells a parable: a town was secure and at rest, with abundant provision coming to it from every side, but it **was ungrateful** (*kafarat bi-*) for Allah's favours, so Allah clothed it with hunger and fear for what they had been doing. (Q16:112)

The Quran speaks of *kufr* as the opposite of thankfulness:

(11) If you remember me, I will remember you. Be thankful to me (to Allah), and do not **be ungrateful to me** (*takfurūnī*). (Q2:152)

[20.] This is a participial form of *kafara*.

[21] It is noteworthy that the expected gratitude is associated with awareness of Allah's benevolent acts, not with Allah's inherent goodness. Allah is never called 'good' in the Quran and 'the good' is not one of his ninety-nine "beautiful names" (Böwering 2002: 317). Moreover, there are references in the Quran to Allah as the source of bad as well as good. For example, Q91:8 speaks of Allah inspiring debauchery in humans and Satan also repeatedly declares that, since Allah has led him astray, he will in his turn devote himself to leading people astray (Q7:16; Q15:39). In Q38:82, Satan explains that he will do this by the power of Allah. Furthermore, there are many quranic references to Allah leading humans astray (cf. Q2:26; Q6:125).

The Quran also often refers to a human tendency to ingratitude, and contrasts this with Allah's many acts of benevolence to humankind, which ought to evoke gratitude:

(12) It is Allah who created the heavens and the earth, and sends down water from the heavens, and brings forth fruits to provide for you. And he has subjected ships to you, to pass through the seas by His command, and he has subjected the rivers to you. And he has subjected the sun and moon to you, both constant (in their courses), and he has subjected night and day to you. He has given you some of all that you have asked for. If you (attempt to) count Allah's favours (to you), you will not be able to. Surely humans are unjust and **ungrateful** (*kaffārun* – emphatic form of *kāfir*). (Q14:32–34)

Such instances of Allah's goodness are included among the "signs" of Allah, which those who commit *kafara* call a lie:

(13) But those who are **ungrateful** (*kafarū*) and call our signs a lie shall be companions of the Fire. They will stay there in it. (Q2:39)

The concept of *kufr* is not only about ungrateful denial: it also conveys the idea of disbelief.

In several passages, forms with *k-f-r* and *'-m-n* 'believe' are contrasted. For example, in the following verse the *kāfirūna* are contrasted with the *mu'minūna* 'believers':

(14) Oh you who believe! If any of you turn back from his faith, Allah will raise up a people whom he will love as they love him: (acting) humbly towards believers (*mu'minīna*), and powerful against **disbelievers** (*kāfirīna*), fighting in the way of Allah … (Q5:54)

The following passage also contrasts *kufr* and belief, and associates *kufr* with rejecting Allah's signs and preferring a different way:

(15) Do those who **disbelieve** (*kafarū*) not see that the heavens and the earth were joined together, and we separated them, and made all living things from water? Will they not believe (*yu'minūna*)? And we set firm mountains on the earth, so that it does not shake, and we placed passes in it (between the mountains) as pathways, so that they might be guided. And we established the sky as a guarded roof. Yet they turn away from its signs. (Q21:30–32)

The idea of rejection – that the person who practices *kufr* knows full well the good things Allah has done, and not only dislikes this, but rejects this knowledge and what it implies – is apparent in the following verse in which the 'People of the Book' (Jews and Christians) are said to deliberately conceal the truth:

(16) You People of the Book, why **reject** (*takfurūna*) the signs of Allah, of which you are witnesses? You People of the Book, why do you mix truth with falsehood, and hide the truth, while you know (what the truth is)? (Q3:70–71)

The rejection can be of a teaching, such as that of a future resurrection, in which case *kafara* could be translated as 'disbelieve'. Note also in the following verse the threat of punishment for those who practise *kufr*:

(17) If you are astonished (at their lack of faith), it is astonishing that they say, "When we have turned to dust, shall we really be in a new creation?" Those are the ones who have **disbelieved** (*kafarū*) in their Lord, and those – iron chains will be on their necks – those are the companions of the Fire, where they will remain. (Q13:5)

Another aspect of *kufr* is rejection of Allah's messengers and their unvarying message of the unity of Allah:

(18) They are amazed that a warner has come to them from among themselves. The **unbelievers** (*kāfirūna*) say "This is a sorcerer, a liar. What! Has he made all the gods into one Allah? That is amazing!" (Q38:4–5)

(19) Praise be to Allah, who created the heavens and the earth, and made darkness and light. Yet those who **disbelieve** (*kafarū*) treat others as equal to their Lord. (Q6:1)

In the following verse this rejection of the Messenger is coupled with the accusation of concealing what they had known to be true:

(20) How will Allah guide a people who have **disbelieved** (*kafarū*) after they believed and have borne testimony that the Messenger is true and that clear signs have come to them. … Surely those who **disbelieve** (*kafarū*) after they believed and then increase in **disbelief** *(kufr)* – their repentance will never be accepted. They are ones who have gone astray. (Q3:86, 90)

Allah does not love someone who chooses the path of *kufr*, rejecting Allah and the Messenger:

(21) Say, "Obey Allah and his Messenger." But if they turn back, Allah does not love **disbelievers** (*kāfirīna*). (Q3:32)

The commission of *kufr* is associated with future punishment:

(22) The penalty of those who **deny** (*kafarū*) their Lord is hell, an evil destination. (Q67:6)

The idea that the *k-f-r* forms express a range of distinct meanings in reference to rejectors of Islam is commonplace in Islamic studies scholarship. For example, Camilla Adang (2001: 220) distinguishes four distinct meanings for *kafara*: 'to ignore or fail to acknowledge', 'to reject' (or 'spurn'), 'to be ungrateful' (or 'thankless') and 'to disbelieve'.

In a recent article, Juan Cole (2020) has argued that *kāfir* should not be translated as 'infidel' or 'unbeliever'. Instead, he proposes multiple polysemies for *k-f-r* forms. When *kafara* is used intransitively, Cole proposes that it has a

polysemous range of meanings which he described as 'fluid' (2020: 627). The distinct meanings he distinguishes for *kafara* include 'be ungrateful', 'reject, deny, disbelieve' (these three glosses are intended to point out a single meaning), 'worship the gods', 'rebel', 'be impious', 'be morally dissolute, be a libertine', 'disobey', 'blaspheme', 'become apostate', 'paganise', and 'commit a heresy'. Cole also proposes polysemous meanings for *kāfir*, which include 'peasant', 'pagan', 'rebel', 'blasphemer', and 'libertine', while *kufr* can, he suggests, in addition to 'disbelief', also mean 'inauthenticity' and 'bad faith'.

One can agree wholeheartedly with Cole's conclusion that 'infidel' is an inadequate all-purpose translation for *kāfir*, but without being persuaded by his reasoning.

Cole's main argument against translating *kāfir* as 'infidel' or 'unbeliever' is that those who are called *kuffār / kāfirūn* in the Quran did believe in a god or gods, so it is wrong to call them unbelievers. For example, he writes, "… the pagans have a religion, but it is simply castigated as a false one, which makes translating *kāfir* as 'infidel' seem odd" (Cole 2020: 619). However, this is just what the English words *unbeliever* and *infidel* mean: they do not mean 'someone who does not believe in a god', but are insider terms for those who do not believe as the insiders believe. To illustrate, one of the meanings the *OED* gives for *infidel* is: "From a Christian point of view: An adherent of a religion opposed to Christianity; esp. a Muslim, a Saracen (the earliest sense in English); also (more rarely), applied to a Jew, or a pagan."[22] Cole seems to be objecting to what these English words actually mean, and wants to use them in accordance with what he believes they ought to mean.

A more serious difficulty in Cole's approach to lexical semantics is his method of semantic analysis. He uses English as his semantic metalanguage, but offers neither semantic decomposition nor stable definitions of individual meanings. Whenever the context suggests a different English translational equivalent for *kafara*, Cole takes this as evidence that yet another polysemous meaning has been unearthed. For example, in Q2:102, where devils are said to commit *kafarū* by diverting people into the practice of magic, Cole rejects Arberry's translation, 'disbelieve', and proposes 'blaspheme' instead:

Of what, however, did this act consist? It does not appear to have been a denial of anything, but rather was a blasphemous activity. The humans were eager to have the teaching of the two angels of Babylon, *Hārūt* and *Mārūt*, which they then desecrated by turning it into dark arts so as to separate spouses from one another.

---

[22] "infidel" *OED Online*. Oxford University Press, December 2021 (accessed 26 March 2022). Cole is not unaware that *kāfir* could be considered an insider term used for outsiders, for he writes: "there is, of course, a sense in which it [the Quran] views them [the *kuffār / kāfirūn*] as outside of and antagonistic to the true faith, part of what translators who used the term 'unbeliever' wished to convey." His counter to this is: "I would argue, however, that there is a key lexical difference between a denier of God and an affirmer of God who gets God wrong" (Cole 2020: 626). This misses the point that, in their actual lexical meaning, the English words *infidel* and *unbeliever* are insider terms used for outsiders.

The demons' instruction harmed people rather than benefited them, and turning to the occult deprived these individuals of any portion of heaven. (Cole 2020: 624).

Here Cole offers no evidence that *kafarū*, in context, means 'blaspheme', other than his judgement that what the demons did was sacrilegious.

In another example, Cole suggests that *kafara* in Q9:74 means 'commit apostasy'. This verse speaks of people who had formerly accepted Islam but later disbelieved (*kafarū*). One might just as well conclude that *kafarū* here means 'make a bad decision', 'act unwisely', 'be double-minded', or 'be losers'.

The other problem with Cole's semantic approach is the lack of stable, testable definitions. This is apparent in Cole's discussion of the phrasal expression *kafara bi-*, which he claims has a single, stable meaning: "this phrasal verb is not polysemous in the Quran for it always means to deny or reject" (Cole 2020: 618). Later, Cole adds 'disbelieve', glossing *kafara bi-* as 'deny, reject, disbelieve' (2020: 634).[23] This scattergun approach to semantic analysis veils the meaning of *kafara* behind a cluster of English words, *deny*, *reject* and *believe*, each of which means something different.

I agree with Cole's conclusion that *kāfir* does not mean 'infidel' and *kafara* does not mean 'disbelieve'. However, my argument is neither that to attribute insider meanings to *infidel*, *unbeliever* or *disbelieve* is illegitimate, nor that *kafara*, when used in reference to rejectors of quranic faith, is multiply polysemous. (Certainly the non-religious uses of *k-f-r* forms include polysemies.) Instead, my argument is that the family of *k-f-r* forms has a complex semantic core which needs to be precisely defined. This precision cannot be achieved by simply listing English translational equivalents.

Contrary to Cole, I propose that the Quran treats the semantic components of the religious uses of *k-f-r* forms as bound together into one concept, not many. It is not that there are different kinds of *kufr*, or different kinds of people known as *kāfir* – *kāfir₁*, *kāfir₂*, *kāfir₃*, etc. – but that the single state of *kufr* has different aspects to it, and the one kind of people known as the *kuffār / kāfirūn* display a range of characteristics, which include not only ignorance, rejection of truth, ingratitude and disbelief, but also wilful disobedience and incurring the wrath of Allah.

In formulating a semantic explication for *kuffār / kāfirūn*, I have treated it as an insider term. *Kuffār / kāfirūn* and the other *k-f-r* terms divide human beings into insiders, who are grateful believers, and outsiders, who are ungrateful disbelievers. This 'insider' meaning can be defined in NSM using a 'people of two kinds' explication. An explication is provided for the plural form, because the large majority of instances are plural. Here, then, is our proposed semantic explication for *kuffār / kāfirūn*:

---

[23] One would need to add 'be ungrateful' to this list because of examples like the following:

Allah presents a parable: a town was at peace and content, abundant provision coming to it from every side. Yet it **was ungrateful** (*kafarū bi-*) for the favours of Allah. So Allah let it taste hunger and fear … (Q16:112; cf. also Q16:72).

kuffār / kāfirūn

a. people of one kind
b. they are not as Allah wants people to be

c. it is like this:
d. there are people of two kinds
e. people of one kind are as Allah wants people to be
f. people of this kind think like this:
g. "I know that it is like this:
h. Allah does many very good things for people
i. Allah does not have to[24] do these things
j. because of this, I want to think very good things about Allah

k. at the same time, people of this kind think like this:
l. "I know that it is like this:
m. Allah says many things to people, these things are true
n. Allah sometimes says these things to some people not like other people
o. afterwards these people say these things to everyone else
p. because of this, people can know what Allah says
q. I want to know what Allah says
r. I want to do everything as Allah wants"

s. people of the other kind are not like this
t. they are not as Allah wants people to be, they are like this:
u. they don't want to think: "Allah does good things for people"
v. they don't want to think something good about Allah because of this
w. they don't want to know what Allah says
x. they don't want to do everything as Allah wants
y. they want to do other things, not as Allah wants people to do
z. when Allah says things to these people, they say
a'. "these things are not true"

b'. it is very, very bad if people are like this
c'. people can know that if people are not as Allah wants them to be, after some time Allah will do something very, very bad to them because of it

In this explication, *kuffār / kāfirūn* encompasses ingratitude, wilful rejection of Allah's messages and his messengers, rebellion against the truth, choosing to act in ways that are contrary to Allah's commands, and rejecting or ignoring Allah's signs and the threat of punishment. The comprehensiveness of this explication means that a disbeliever in Islam is, as a matter of course, considered to be guilty of a wholesale rejection of Allah's commands, of his messengers, and of the truth. This is indeed what the concept of *kuffār / kāfirūn* conveys.

[24.] The predicate *have to* in *X has to do Y* is shorthand for *X can't not do Y*. See Goddard (2014) for a broader discussion of modal verbs of necessity.

Equivalent semantic components will need to be included in semantic explications of the verb *kafara* and the noun *kufr*.

A question that arises is whether some of these additional elements added to the meaning of *kafara* could be incidental inferences that might arise in particular contexts, but are not intrinsic to the semantic structure of *kuffār / kāfirūn*. If someone is said to do *kafara* or be guilty of *kufr*, would all the elements of this explication be attributed to them? I believe they would. Not only have the exemplar verses cited here been carefully chosen to stand for many other similar verses – they are by no means cherry-picked – but to say someone is a *kāfir* is indeed to accuse the person of rejecting Allah's messengers by calling them liars, of denying the truth of Allah, of disobedience, of coming under the wrath of Allah, and of ingratitude. In the divided worldview of the Quran, anyone who refuses to embrace Islam and rejects the Messenger and his message is guilty of all this. The verses of the Quran reinforce these judgements repeatedly.

The combination of qualities in this semantic explanation reflects the way the Quran construes disbelief as a full rejection of the quranic Messenger and his message, in its entirety, by people who should know better. By attributing ingratitude and denial to anyone who does not accept the Quran's message, all disbelievers in Islam are stigmatised.

This does not mean, however, that all elements of this semantic explication are in focus in every instance in which *kāfir*, *kafara* or *kufr* are used. Consider, for example, the uses of *kafara* in Q2:102, which Cole took to have the sense of 'blaspheme':

> (23) …and they follow what the devils recited for Solomon's kingdom. Solomon did not **disbelieve** (*kafara*), but the devils **disbelieved** (*kafarū*). They taught humans sorcery, and what was sent down to Babylon's two angels, Hārūt and Mārūt. Whenever they taught anyone they said, "We are but a test; do not **disbelieve** (*takfur*)." From those two they (the humans) learned how to divide a man from his spouse. Yet they did not harm anyone in this way, except by Allah's permission. And they learned what harmed them, and what did not profit them. And they knew full well that whoever buys it [magic] will have no share in the world to come. Evil is the price for which they sold themselves. If only they had known! (Q2:102)

What aspects of *kufr* are in play here? In Q2:102, I propose that the reason the devils of Solomon's time are said to *kafarū* is that they were enticing people to use magic to fulfil their desires instead of looking to what Allah says, doing what Allah wants, and depending upon Allah's benevolent provision. The human magic practitioners were **not** thinking, "Allah does many very good things for people" (lines h, u in the explication), "I want to know what Allah says" (lines q, w), or "I want to do everything as Allah wants" (lines r, x). Instead, "they want to do other things, not as Allah wants people to do" (line y). In all this they were acting as *kuffār / kāfirūn*, as defined in the semantic explication.

It is also relevant to consider the immediately preceding and following verses in Q2:100–105. The preceding verses make clear that Q2:102 is about 'People of the Book' (in this case, Jews) who had discarded their covenant and rejected what their messenger brought, "throwing away the book of Allah" (lines z–a'). Furthermore, Q2:103 speaks of these past People of the Book abandoning faith and not refraining from evil (lines w–y); Q2:104 speaks of "people of Faith" in the present moment speaking duplicitously and disrespectfully, rejecting the quranic Messenger (lines z–a'); and Q2:105 states that such people without faith "do not want anything good to come down to you from your Lord" (line u). This surrounding context makes clear that these people were rejecting what a messenger had brought to them from Allah, which is a key element of the semantic explication of *kuffār / kāfirūn*.

These points all align with the semantic decomposition of *kuffār / kāfirūn* offered above. The humans' resort to magic functions as but one element in an extended discussion of *kufr* which runs through the whole passage of Q2:101–105.

We conclude that the use of *kafara* in Q2:102, in its context, is consistent with key semantic elements in the proposed explication of *kuffār / kāfirūn*. The sustained focus throughout this passage is on the *kufr* of these Jews in Solomon's time, so there is no need to posit a separate polysemous meaning 'blaspheme' to account for the instance in question.

Note, however, that in Q2:102 not **all** the semantic elements of *kuffār / kāfirūn* are in focus. Ingratitude (lines h–j, v) is not in focus in the surrounding passage. On the other hand, rejection of Allah's truth and rejection of his messengers are in focus in the surrounding verses, but not in Q2:102.

Our explication of *kuffār / kāfirūn* is not unnecessarily overloaded. There are some aspects of *kufr* which have not been included in the explication because they are entailed by the meaning as defined. For example, arrogance is repeatedly attributed to the *kuffār / kāfirūn*, as expressed by a variety of different words such as *astakbara* 'be arrogant' and *'alī* 'high, superior' (Izutsu 1966: 142–152). However, in the Quran arrogance is implied by the meaning of *kafara*: human beings who take it upon themselves to deny the truth of what Allah has spoken are by definition arrogant, since they rate their own judgement above their all-powerful and all-knowing creator's. Since arrogance is implied by the semantic explication of *kafara*, it does not need to be written into it.[25]

Kufr is a complex and unique concept, tailor-made to fit the theologically divided worldview of the Quran. It is therefore not surprising that in the religious practice of Muslims around the world, the word *kāfir* is normally not translated but is borrowed directly from Arabic. The term is universally considered to be derogatory.

In English translations of the Quran, it seems impossible to avoid using a variety of terms for forms derived from the root *k-f-r*, depending upon the context.

[25.] The same can be said of the whole range of attributes of *kufr* which Izutsu explores in chapters 7 and 8 of *Ethico-Religious Concepts in the Qurʾān*.

For any one instance, a translator must choose whether to emphasise ingratitude, rejection, denial or disbelief. The cost of having to make this choice in translations of the Quran is that the unity of the concept of *kufr* is concealed from the reader. This unity can only be retrieved through working with the Arabic text.

4.4. ittaqā

Our final lexical study is of the verb *ittaqā* (the root is *w-q-y)*. This is one of the most difficult quranic words to translate into English. The derivative abstract noun is *taqwā*, which Izutsu, who produced two monographs on the semantics of quranic concepts (Izutsu 1964, 1966), translated as 'fear of God', a choice which is no doubt influenced by the biblical 'fear of the Lord':

> … the idea of *taqwā* 'fear of God' … is indeed the central element of the Islamic conception of religion in general. (Izutsu 1966: 120)

A form of the verb *ittaqā* appears, for example, in the following verse, where it is translated as 'protect yourselves':

> (24) "Oh People! Serve your Lord who created you, and those that were before you, so that you may **protect yourselves** (*tattaqūna*)." (Q2:21)

The agent noun formed from the root *w-q-y* is *muttaqīn*, occurring only in the plural.

The difficulty *ittaqā* presents for translators may be gauged by the diversity of English renditions of the final word in this single verse, which include 'learn righteousness' (Ali 1946), 'attain piety' (Munshey 2016), 'guard against (evil)' (Shakir 1985), 'ward off (evil)' (Pickthall 1976), 'fear (Allah)' (Palmer 1880, Sale 1734, Rodwell 1876), 'be godfearing' (Arberry 1998), 'be saved' (Khalifa 1981), 'become Al-Muttaqoon (the pious)' (al-Hilālī & Khān 1998) and 'be Godwary' (Reynolds 2018).

Most translators use a variety of English terms to translate this one Arabic word. Other translations of the same word in English Quran translations include 'the righteous', 'act rightly', 'do right', 'reverence (Allah)', 'practise self-restraint', 'be careful of', 'be wary of', 'beware', 'do right', and 'keep your duty to'.

The root *w-q-y* occurs frequently in the Quran, mostly as the form VIII verb *ittaqā*, which can be used intransitively or transitively. The most basic verb derived from this root (form I) is *waqā,* which ditransitively means 'protect someone against something' and intransitively means 'be wary, cautious'. Occurring ten times more frequently than form I *waqā*, the form VIII *ittaqā* is in the top 2% of most frequent lexemes in the Quran. Verbs of form VIII are typically reflexive or reciprocal in meaning, but they can also signify that the subject is doing an action for their own benefit (Wright 1896–1898: I:42). This suggests that *ittaqā* could be glossed as 'he protected himself (from)'.

Throughout the Quran it is Allah who is almost invariably used as the object of the verb *ittaqā*, the one against whom people guard or protect themselves.

Although some translators render the common phrase *ittaqū allāha* as 'you (pl.) fear Allah', the contexts where *ittaqā* is used are mostly not about fear or other negative emotions, but about being rightly directed or guided. Consider, for example, the following verses:

(25)   Allah would never lead a people astray after guiding them, until he makes clear to them what they should be **guarding themselves against** (*yattaqūna*). Surely Allah knows everything. (Q9:115)

(26)   Eat of what you have taken as booty, as is lawful and good, and **guard yourself against** (*ittaqū*) Allah; surely Allah is all-forgiving, all-compassionate. (Q8:69)

The focus in these verses is on following the guidance of Allah. In Q8:69, *ittaqā* describes the lawful eating of food taken as booty as an act of "guarding yourself against Allah": this is guarding yourself because to do lawful acts keeps one on the right sight of Allah.

It is not that the Quran lacks vocabulary for fearing others or Allah. The two verbs *khāfa* (root *kh-w-f*) and *khashiya* (root *kh-sh-y*) can be translated as 'fear'. In the Quran, *khāfa* (root *kh-w-f*) is a negative, undesirable emotion, which describes human feelings towards Allah only a few times (e.g. Q5:28, 94). Although occasionally it is said that believers fear (*khāfa*) the punishment of Allah (e.g. Q13:21), most of the instances of *khāfa* are affirmations that the rightly guided will be free from fear (e.g. Q3:170: "on them will be no fear"). *Khāfa* is thus for the most part an undesirable emotion of disbelievers. In contrast, it is stated that "the allies of Allah (there is) no **fear** (*khawfun*) on them" (Q10:62).

In contrast to *khāfa*, the verb *khashiya* (root *kh-sh-y*) can be used to refer to a godly fear of Allah, often in contrast to fearing people (e.g. *fa-lā takhshawu al-nāsa wa-akhshawnī,* "do not fear people but fear me" (Q5:44)). A famous verse in Q33 reveals that Muhammad was right to have married Zainab, the former wife of his foster son Zaid. *Ittaqā* is used initially as an instruction from Muhammad to his foster son Zaid to keep his wife, which was at a time when Muhammad thought this was the right thing to do. However, this was subsequently corrected by a revelation that it was permissible for a man to marry the divorced wife of his foster son. Indeed, it was said that Muhammad himself ought to do so in order to demonstrate that this was permitted for believers in general. In this passage, Allah says that Muhammad had been holding back from marrying Zainab because he feared (*khashiya*) people instead of fearing Allah:

(27)   (Remember) when you said to the one Allah had favoured and you had favoured (to Zaid), "Keep your wife and **guard against Allah** (*wa-attaqi allāha*)." But you hid in your heart what Allah was about to reveal, and you **feared** (*takhshā*) the people, when Allah had more right for you to **fear him** (*takhshāhu*). (Q33:37)

A fundamental problem with translations that use *fear* is that *ittaqā* is not actually an emotion at all. It is something one does, not something one feels. This

contrasts with experiencing the (biblical) "fear of the Lord", which CS Lewis has described as a feeling of "numinous awe" (Lewis 1940: 5).

If 'fear' is problematic in translating *ittaqā*, the alternative of 'guard oneself' or 'protect oneself' is equally problematic. One issue is a potential conflict with the theological principle that nothing can thwart Allah's will. The English words *guard* and *protect* imply an effective exercise of power to thwart or defend against an attempted act of hostility: someone wants to do something bad to someone else, which is thwarted. That *guard* and *protect* combine with the preposition *against* is significant in this context. Thus the OED defines *protect* as "to support or assist against hostile or inimical action".[26] However, the Arabic *ittaqā* is about being careful to avoid transgression of Allah's laws. It is not about defending against and thwarting Allah's hostility, which from a quranic perspective is a theological impossibility, due to Allah's omnipotence.

Another difficulty with *guard* and *protect* as translations of *ittaqā* is that they do not capture the sense of righteousness and goodness that *ittaqā* projects. The positive good *ittaqā* conveys is reflected in translations that use the words *pious* and *careful.*

In the light of these observations, here is a proposed explication of *ittaqā,* third person masculine singular, perfect, used intransitively:

> *X ittaqā* (intr.)
> a.   someone (X) thought like this:
> b.   "something very bad can happen to me
> c.   I don't want this
> d.   it will not happen if I do something good
> e.   I want to do it"
>
> f.   because this someone thought like this, this someone did this good thing
>
> g.   people can think about it like this:
> h.   "because this someone did this good thing,
>   this very bad thing will not happen to him"
> i.   at the same time, they can think like this:
> j.   "this someone is someone good"

Here is an explication of a transitive use of the same verb with Allah as the object:

> *X ittaqā Allah* (tr.)
> a.   someone (X) thought like this:
> b.   "Allah can do something very bad to me"
> c.   I don't want this
> d.   it will not happen if I do something good
> e.   I want to do it"

[26.] "protect, v." *OED Online.* Oxford University Press, December 2021 (accessed 15 February 2022).

f. because this someone thought like this, this someone did this good thing
g. people can think about it like this:
h.   "because this someone did this good thing,
            this very bad thing will not happen to him"
i. at the same time, they can think like this:
j.   "this someone is someone good"

The semantic explication of *ittaqā* integrates a number of key components. One is the component of 'fear' (lines b–c): the person thinks something bad can happen to them if they do not do something. Another is the 'piety' element: the person wants to do something good and believes it is good to do this, and other people think the person is good for doing this good thing (lines d–e, h–j). Another is the 'guarding' aspect: because of what the person does, something bad will not happen to him. Note that the explication does not include the semantic prime *feel*: *ittaqā* involves doing and thinking, not feeling.

Note too that this explication does not suggest any hostility on Allah's part, which could be implied by a translation with English *protect* and *guard*. It is not said or implied that Allah has an intent to do something bad to the person, nor that the will of Allah could be opposed or resisted in some way. Nor is it implied that the person is only acting under compulsion: rather, the person wants to do what is good. They are not just acting prudently, but acting piously.

How, then, should one render *ittaqā* in an English translation of the Quran? The heart of the meaning is about doing something good in order to prevent harm to oneself, in accordance with Allah's direction. The problem with translations like 'be pious' or 'learn righteousness' is that they completely miss out the element of danger (line b). They are also too passive, downplaying the idea that the person is doing something. A possible translation is 'be cautious (of)', but no one translation can adequately capture all the facets of the meaning of *ittaqā*.

5. Conclusion

Words matter. To understand a culture one needs to rightly discern the meanings of its key words. While the practice of translation can offer a window into the world of another culture, it inevitably also frames and thus distorts the source text by means of the words of the target language. The instruments deployed to liberate meanings from the bondage of one language and make them known in another language are the very tools of their distortion and veiling.

Here we have considered certain key concepts of the Arabic Quran, a text which has exerted a profound and enduring influence to shape the languages and cultures of the two billion people in the world today who follow the religion of Islam.

The scholarly discipline of Islamic studies, as practised in the Western tradition, has relied on a lexicon infused with biblical concepts for understanding, interpreting and translating Islamic texts. To set aside the biblical frame, we have explicated a handful of Islamic terms using the tool of Natural Semantic

Metalanguage, applying the technique of what Goddard and Wierzbicka have termed 'experimental semantic analysis' (Goddard & Wierzbicka 2014: 11). We have argued that terms that may be very difficult to translate into English can be precisely defined using NSM semantic decompositions. The results achieved are provisional and can no doubt be improved upon. Nevertheless, I hope they have demonstrated the power of NSM semantic analysis and its value for facilitating cross-cultural awareness and understanding in the important domain of religious belief.

With that conviction, and in honour of a dear friend Igor ('Yehoshua') Mel'čuk, I submit these findings, in the hope that others will improve what has been here begun.

Acknowledgements

REFERENCES

Adang, Camilla P. 2001. Belief and Unbelief. *Encyclopaedia of the Qurʾān* 1. 218–226. Leiden: Brill.

Al-Jallad, Ahmad. 2015. *An Outline of the Grammar of the Safaitic Inscriptions*. Leiden: Brill.

Ali, Abdullah Yusuf. 1946. *The Holy Qur'an: Translation and Commentary*. Birmingham: Islamic Propagation Centre International.

Arberry, Arthur J. 1998. *The Koran Interpreted*. Oxford: Oxford University Press.

Bromhead, Helen & Zhengdao Ye (eds.). 2020. *Meaning, Life and Culture.* Canberra: ANU Press.

Cole, Juan. 2020. Infidel or paganus? The polysemy of kafara in the Quran. *Journal of the American Oriental Society* 140 (3). 615–635.

Durie, Mark. 2018. *The Qur'an and its Biblical Reflexes: Investigations into the Genesis of a Religion*. Lanham, MD: Lexington Books.

Goddard, Cliff. 2014. Have to, Have got to and Must: NSM analyses of English modal verbs of 'necessity'. In Maite Taboada & Radoslava Trnavac (eds.), *Nonveridicality and evaluation: Theoretical, computational and corpus approaches*, 50–75. Leiden: Brill.

Goddard, Cliff & Anna Wierzbicka. 2014. *Words and Meanings: Lexical Semantics Across Domains, Languages and Cultures*. Oxford: Oxford University Press.

Habib, Sandy. 2011. Angels can cross cultural boundaries. *RASK (International Journal of Culture and Communication)* 34. 49–75.

Habib, Sandy. 2017. The meanings of 'angels' in English, Arabic, and Hebrew. In Zhengdao Ye (ed.), *The semantics of nouns*, 89–119. Oxford: Oxford University Press.

al-Hilālī, Muḥammad Taqī-ud-Dīn & Muḥammad Muḥsin Khān. 1998. *The Noble Qurʾan: The English Translation of the Meanings and Commentary*. Madinah: King Fahd Complex.

Hughes, Aaron W. 2012. *Abrahamic Religions: On the Uses and Abuses of History*. Oxford: Oxford University Press.

Hughes, Aaron W. 2022. *An Anxious Inheritance: Religious Others and the Shaping of Sunnī Orthodoxy*. Oxford: Oxford University Press.

Izutsu, Toshihiko. 1964. *God and Man in the Qurʾan: Semantics of the Qurʾanic Weltanschauung*. Tokyo: The Keio Institute of Cultural and Linguistic Studies.

Izutsu, Toshihiko. 1966. *Ethico-Religious Concepts in the Qurʾān*. Montreal: McGill University Press.

Jeffery, Arthur. 1938. *The Foreign Vocabulary of the Qurʾān*. Baroda: Oriental Institute.

Jeffery, Arthur. 1950. The Qurʾān as scripture. *The Muslim World* 40. 41–55, 106–134, 185–206, 257–275.

Khalifa, Rashad. 1981. *Quran: The Final Scripture Authorized English Version*. Tucson, AZ: Islamic Productions.

Levisen, Carsten & Sophia Waters (eds.). 2017. *Cultural Keywords in Discourse*. Amsterdam: John Benjamins.

Lewis, Bernard. 2002. *What Went Wrong: Western Impact and Middle Eastern Response*. New York: Oxford University Press.

Lewis, Clive Staples. 1940. *The Problem of Pain*. Glasgow: Collins.

Marshall, David. 2014. *God, Muhammad and the Unbelievers: A Qurʾanic Study*. London: Routledge.

Mel'čuk, Igor. 2018. Anna Wierzbicka, Semantic Decomposition, and the Meaning-Text Approach. *Russian Journal of Linguistics* 22 (3). 521–538.

Munshey, Munir. 2016. *The Entire Noble Qur'an*. Available at answeringchristianity.com.

Ortega y Gasset, José. 1959. The difficulty of reading. *Diogenes* 7 (28). 1–17.

Palmer, Edward Henry. 1880. *The Qur'ân*. Oxford: Clarendon.

Pickthall, Marmaduke. 1976. *The Glorious Koran: A Bi-Lingual Edition with English Translation, Introduction and Notes*. London: Allen & Unwin.

Reynolds, Gabriel Said. 2018. *The Qurʾān and the Bible: Text and Commentary*. New Haven: Yale University Press.

Rodwell, John Medows. 1876. *The Koran*. London: B. Quaritch.

Sale, George. 1734. *The Koran, Commonly Called the Alcoran of Mohammed, Tr. Into English Immediately from the Original Arabic; with Explanatory Notes, Taken from the Most Approved Commentators. To Which is Prefixed a Preliminary Discourse*. London: J. Wilcox.

Shakir, M.H. 1985. *The Holy Quran: Arabic Text and English Translation*. Beirut: Ets. al-Wafaa.

Wansbrough, John. 1977. *Quranic Studies: Sources and Methods of Scriptural Interpretation*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configuraations*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1997. *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1999. *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge: CUP.

Wierzbicka, Anna. 2001. *What Did Jesus Mean? Explaining the Sermon on the Mount and the Parables in Simple and Universal Human Concepts*. Oxford: Oxford University Press.

Wierzbicka, Anna. 2014. *Imprisoned in English: The Hazards of English as a Default Language*. Oxford: Oxford University Press.

Wierzbicka, Anna. 2019. *What Christians Believe: The Story of God and People in Minimal English*. New York: Oxford University Press.

Wierzbicka, Anna. 2020. Addressing God in European languages: Different meanings, different cultural attitudes. *Russian Journal of Linguistics* 24 (2). 259–293.

Wierzbicka, Anna. 2021. "Semantic Primitives", fifty years later. *Russian Journal of Linguistics* 25 (2). 317–342.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. trans. G. E. M. Anscombe. 2nd edn. Oxford: Blackwell.

Dictionaries, Commentaries, Encyclopedias and Grammars

Adang, Camilla P. 2001. Belief and unbelief. In Jane Dammen McAuliffe (ed.), *Encyclopaedia of the Qurʾān*, 1. 218–226. Leiden: Brill.

Ambros, Arne A. & Stephan Procházka. 2004. *A Concise Dictionary of Koranic Arabic.* Wiesbaden: Reichert.

Biella, Joan Copeland. 1982. *Dictionary of Old South Arabic: Sabaean Dialect.* In Frank Moore Cross (ed.), Harvard Semitic Studies 25. Chico, CA: Scholars Press.

Böwering, Gerhard. 2002. God and his attributes. In Jane Dammen McAuliffe (ed.), *Encyclopaedia of the Qurʾān* 2, 316–331. Leiden: Brill.

Hawting, Gerald R. 2002. Idolatry and idolaters. In Jane Dammen McAuliffe (ed.), *Encyclopaedia of the Qurʾān* 2, 475–480. Leiden: Brill.

al-Maḥallī, Jalāl al-Dīn & Jalāl al-Dīn al-Suyūṭī. 2007. *Tafsīr al-Jalālayn.* In Ghazi bin Muhammad bin Talal (ed.). Trans. Feras Hamza. Amman, Jordan: Royal Aal al-Bayt Institute for Islamic Thought.

Mir, Mustansir. 2004. Polytheism and Atheism. In Jane Dammen McAuliffe (ed.), *Encyclopaedia of the Qurʾān* 4, 158–162. Leiden: Brill.

OED Online. http://www.oed.com/ Oxford: Oxford University Press.

Wright, William. 1896–1898. *A Grammar of the Arabic Language.* 3rd edn. 2 vols. Cambridge: CUP.

Bionote:
Mark DURIE is an Australian linguist, theologian and pastor. He is a Senior Research Fellow at the Arthur Jeffery Centre for the Study of Islam at Melbourne School of Theology, Founding Director of the Institute for Spiritual Awareness, and an Honorary Associate Professor at the Australian National University. His most recent book is *The Qur'an and Its Biblical Reflexes: Investigations into the Genesis of a Religion* (Lexington, 2008).
e-mail: mdurie@jefferycentre.mst.edu.au
https://orcid.org/0000-0003-2990-0937

Сведения об авторе:
Марк ДЬЮРИ – австралийский лингвист, богослов и пастор. Является Старшим научным сотрудником Центра им. Артура Джеффери по изучению ислама при Мельбурнской школе теологии, директором-основателем Института духовного самосознания при Австралийском национальном университете. Автор книги *The Qur'an and Its Biblical Reflexes: Investigations into the Genesis of a Religion* (Lexington, 2008).
e-mail: mdurie@jefferycentre.mst.edu.au
https://orcid.org/0000-0003-2990-0937

Research article / Научная статья

Emotions and attitudes in present day Russian through the prism of new words: Cultural semantics of *zhest'* and related concepts

Anna GLADKOVA ⓘ ✉

Australian National University, Canberra, Australia
Higher School of Economics, Nizhny Novgorod, Russia
✉angladkova@gmail.com

Abstract
Contemporary Russian lexicon is characterized by rapid change which involves borrowings, the use of new words and expressions as well as the development of new meanings from the existing word forms. The new meanings are indicative of new attitudes or the reinforcement of the existing ones. In this context, the paper considers the recently emerged colloquial use of the word *zhest'* (from the primary meaning of *zhest'* 'tin' as a type of metal) and the increase of use of the words *zhestkii* 'hard/tough/firm' and *zhestko* 'firmly/toughly' as examples of 'internal' language processes. The word *zhest'* is a colloquial word mainly used in youth slang, but also infiltrating other types of discourse. We analyze its use as an interjection, as well as a noun in predicative and attributive functions. *Zhestkii* and *zhestko* are shown to rise in use and to be prevalent in the political discourse as a sign of power. The paper aims to trace the rise in frequency of the words under the analysis, study their semantics and establish links between the meaning of the words and broader Russian cultural themes. This kind of linguistic analysis with focus on cultural aspects allows us to identify culturally prevalent ideas in present day Russian. The paper uses the Natural Semantic Metalanguage (NSM) as a method of semantic analysis relying on data from the Russian National Corpus. Based on a detailed semantic analysis, the paper proposes NSM explications of *zhest'*, *zhestkii* and *zhestko*, identifies connections between their meanings and the cultural themes of 'emotionality', 'not being in control' and 'strait talk', and recognizes the increased cultural salience of these words in present day Russian. The study uncovers trends of the contemporary Russian language uses and can be applied in culture-enhanced language teaching and cross-cultural training.
Key words: *Russian, language change, Natural Semantic Metalanguage (NSM), youth slang, political discourse, emotions*

Эмоции и отношения в современном русском языке через призму неологизмов: культурная семантика слова *жесть* и связанных с ним понятий

А.Н. ГЛАДКОВА ⒾⅮ ✉

Австралийский национальный университет, Канберра, Австралия
Высшая школа экономики, Нижний Новгород, Россия
✉angladkova@gmail.com

Аннотация
Современный русский лексикон подвержен стремительным изменениям, таким как заимствования, использование новых слов и выражений, а также появление новых значений у существующих словоформ. Новые значения являются проявлениями новых или усилением существующих культурных представлений. В статье рассматривается недавно возникшее разговорное употребление слова *жесть* (от первичного значения *жесть* как вида металла) и рост употребления слов *жесткий* и *жестко* как примеров «внутренних» языковых процессов. Слово *жесть* является разговорным словом и употребляется, в основном, в молодежном сленге, но также проникает в другие типы дискурса. В статье рассматриваются употребления слова *жесть* в качестве междометия, а также существительного в предикативных и атрибутивных функциях. *Жесткий* и *жестко* употребляются в политическом дискурсе как знаки власти. Цель статьи – проследить увеличение частотности анализируемых слов, изучить их семантику и установить связи между значением слов и культурными темами русского языка. Проведение данного лингвистического анализа с вниманием к культурным аспектам значения позволяет выделить культурно-значимые идеи в русском языке сегодня. Естественный семантический метаязык (ЕСМ) используется в качестве метода семантического анализа, языковые примеры взяты из Национального корпуса русского языка. На основе детального семантического анализа в статье предлагаются ЕСМ толкования слов *жесть*, *жесткий* и *жестко*, выявляется связь их значений с культурными темами «эмоциональность», «отсутствие контроля» и «откровенный разговор», и отмечается возросшая культурная значимость этих слов в современном русском языке. Исследование раскрывает тенденции использования современного русского языка, и его результаты могут быть применены в преподавании языка с углубленным изучением культуры и обучении межкультурной коммуникации.
Ключевые слова: *русский язык, изменение языка, Естественный семантический метаязык (ЕСМ), молодежный сленг, политический дискурс, эмоции*

1. Introduction

Contemporary Russian lexicon is characterized by rapid change which involves borrowings, the use of new words and expressions as well as the emergence of new meanings from the existing word forms (e.g., Gladkova 2020, Levontina 2015, 2016, Lerner & Zbenovich 2013, Krongauz 2008, 2013, Krysin 2014, Larina et al. 2020, Ozyumenko & Larina 2018). While the use of borrowings, mainly from English, is revealing of globalization processes and is consistent with the tendencies in numerous languages, the emergence of new meanings and words originating from Russian could be considered part of language- and culture-specific processes[1].

A recent example of 'internal' language processes is the development of the colloquial use of the word *zhest'* (from *zhest'* 'tin' – a type of metal) and the increase of use of the words *zhestkii* 'hard/tough/firm' and *zhestko* 'firmly/toughly'. There is no unanimity among scholars on the etymological origin of these words, but some sources suggest that all these words are etymologically related (cf. Fasmer 2004 online, Krylov 2005 online).

Zhest' as a colloquial word is loaded with emotional attitude and is usually used as a response token in situations when the Addressee learns about some 'tough' situation the Speaker has been in. The word has become prevalent in youth jargon, however, it has already left the domain of youth slang and has penetrated official or semi-official settings. Some examples can be found in Russian political discourse.

Along with the emergence of the new meaning of *zhest'*, we also observe the rise of use of the words *zhestkii* and *zhestko*. The words are used in colloquial speech as well as in official register and are becoming prevalent in political discourse.

In this paper, we attempt to trace the rise of frequency of use of the words *zhest'*, *zhestkii*, *zhestko*, analyze their semantics and establish cultural links between the meaning of the words and Russian cultural themes. We will use the Natural Semantic Metalanguage (Wierzbicka 1996, 2021, Goddard & Wierzbicka 2002, 2014, Gladkova & Larina 2018a,b) as a method of semantic analysis, also adhering to the principles of paraphrase, semantic transcription and decomposition using a semantic metalanguage (Mel'čuk 2012, 2018), while following the principles of Cultural Semantics (Gladkova 2010, Gladkova & Romero Trillo 2014, Levisen 2012) and Ethnopragmatics (Goddard 2006).

---

[1] One could anticipate that post-2022 military conflict in Ukraine, de-globalization or/and 'multipolarity' processes could prevail in Russia and have their impact on language. However, this would be the matter of future research.

2. Cultural element of meaning in the lexicon

The idea about the interconnection between language and thought on the one hand and language and culture on the other hand has prevailed in linguistics for over two centuries (e.g., Humboldt (1971[1836], 1988 [1836], 1997) Sapir (1949) Whorf (1956)). The approach known as Ethnopragmatics (Goddard (ed.) 2006) with its 'sister branches' Cultural Semantics (Gladkova 2010, Levisen 2012) and Applied Ethnolinguistics (Peeters 2015) propose a framework that attempts to underpin in detail the cultural element of linguistic meaning. Ethnopragmatics aims to articulate culture-internal perspectives on the 'how and why' of speech practices in the languages of the world using universal human concepts (Goddard & Wierzbicka 2014). It also describes and explains people's ways of speaking relating them to indigenous values, beliefs, attitudes, social categories, emotions, and so on (Goddard 2006).

Cultural Semantics and Ethnopragmatics postulate that certain words are culturally more salient than others (e.g., Gladkova 2010, Levisen 2013, Wierzbicka 1997). Wierzbicka (1997: 15–16) calls such words 'cultural keywords' and defines them as "words which are particularly important and revealing in a given culture." Cultural keywords are salient in the collective psyche of a society and their meanings resonate with meanings of other linguistic units and cultural practices. The cultural keywords are commonly characterized by relatively high frequency, relation to other culturally important concepts, and use in culturally prominent discourse (proverbs, songs, among others).

Goddard (2018: 165) further elaborates that the borderline between cultural keywords and other types of culturally salient vocabulary is not always easy to identify:

> A cultural key word is a kind of focal point for cultural ways of thinking, acting, feeling, and speaking. I must admit that there are many culturally important words in any language, and that sometimes it is hard to draw a strict line between cultural key words and other very important cultural words. I am not even sure that there is an absolutely strict line there. But the concept of "cultural key words" is still a useful concept, a way of directing attention to the fact that some words are tremendously important to a culture.

While the research on cultural keywords has successfully identified words of this kind in different cultures (e.g., Levisen 2012, Levisen & Waters 2017, Wierzbicka 1997), the question of the dynamics of meaning among cultural keywords and cultural themes has not been explored in depth before. This article attempts to study the meaning of the words *zhest'*, *zhestkii* and *zhestko* and identify their cultural role. Adhering to the principles of Ethnopragmatics and Cultural Semantics, we will explore the meaning of the words in question using the universal human concepts as they are identified within the Natural Semantic Metalanguage.

3. Data and methodology

The paper relies on data available in the Russian National Corpus – an open access online resource with over 1.5 billion words. For the purpose of our study, we will draw examples of use from three subcorpora: main (written sources), oral and newspaper. We will mainly rely on examples of use after the mid-1980s to record the linguistic change after the collapse of the Soviet Union which marked a significant economic and social turning point in the country's history. Therefore, the amount of analyzed material is around 915 mln. words.

To trace the dynamics of frequency of use we will rely on the data from Google Books as reflected in the Google Books Ngram Viewer. We acknowledge limitations of the Google Books data as they are restricted to books and, consequently, have limited representation of spoken data. However, we believe that these data are still useful in understanding the prevalent tendencies in the frequency of use.

The Natural Semantic Metalanguage is an approach to studying natural language meaning relying on universal human concepts. It originates from Leibniz's idea that linguistic meaning can be explained in terms of a set of irreducible concepts available in natural language. The empirical research of identifying primitive meanings by Anna Wierzbicka, Cliff Goddard and colleagues has resulted in a list of 65 semantic primitives and their syntactic properties. They are believed to be identifiable in all languages (Goddard & Wierzbicka 2014). Russian and English exponents of the primes are listed in Table 1.

Table 1. **Exponents of semantic primes in English and Russian** (from Gladkova and Larina (2018a))

YA, TY, KTO-TO, CHTO-TO~VESHCH', LYUDI, TELO	substantives
I, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	
ROD~VID, CHAST'	relational substantives
KINDS, PARTS	
ÉTOT, TOT ZHE, DRUGOI	determiners
THIS, THE SAME, OTHER~ELSE	
ODIN, DVA, NEKOTORYE, VSE, MNOGO, MALO	quantifiers
ONE, TWO, SOME, ALL, MUCH~MANY, LITTLE~FEW	
KHOROSHII~KHOROSHO, PLOKHOI~PLOKHO	evaluators
GOOD, BAD	
BOL'SHOI, MALEN'KII	descriptors
BIG, SMALL	
ZNAT', DUMAT', KHOTET', NE KHOTET', CHUVSTVOVAT', VIDET', SLYSHAT'	mental predicates
KNOW, THINK, WANT, DON'T WANT, FEEL, SEE, HEAR	
GOVORIT'~SKAZAT', SLOVA, PRAVDA	speech
SAY, WORDS, TRUE	
DELAT', PROISKHODIT'~SLUCHAT'SYA, DVIGAT'SYA	actions, events, movement
DO, HAPPEN, MOVE	
BYT' (GDE-TO), BYT'~EST', BYT' (KEM-TO/CHEM-TO)	location, existence, specification
BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING)	
MOI/MOYA/MOE	possession
(IS) MINE	

ZHIT', UMERET'	life and death
LIVE, DIE	
KOGDA~VREMYA, SEICHAS, DO, POSLE, DOLGO, KOROTKOE VREMYA, NEKOTOROE VREMYA, MOMENT	time
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	
GDE~MESTO, ZDES', NAD, POD, DALEKO, BLIZKO, STORONA, VNUTRI, KASAT'SYA	place
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH	
NE, MOZHET BYT', MOCH', POTOMU CHTO, ESLI	logical concepts
NOT, MAYBE, CAN, BECAUSE, IF	
OCHEN', BOL'SHE~ESHCHE	intensifier, augmentor
VERY, MORE	
KAK~TAK	similarity
LIKE~AS	

Apart from the 65 primitive meanings, the NSM toolkit includes semantic molecules, that is intermediate concepts consisting of semantic primitives and, possibly, other semantic molecules which are required for explications.

NSM as a tool of sematic analysis involving semantic paraphrase, semantic transcription and decomposition using a semantic metalanguage (Mel'čuk 2012, 2018) allows us to arrive at explications of meaning, which are precise, substitutable and comprehensible (Goddard 2018).

4. Results

4.1. Zhest' *as a colloquial word*

4.1.1. An overview of current uses

I will start with illustrating the use of the new meaning of *zhest'* with several examples from the Russian National Corpus:

> (1) – *Nu / na samom dele / glaza ochen' sil'no bolyat / potomu chto nu vot u menya byl den' – ya sela v desyat' utra / ponyatnoe delo s pereryvami / no vse ravno zakonchila v chetyre utra.*
> – *Zhest'!*
> '– Well, indeed, the eyes are really aching because what a day I had – I started at ten in the morning, with breaks of course, but finished at four in the morning anyway.
> – Zhest'! (Rough!)'

> (2) – *[…] i ya v pyat' utra prosypayus' prosto v kholodnom potu.*
> – *Zhest'!*
> – '[…] and I wake up in the morning all in cold sweat.
> – Zhest'! (Rough!)'

> (3) – *Nu / koroche u menya nogi potom tak otvalivalis' / èto zhopa.*
> – *Ty na kablukakh byla?*
> – *Konechno / na kablu … Tok na … Da / na kablukakh / na vot etikh / sapogakh.*
> – *Zhe-e-est'!*

'– Well, my feet were falling off after that …
– Were you wearing heels then?
– Of course, heels. Yes, heels, these high boots.
– Zhest'! (Rough!)'

As these examples demonstrate, the word is prevalent in youth jargon, however, it has already left the domain of youth slang and is being used by older people in official or semi-official settings at times. It has even infiltrated the Russian political discourse. One of such examples is a notorious slogan by Valentina Matvienko "*Sport – èto zhest'*" 'Sport is *zhest'*". According to Fontanka.ru, the then governor of St. Petersburg Valentina Matvienko used this slogan during her annual address to the deputies of the Legislative Assembly in 2007. When explaining why she was using a term from the youth slang she said that she had heard this word and got interested in its meaning. Upon learning that it means something like 'breakthrough, boiling and moving ahead' she decided that the word *zhest'* can be applied to sport.[2]

There is also an example of use of *zhest'* by Maria Zakharova, the Ministry of Foreign Affairs Spokesperson, as listed in the Russian National Corpus. According to *Parlamentskaya Gazeta*, Maria Zakharova commented in her Telegram-channel on a large-scale disruption in the work of Facebook, WhatsApp and Instagram in 2021 as follows:

(4) *Zhest'. K voprosu o tom, nuzhny li nam svoi sotsseti i internet-platformy.*
    'Zhest'. To the question whether we need our own social networks and internet platforms.'
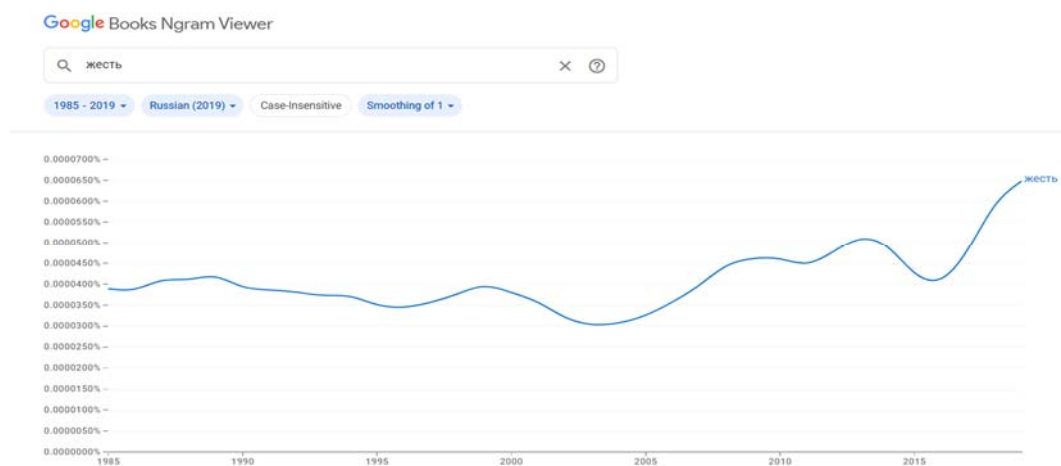
Figure 1. The frequency of occurrence of *zhest'* in Russian Google Books for 1985–2019

---

[2] The quote is from https://www.fontanka.ru/2007/05/23/067/. As our analysis further demonstrates, this use of *zhest'* by Matvienko is not consistent with the majority of uses because the word normally has negative connotations. However, this example dates to 2007 when the new meaning of *zhest'* was just emerging. Therefore, it is possible that Matvienko's use coincided with some initial uses of *zhest'* which did not prevail over time.

Google Books Ngram Viewer reports a more than two-fold increase in usage frequency of the word *zhest'* in the Russian language sources since 2004 (see Figure 1). While in this data it is impossible to differentiate between the use of *zhest'* in the original and the derived meanings, we hypothesize that this tendency is indicative of the development of the new meaning of *zhest'* and its increase in use.

Let us now delve into exploring the origins of the word and analyzing its meaning in greater depth.

4.1.2. Possible origins

In the original meaning *zhest'* 'tinplate' refers to sheets of steel coated with a thin layer of tin to impede rusting which are widely used in the manufacturing of tin cans. According to Fasmer's Etymological Dictionary (Fasmer 2004 online), *zhest'* as a name for a type of metal originates from Turkic or Mongolian *žes* meaning "copper, brass". This dictionary also relates the second outdated meaning referring to frozen soil to the words *zhestokii* 'cruel' and *zhestkii* 'hard/firm'. Krylov's Etymological Dictionary tentatively links *zhest'* with Turkic languages (*zhes*) and suggests that it changed its form under the influence of the Russian word *zhest'* 'frozen soil/hard soil' with the same derivational base as *zhestkii* (or *zhestyi*) 'hard/firm' (Krylov 2005 online). *Zhestkii*, in its turn, is listed here as a borrowing from Old Slavonic *zhestyi*, possibly related to German *kes* – hard soil. Krylov also links *zhestokii* 'cruel' to *zhestkii* and notes a close link between these notions. Therefore, there is likely to be an etymological link between *zhest'* as a type of metal which is hard with the word *zhestkii* (etymologically relating to hard or frozen soil) as well as the word *zhestokii* 'cruel'.

The Russian National Corpus oral subcorpus records the first uses of *zhest'* as a colloquial expression in 2006. Here is one of the examples dating 2006:

(5) *– Na her ėti shtuki krutyatsya?*
 – Nu / akkordy vybiraesh' / to est' / ty smotrish' / dopustim / fa… Ty vidish'
 / na kakom ladu ėto vse zazhimaetsya…
 – Zhest'! Kto ėto pridumal?
 – Ya dumayu / umnyi chelovek pridumal …
 '– Why are these things rolling?
 – Well, you select the chord, that is you see, say F… You see on which harmony it is gripped.
 – Zhest'! Who came up with it?
 – I think a clever person did!'

In this example *zhest'* is used to express astonishment and awe at the complexity of the object the Speaker encounters.

The emergence of colloquial use of the word *zhest'* could be associated with the release of the film with the same title by the Russian producer Denis Neimand in 2006. The film is a psychological thriller where a journalist Marina joins a police inspector in the search for a maniac teacher as part of her last task for the newspaper

she decides to quit. The action takes place on the vast territory of the abandoned summer house estate next to a large industrial city Azovstal' with no mobile connection and locals living according to their own laws.

The title of the film creatively interprets the use and the etymology of the original meaning of the word *zhest'* as a metaphor for human life. The film is not about 'tinplate' as sheets of steel, but rather human life which is, as tinplate, hard, flexible (changeable), and sharp at the ends. The title also highlights the link of *zhest'* with the words *zhestkii* 'hard' and *zhestokii* 'cruel' as the film contains numerous scenes of cruelty and violence. It is worth noting that several other Russian films released at about the same time – *Zhara* (2006) 'Heat' and *Zhmurki* (2005) 'Blind man's buff' – also used the new word in their scripts.

4.1.3. Meaning analysis

Krysin (2014: 550–551) describes *zhest'* as an evaluative jargon word expressing different emotions – delight, surprise, etc. in the role of an interjection. In our analysis we will divide the use of *zhest'* into two groups – one as an interjection (*zhest'₁*) and the other one as a noun in a predicative or attributive function (*zhest'₂*). We will demonstrate that the meaning of *zhest'* in both uses is not limited to emotions only and entails more complex attitudes.

4.1.3.1. *zhest'₁*

In the first use *zhest'* is used as a response token in a conversation. It is normally used after the Speaker talks about some unusual situation, often involving experiencing of hard conditions (from the point of view of the Speaker). The Addressee replying with *zhest'* demonstrates emphatic understanding of the toughness of the situation the Speaker had to endure and the resultant emotion. An element of surprise is also being conveyed.

We will quote some examples from the Corpus. In the following example a young person is telling about her experience of being exposed to the behavior of older people in a café or a restaurant which is different from her expectations and her friend replies with *zhest'*:

> *(6) – ... My sideli v "Kul'te" / i tam znaesh' / takoe / kak budto iz kakogo-to sosednego ofisa prishla kompaniya lyudei / i lyudi let po tridcat' pyat' / po sorok / takie teti-zhaby / i odna iz nix sidela bosikom / a kogda shla kuda-to / odevala takie plastikovye tapochki prosto. i onu chetyre raza stavili svoyu lyubimuyu pesnyu "Disko-partizany" / znaesh'?*
> *– Gospodi / zhest' kakaya-to!*
> *– My tak ugorali vashche / da.*
> – 'We were sitting in 'Cult' and there, you know, a company of people came from a nearby office, and they were thirty-five or forty, women-toads, you know. And one of them was sitting barefoot and she put on plastic slippers when she went somewhere. And they played their favorite song "Disco partisans" four times, you know?'

– Oh my God! What a zhest'!
– Yeah, we were laughing.'

The following example is about a young person complaining about the unexpected change of exam materials two weeks before the exam, which warrants the response *zhest'* from a friend:

(7) – *[…] za dve nedeli do ėkzamena pomenyali voprosy. Vmesto 57 voprosov / 70 / drugie voprosy / drugie proizvedeniya … Ne znayu / kak mne teper' sdavat' literaturu …*
    – *Zhest'!!! Razve tak mozhno?*
    – 'Two weeks before the exam they changed the questions. 70 questions instead of 57, different questions, different books … I don't know how I can pass Literature now…
    – Zhest'!!! How can one do it?'

Telling about enduring psychological pressure from another person also calls for the response *zhest'*:

(8) – *Da vse tak zhe … slava bogu / my s nei redko teper' vidimsya.*
    – *Ona prodolzhaet gruzit' ėmocional'nymi problemami?*
    – *Estessno / kuda tam bez nix.*
    – *Zhest'.*
    '– It's the same… Thanks God we meet rarely with her these days.
    – Is she continuing to load you with emotional problems?
    – Of course. How else?
    – Zhest'.'

In the following example *zhest'* is used two times to respond to a radical haircut. The person who had the haircut is also called *zhestkach* – another noun derived from *zhest'* which can be used to refer to tough situations or people.

(9) *S1 – Chego-chego / postrigli.*
    S2 – Zhest' / obkromsali nereal'no/
    S3 – Da ty posmotri na ego chelku / ussat'sya mozhno.
    S2 – Kha / zhest'! Ty tipa pod Natal'yu Oreiro kosit' nachal / zhestkach!
    S1 – Ty voobshche by pomolchal […].
    S1 – 'What? I had a haircut.
    S2 – Zhest'! They cut your hair radically.
    S3 – Look at his fringe! You can piss yourself!
    S2 – Ha, zhest'! Do you want to look like Natalia Oreiro? You are a tough guy!
    S1 – Shut up! […].'

Having to queue for a long time also invites the response *zhest'*:

(10) – *[…] Che / gulyali potom eshche?*
    – *Ne / v ocheredi do polvdenadcatogo pochti stoyali.*
    – *Zhest' kakaya. Skol'ko vas bylo?*
    – *Ya / Bob i Diman. No vzyali na semerykh bilety.*
    '– […] Did you walk after that?

– No, we were standing in the queue till 11.30 almost.
– What a zhest'. How many were you?
– Myself, Bib and Diman. But we bought tickets for seven people.'

As we can observe in the above quoted examples, the situations when *zhest'* is used are the following:
- having to work for a day without rest,
- having to run in boots on high heels,
- having to share space with people of different age group, habits and interests,
- having to interact with a person who has psychological problems,
- changing exam questions two weeks before the exam,
- having a radical haircut.

Most examples are limited to youth conversations and therefore embody the views of this age group. Experiencing discomfort, pressure, hard work, unfair treatment are regarded as disturbing everyday experience and trigger the response *zhest'*.

We will propose the following explication:

Zhest'₁!
(a)   I now know: something happened
(b)   I think: things like this don't happen at many times
(c)   when this happened, you felt something very bad
(d)   I know how you felt
(e)   I don't want you to feel like this
(f)   when I think about it now, I can't not feel something
(g)   I want you to know this
(h)   because if this, I say this word {*zhest*}

This explication follows the patterns for explicating interjections as presented in (Wierzbicka 1992, Goddard 2014, 2015, Gladkova et al. 2016) and contains the elements of (a) awareness of the situation, (b) – expression of the unusual character of the situation, (c) – demonstration of the awareness of the consequent negative feeling of the interlocutor, (d) – emphatic expression of the awareness of the feeling of the other person, (e) – expression of support or pity to the interlocutor in relation to this situation, (f) – an emotional response to the situation, (g) – summary of the communicative effort, and (h) – the word utterance component.

4.1.3.2. *zhest'₂*

Zhest'₂ is used in the predicative or attributive function as a colloquial word. However, unlike *zhest'₁*, it does not necessarily function as a response token, but it is used by the Speaker to express surprise or astonishment about some 'tough' conditions.

As *zhest'₂* is also prevalent in the youth slang or conversational discourse, it is common in discussions relating to studies. In example (11) the teacher is described

as *zhest'* as she was very strict and in (12) – the homework (voluminous and challenging).

(11) – *A u vas strogaya uchitel'nica?*
– *Nea … dobraya.*
– *A u nas voobshche Vera Vanna vo 2 shkole prosto zhest' byla zhenshchina. Znaesh' / na fizike dazhe spisat' nel'zya. Znaesh' / esli povernesh'sya zatylkom vo vremya kontrol'noy / vse / ona libo podoidet i postavit minus ball / potom minus dva / a potom i vovse mozhet zabrat' listochek. Voobshche / spisat' slozhno bylo.*
'– Do you have a strict teacher?
– No … a kind one.
– And our Vera Vanna in school number 2 was simply *zhest'* woman. You know, it was impossible to cheat during Physics class. You know, if you turn your head during a test she would come over and deduct one point, then two points, and then she could even take away the paper. It was difficult to cheat.'

(12) *Odno mogu skazat' / to / chto ona zadaet – zhest'! Osobenno to / chto nam kakie-to referaty nado gotovit'. Uzhas!*
'I can say one thing – the homework she gives is *zhest'*! Especially the term papers we need to prepare. Crazy!'

Very cold conditions outside are described as *zhest'* in the following two examples:

(13) *Slushai / smotri / blin / tam takoi sneg / takaya zhest' prosto / belki ne mogut … Poshli v snezhki igrat'? Belki ne mogut zhit' v takom kholode.*
'Look at the snow outside. It is simply *zhest'*. Squirrels can't … Let's go and play snowballs? Squirrels can't live in such cold.'

(14) – *Blin … da gde zh avtobus-to / a?*
– *Da heze! A che / speshish'?*
– *Da dubak – vashche zhest'! [...]*
– *Ta zhe fignya!*
'– [Swearing]. Where is the bus?
– I don't know. Are you in a hurry?
– The cold is *zhest'*. […]
– Same crap here.'

The conditions or situations that can be characterized as *zhest'* are numerous and diverse and include, among others, the negative effect of divorce on children in the 1970s (example 15), violence in detention centers (example 16), unhuman conditions in public transport (example 17), tourist's being cheated by locals (example 18), lack of medical help (example 19). In such examples we can observe the resonance between the new meaning of the word *zhest'* and a similar-sounding word *zhestoko* 'cruel'. One could tentatively argue that there is an element of cruelty in the following situations:

(15) *Vy pomnite vot razvody / naprimer / semidesyatykh? Nu / ėto / ėto byla krovavaya zhest'. Kakie deti? Kogo interesovali chuvstva detei?*
'Do you remember, for example, divorces of the seventies? Well, it was bloody *zhest'*. What children? Who was interested in children's feelings?'

(16) *Ta, kak seichas govoryat, zhest', kotoraya tvoritsya v tsentrakh zaderzhaniya, – v novinku dazhe dlya tekh, kto uzhe sidel ran'she. [...] Situatsiya pomenyalas' v khudshuyu storonu, i estestvenno, chto ėto politicheskij zakaz, a ne prosto sledstvie ėkonomicheskogo krizisa. Seichas centry izolyacii pravonarushitelei v Minske prevratilis' v takie pytochnye, kak gestapo vo vremya voiny.*
'The zhest', as they call it now, that takes place in detention centers is a novelty even for those who have had sentences in the past. […] The situation has changed for the worse and, naturally, it is a political order and not simply consequences of the economic crisis. Now isolation centers in Minsk turned into torture rooms like gestapo in war times.'

(17) *Ranee ėta zhe turistka opisala otpusk v Egipte slovami "dikii stress" i "ėto kakaya-to zhest'" iz-za togo, chto mestnye zhiteli vse vremya pytalis' ee obmanut'.*
'Earlier the same tourist described her holiday in Turkey with the words 'utter stress' and 'it is some *zhest'*' because locals were always trying to cheat her.'

(18) *Ikh vid privel ee v uzhas. Po slovam Manek, ėlektrichka bol'she napominaet transport dlya perevozki skota. Na snimkakh vidno, chto vagony pokryty rzhavchinoi, kraska oblupilas', a okna potemneli ot gryazi. "Rebyata, nu ėto real'naya zhest'! Nu kak na vot takom mozhno vozit' lyudei?"*
'Their [trains] look made her terrified. According to Manek's words, the train looks more like transport for animals. It is clear from the shots that the wagons are covered in rust, the paint is peeling off and the windows are dark from the dirt. 'Guys, this is real *zhest'*! How can one transport people in it?'

(19) *U nas v gorode s naseleniem 150 tysyach chelovek vsego odna detskaya poliklinika. Ėto zhest'.*
'In our town with the population of 150 thousand people there is one children's policlinic. It is *zhest'*'.

We will attempt to generalize that calling something *zhest'* involves recognizing that the event or the situation is not a typical experience, but a negative one. It causes negative emotions of the person who has to go through it. By using the word *zhest'* the Speaker marks his or her negative attitude to the situation and the desire for it not to happen. The explication looks as follows:

Something X is zhest'₂
(a) this something is like this:
(b)   something happens, it doesn't happen at many times
(c)   when other people think about it, they can know that this is something very bad

(d)   they can feel something very bad because of this
(e)   they don't want to feel like this
(f)   they don't want things like this to happen

To summarize, *zhest'₂* is an emotional and attitudinal term expressing attitude relating to the previous proposition. It underlies the unusual character of the event (component b) and it is negatively evaluated (component c). The event causes a negative feeling (d). The explication also includes a natural want of people not to feel like this (e) and for things like this not to happen (component f).

We will now turn to the analysis of the words *zhestkii* and *zhestko* as conceptually related to the colloquial use of the word *zhest'*.

4.2. Zhestkii and zhestko

Along with the emergence of the new meaning of *zhest'*, we also observe the rise of use of the words *zhestkii* and *zhestko*. The words are used in colloquial speech as well as in official register and are becoming prevalent in political discourse. The data from the Google Ngram Viewer suggest that the words *zhestkii* and *zhestko* have experienced a steady rise in use after 1985 with a temporary decline in 2008–2014 (see Figure 2).
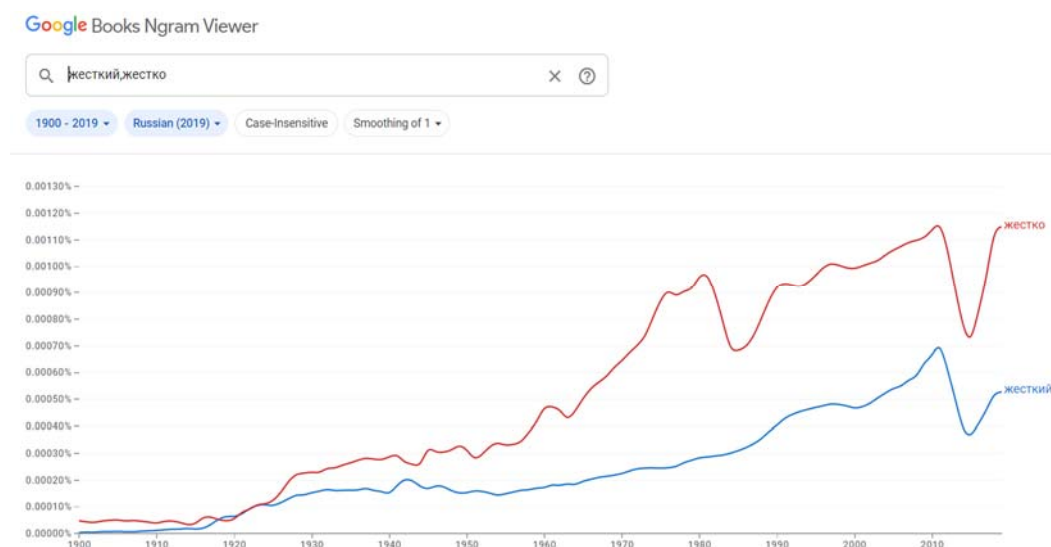
Figure 2. The frequency of occurrence of *zhestkii* and *zhestko* in Google Books for 1900–2019

4.2.1. Zhestkii

4.2.1.1. Zhestkii₁

In the context of our study, we are most interested in the metaphorical use of *zhestkii* 'hard/firm', that is *zhestkii₂*. However, its meaning cannot be understood without the analysis of its original meaning as a word describing physical property (*zhestkii₁*).

Zhestkii₁ is an adjective referring to a physical property of objects which display resistance to touch. The most prototypical use of the word in its primary sense could be considered the characteristics of objects that a person could lie or sit on: *krovat'* 'bed', *lozhe* 'bed' (lit. 'place for resting in a lying position'), *koika* 'bunk', *pastel'* bed', *meditsinskaya kushetka* 'medical couch', *matras* 'matrass', *divan* 'sofa', *kreslo* 'armchair', *stul* 'chair'. As a way of extension, the word is also applicable to supporting constructions that have a firm structure and that prevent things they are intended for carrying or supporting from moving, such as *karkas* 'frame', *tara* 'container', *konteiner* 'container', *kabina* 'cabin'. Another group of words that *zhestkii* is applicable to is human hair – *volosy* 'hair', *resnicy* 'eye lashes', *kudri* 'curls', *brovi* 'eyebrows', *usy* 'moustache', *shchetina* 'bristle', *boroda* 'beard', as well as parts of body, such as *pal'cy* 'fingers' and *ladon'* 'palm'. *Zhestkii* could also refer to some types of food, such as a cooked piece of meat or an apple, as well as fabric and shoes – *kosynka* 'head scarf', *kanva* 'canvas', *obuv'* 'shoes'.

To generalize, *zhestkii* refers to a property of objects that could be identified if a person comes in contact with the object. Goddard and Wierzbicka (2014) in their analysis of *hard* argue that the property of *hard* is identified by a person touching an object by hand. Extending on their approach and taking into account the prototypicality of lying or resting on things that are described as *zhestkii*, we suggest that this quality is identified by parts of human body. A hand is one of such parts, but it could also be the back or the bottom (in case of sitting or lying). We would propose the following explication:

Something is zhestkii₁ (krovat' 'bed', stul 'chair')
(a)   this thing is like this:
(b)   if someone's part of the body touches this thing
(c)   this someone can feel something in this part of the body because of this
(d)   because of this, this someone can know something about this thing
(e)   because of this, this someone can think like this:
(f)   if someone's part of the body touches this thing,
(g)   this part of the body can't move as this someone's wants
(h)   because of this, this someone can feel something bad in this part of the body

Some of the examples suggest a lengthy lasting effect of contact with a *zhestkii* object if a person sits or lies on it for some time:

(20) *Takaya krovat' zhestkaya, vse telo bolit.*
     'The bed is so hard, the whole body is aching.'

However, given that this effect can happen due to coming in touch with objects by sitting or lying on them and might not apply to things one could get in touch by hand (e.g., hair), we will not propose the component 'for some time' as an invariant of meaning.

4.2.1.2. *Zhestkii₂*

As Goddard and Wierzbicka (2014) rightly note, the metaphoric use of words of physical properties could be more frequent than the use of the words in their original meaning. It is the case with *zhestkii₂* which builds on the meaning of *zhestkii₁* and applies to a wide range of situations where a person has some sort of restrictions in his/her actions. Such uses include: *ramki* 'framework', *grafik* 'schedule', *kontrol'* 'control', *pravila* 'rules', *konkurentsiya* 'competition', *ogranicheniya* 'restrictions', *usloviya* 'conditions', *kriterii* 'criteria', *trebovaniya* 'requirements', *dieta* 'diet', *davlenie* 'pressure', *poryadok* 'order', *reshenie* 'decision', etc. Some examples from the corpus:

(21) *[...] bez zhestkoi konkurencii vryad li mestnye proizvoditeli stremilis' by k tomu vysokomu urovnyu, kotorogo dostigli segodnya.*
'It is unlikely that the local producers would have aimed at this high level without tough competition.'

(22) *U torgovykh setei sushchestvuyut zhestkie grafiki postavki produktov, i im ne nuzhny zapasy, poskol'ku ikh prosto negde skladirovat'.*
'Retail chains have strict delivery schedules, so they do not need reserves because they have nowhere to store goods'.

(23) *On principial'no otkazyvalsya rabotat' v zhestkikh ramkakh sotsrealizma, vypolnyat' sotsyal'nyi zakaz.*
'He categorically refused to work within the rigid framework of social realism and deliver social order.'

We propose the following explication of *zhestkii₂*:

Something X is zhestkii₂
(a)   this thing is like this:
(b)   when this thing happens,
(c)   it is like someone touching something *zhestkii₁*[m]
(d)   because of this, this someone can think:
(e)   I can't do many things as I want
(f)   because of this, this someone can feel something bad

This explication embeds the meaning of *zhestkii₁* as a derivational molecule. At the same time, it reinforces and builds on the components present in *zhestkii₁* – the prototypical cognitive scenario involving a person realizing that s/he is restricted by something (I can't do many things as I want, although the 'move' component is absent) and a consequent negative feeling.

We hypothesize that this meaning is growing in use by ordinary speakers as a reflection of the challenging reality. At the same time, it is commonly used by officials to 'justify' their 'tough' actions in the challenging conditions.

The metaphoric use of *zhestkii* is common in the current Russian President's discourse as reported in mass media (newspaper) subcorpus. An example from the New Year 2022 President's address to the nation:

(24) *Vsekh nas seichas ob"edinyaet nadezhda na dobrye peremeny, no my ponimaem, chto ikh nevozmozhno otorvat', otdelit' ot sobytii ukhodyashchego goda. My stolknulis' s kolossal'nymi vyzovami, no nauchilis' zhit' v takikh zhestkikh usloviyakh, reshat' slozhnye zadachi, i smogli ėto sdelat' blagodarya nashei solidarnosti.*
'We are all united by the hope for good changes, but we understand that they cannot be separated from the events of the previous year. We have faced colossal challenges, but learned to live in such tough conditions, solve complex problems and we have been able to do it due to our solidarity.'

Another example is from the Address to the Federal Assembly regarding accepting the law on financial responsibility of producers for their ecological harm:

(25) *Takoi podkhod ochen' prostoi. Kak on zvuchit: poluchil pribyl' za schet prirody – uberi za soboi. Zdes' nuzhno deistvovat' zhestko. Rosprirodnadzor, drugie kontroliruyushchie organy dolzhny vypolnyat' vozlozhennye na nikh obyazannosti.*
'This approach is very simple. It sounds like this: if you have gained profit at the nature's cost, clean after yourself. One needs to act firmly here. Rosprirodnadzor and other controlling organizations should perform their duties.'

Some other examples as reported by other politicians:

(26) *Peskov zayavil, chto Putin zhestko otreagiroval na slova Sokurova.*
'Peskov stated that Putin reacted firmly to Sokurov's words.'

(27) *[…] prezident predupredil o zhestkom otvete na kakie-libo provokacii.*
'The President warned about a firm response to any kind of provocations.'

(28) *"Ėto zhestkoe poruchenie prezidenta RF", – napomnila spiker […].*
''It is a firm task by the Russian President,' the Speaker reminded.'

(29) *"Prezident osobo podcherknul, chto stoimost' etikh uslug ne dolzhna byt' zavyshennoi, nuzhno ee zhestko kontrolirovat'", – otmechal ranee prem'er-ministr Mikhail Mishustin.*
''President especially emphasized that the cost of these services should not be inflated, it needs to be controlled firmly', Prime Minister Mikhail Mishustin noted earlier.'

(30) *"[…] Potomu chto poruchenie prezidenta bylo konkretnoe, zhestkoe, pravil'noe, i nado, chtoby ljudi ėto pochuvstvovali", – otmetila predsedatel' Soveta Federacii.*
''Because the President's task was concrete, firm and correct, and it is important that people feel it', – Chairperson of the Federation Council noted.'

It is also common in the speech of other government officials:

(31) *Gref schitaet, chto vlasti strany budut priderzhivat'sya "ochen' zhestkoi makroekonomicheskoi stabil'nosti" i narashchivat' usiliya dlya smyagcheniya investitsionnogo klimata.*
'Gref believes that the country's government will follow "a very strict macroeconomic stability" and increase efforts for softening the investment climate.

(32) *Popova otmetila, chto ne vezde soblyudaetsya poruchenie, v sootvetstvii s kotorym rezul'taty PCR-testa na COVID-19 dolzhny predostavlyat'sya klientu v techenie sutok, i rasporyadilas' zhestko reagirovat' na zaderzhki.*
'Popova noted that not everywhere the instruction to provide the client with the COVID-19 pcr-test result within 24 hours is followed and ordered to react firmly to the delays.'

(33) *Èl'vira Nabiullina skazala, chto Centrobank budet provodit' zhestkuyu denezhno-kreditnuyu politiku.*
'Elvira Nabiullina said that Central Bank would follow a strict monetary and credit policy.'

(34) *Po mneniyu Matvienko, neobkhodimo "zhestche trebovat' s sobstvennikov investirovat' v soderzhanie teplovykh setei, chtoby snizhat' poteri tepla pri teplosnabzhenii".*
'In Matvienko's opinion, it is necessary to "firmly demand owners to invest into maintenance of heating network to reduce the losses of heat in heat supply".'

4.2.2. Zhestko

Zhestko is an adverb derived from the adjective *zhestkii.* Its use in the 'original' physical properties meaning (*zhestko₁*) is even more limited than the use of the adjective. We will quote only a couple of examples from the corpus:

(35) *V zatylok zhestko upersya stvol avtomata.*
'The barrel of the machine gun rested hard on the back of his head.'

(36) *Ne sleduet razbirat' korpusnuyu mebel' [...], esli ee chasti zhestko soedeneny (s pomoshch'yu kleya).*
'Do not disassemble cabinet furniture if its parts are rigidly connected (with glue).'

The extended use of *zhestko* (*zhestko₂*) prevails over the use of its original meaning. The predominant group of words collocating with *zhestko* is speech act verbs – *skazat'* 'say', *govorit'* 'say/speak', *konstatirovat'* 'state', *otzyvat'sya* 'say/evaluate', *zayavit'* 'declare', *sprosit'* 'ask', *proiznesti* 'say/announce', *otchitat'* 'reprimand', *velet'* 'order', *otvetit'* 'respond'. It also actively collocates with verbs of action which control or restrict actions of others, such as

kontrolirovat' 'control', *ogranichivat'* 'restrict', *regulirovat'* 'regulate', *prinyat' reshenie* 'decide', *konkurirovat'* 'compete'. Some examples from the corpus:

(37) *Nas zhenskie podrobnosti ni s kakogo boka ne interesuyut, tovarishch prokuror, – zhestko obrezal Nejman.*
'We are not interested in women's details from any side, Comrade prosecutor, – Neiman cut off harshly.'

(38) *Posemu vsyakogo roda nauchnaya deyatel'nost' po issledovaniyu chelovecheskogo mozga budet vsegda zhestko kontrolirovat'sya.*
'Therefore, any kind of research activity on human brain will always be strictly controlled by the state.'

The explication would be the following:

Someone X did something Y zhestko
(a) someone X did something Y to someone else Z
(b) when X did it, someone else Z could think about it like this:
(c) it is like touching something *zhestkij₁*[m]
(d) because of this, this someone could think:
(e) I can't do many things as I want
(f) because of this, this someone can feel something bad

The explication refers to the explication of *zhestkii₁* and repeats some of the components of *zhestkii₂* explication.

5. Discussion

As Levisen and Waters (2017: 6) argue in their introduction to the book *Cultural Keywords in Discourse*,

> One of the truisms of traditional pragmatics was that meaning sits in contexts and intentions, and not, as such, in words. But words, and especially cultural keywords, have context-governing potential. Of course contexts can influence meaning, but the opposite can also be true – words can create contexts. […] Once invoked, words can activate and guide people's interpretations and direct their conversations and discourses.

These words can be applicable to the current use of the Russian words *zhest'*, *zhestkii* and *zhestko* which are examples of how words are both influenced by and at the same influence the reality.

The emergence and the rise of the new meaning of the word *zhest'* in the colloquial sense reflects attitude to the reality full of rapid change, unpredictability, lawlessness and, at times, cruelty. Once emerged, the meaning got stabilized and is now experiencing rise in use and gradual spread in other conversational domains. Its uses started creating a reality on its own where events and people get characterized as *zhest'*.

It is intriguing to realize that the existing meanings *zhestkii* and *zhestko* undergo a similar rise in use which underlies close conceptual links between these

three words. Their current spread in the political discourse is also revealing. On the one hand, it reflects the reality of 'tough' conditions and, on the other hand, it implicitly serves as a justification for political actions that 'create' such reality (cf. Ozyumenko & Larina 2021).

Conducting a detailed semantic analysis of the words in question and developing their explications in universal human concepts following the principles of semantic decomposition, allow us now to hypothesize about the relation of their semantics to several cultural themes. Firstly, they are linked with the theme of emotionality (Wierzbicka 1999, Zalizniak et al. 2005) and the emerging emotionalization of discourse (Lerner & Rivkin-Fish 2021, Alba-Juez & Larina 2018, Zappettini et al. 2021). As Lerner and Rivkin-Fish (2021: 5) put it, many contemporary theorists stress "the pervasive presence of emotionality in contemporary culture where emotions become more important and formative than anything else". We can observe the trend of 'emotionality' in the meanings of the words under question, especially the new meaning of *zhest'*. *Zhest'* is a word involving an emotional response to another person, who is also recognized to be experiencing strong emotions. Therefore, the meaning of this word is highly consistent with the tendency to focus on one's emotions, analyze them, discuss them, and relate to them.

Secondly, the use of the words reinforces the idea of 'not being in control' which was previously proposed by Wierzbicka as a Russian cultural theme (Wierzbicka 1992). The words *zhestkii* and *zhestko* have the elements of meaning 'not being able to do something as one wants' and it is intriguing to see them becoming prominent in political discourse.

Thirdly, the use of the words *zhestkii* and *zhestko* could be linked with the cultural theme of 'direct and forceful talk' and expression of opinion. Multiple scholars note this tendency of Russian discourse and contrast it with Anglo-Saxon speech practices of more 'subtle' and 'indirect' expression (e.g., Gladkova 2015, Larina 2005, Wierzbicka 2012, Prohorov & Sternin 2002). We will quote Yale Richmond, a former US Foreign Service Officer who spent twenty years in Russia, and characterizes the manner of speaking which he finds specifically Russian as follows:

> Straight talk is appreciated, even when it leads to disagreement. When disagreement does occur, Russians appreciate honesty rather than attempts to paper over differences. It is far better to level with them and be certain that they fully understand your position. They respect adversaries who are straightforward and sincere in expressing views that diverge from their own. (Richmond 2003: 143)

The use of the word *zhestkii* and *zhestko* with speech acts reveals similar tendencies in Russian-specific ways of talking.

Finally, the emergence of the new meaning of the word *zhest'* could be considered in the light of metaphorical use of words referring to metal. Other names of metal and their derivatives that are used metaphorically in Russian are *zheleznyi*

'iron' (as well as *zhelezno* 'for sure/firmly', also its variant *zhelezobetonnyi* 'reinforced concrete'), *stal'noi* 'steel', also *zolotoi* 'gold', *serebryanyi* 'silver', *titan* 'titan'. It is particularly interesting to consider the extension of the meaning *zhest'* against *zheleznyi* 'iron' and *stal'noi* 'steel', all of them being hard metals or alloys.

Both *zheleznyi* and *stal'noi* are productive in their metaphorical use. *Zheleznyi* collocates with the words *ruka* 'hand', *distsiplina* 'discipline', *kontrol'* 'control', *zanaves* 'curtain', *volya* 'will', *zhenshchina* 'woman', *tverdost'* 'firmness', *rukopozhatie* 'handshake', *paren'* 'guy'. *Zhelezno* (adverb) has only metaphorical use as a colloquial word to emphasize the definite character of something:

(39) *V obshchem ėto chas obeda / ikh ne otryvat' / ėto voobshche zhelezno.*
    'It's lunch time and they shouldn't be distracted; it is like iron.

(40) *– Kos! Tebe ekhat'!*
    – Ya odin ne poedu / zhelezno!
    '– Kos, you are to go!
    – I will not go alone, that's for sure'.

(41) *– Ty chto / ne predupredil?*
    – Da predupredil / zhelezno!
    '– What? You haven't warned them?
    – I have! Dead sure!'
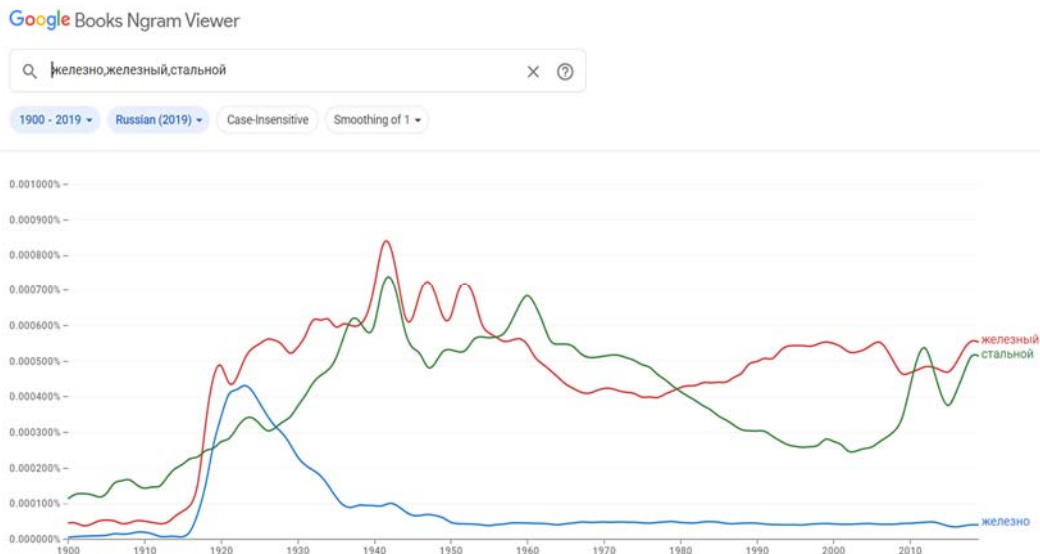
Figure 3. The frequency of use of *zheleznyj*, *zhelezno*, *stal'noj* in Google Ngram View in 1900–2019

The graph in Figure 3 indicates a significant rise in use of the words *zheleznyi* and *stal'noi* from the pre-revolution time with the steady use till around the 1960s. The rise is consistent with the industrial production of the materials. After the 1960s the variation in the frequency of use is not that significant. At the background of frequency of the words *zheleznyi*, *stal'noi*, the rise of frequency of *zhest'* in the recent years is more dramatic (see Fig. 1). This rise could coincide with the rise in

the demand and the production when tin packaging for food and beverages became widespread. At the same time, it is interesting to observe the change in the type of material and, consequently, qualities that are associated with the different types of metals. Iron and steel are very strong metals which are very endurable and hard to bend. Consequently, the metaphorical use of the words emphasizes stability, firmness, and ability to stick to one's principles and words. Tin, on the other hand, has different qualities and use from iron and steel. It is characterized by flexibility and sharpness. It is the effect of emotional and psychological pressure and the ability to withstand it that is reflected in the meaning of *zhest'* as a colloquial word.

To sum up, the emergence of the new meaning of *zhest'* suggests, on the one hand, a certain lacuna in the Russian lexicon which has been filled. On the other hand, it also indicates the sufficient accumulation of the experience embedded in the word *zhest'* (unfair, sometimes, cruel treatment causing a strong emotional reaction and a desire to oppose it) that 'requires' being 'recorded' in language. The new meaning of *zhest'* is semantically and conceptually linked to the existing notions *zhestkii* and *zhestko* 'firm/hard', and it is also linked to the notion *zhestokii* and *zhestoko* 'cruel'. The sound similarity and symbolism might be playing a role in strengthening this connection. At the same time, we also observe a rise in frequency of use of the words in question. This overall situation suggests that the conceptual field relating to '*zhestkost'*' 'firmness' seems to be reinforced in contemporary Russian. It is not a new semantic field, but in our days, we observe the rise of its cultural salience.

6. Conclusion

The word *zhest'* has been used in a new meaning for more than a decade in Russian. While being mainly limited to youth jargon, it is already leaving this domain and infiltrating the official or semi-official speech. A detailed analysis based on data available in the Russian National Corpus using the Natural Semantic Metalanguage demonstrates that two major uses of the word can be distinguished – as an interjection and a noun. In the first use, the meaning has the components of realisation of something very bad happening to the interlocutor, emphatic understanding of the emotional condition of the interlocutor and the communication to the interlocutor of this realisation and a consequent feeling. In the second use, the word has the components of realisation of something very bad happening, the unusual character of such events, the desire not to feel like this and for things like this not to happen.

A detailed semantic analysis also identifies semantic components of the words *zhestkii* and *zhestko* and demonstrates their conceptual relation to the new meaning of *zhest'*. The increase of frequency of use of the words can be hypothesized to correlate with the increase of the cultural significance of the words.

Acknowledgements

REFERENCES

Alba-Juez, Laura & Tatiana Larina. 2018. Language and emotion: Discourse-pragmatic perspectives. *Russian Journal of Linguistics* 22 (1). 9–37. https://doi.org/10.22363/2312-9182-2018-22-1-9-37

Fasmer, Maks. 2004. *Etimologicheskii slovar' russkogo yazyka* [Etymological Dictionary of Russian]. 4 vols. https://gufo.me/dict/vasmer (accessed 20 November 2022).

Gladkova, Anna. 2010. *Russkaya kul'turnaya semantika: emocii, cennosti, zhiznennye ustanovki* [Russian cultural semantics: emotions, values and attitudes.] Moscow: Languages of Slavonic Cultures.

Gladkova, Anna. 2017. Communication Modes, Russian. In Young Yun Kim (ed.)*, The International Encyclopedia of Intercultural Communication*. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781118783665.ieicc0147

Gladkova, Anna. 2020. When value words cross cultural borders: English *tolerant* vs. Russian *tolerantnyi*. In Bert Peeters & Kerry Mullan (eds.), *Studies in Ethnopragmatics, Cultural Semantics and Intercultural Communication. Minimal English* (and Beyond), 73–94. Singapore: Springer.

Gladkova, Anna & Tatiana Larina. 2018a. Anna Wierzbicka, words and the world. *Russian Journal of Linguistics* 22 (3). 499–520. http://dx.doi.org/10.22363/2312-9182-2018-22-3-499-520

Gladkova, Anna & Tatiana Larina. 2018b. Anna Wierzbicka, culture and text. *Russian Journal of Linguistics* 22 (4). 717–748. http://dx.doi.org/10.22363/2312-9182-2018-22-4-717-748

Gladkova, Anna**,** Ulla Vanhatalo & Cliff Goddard. 2016. The semantics of interjections: An experimental study with Natural Semantic Metalanguage. *Applied Psycholinguistics* 37 (4). 841–865. http://dx.doi.org/10.1017/S0142716415000260

Goddard, Cliff. 2006. Ethnopragmatics: A new paradigm. In Cliff Goddard (ed.), *Ethnopragmatics: Understanding discourse in cultural context,* 1–30. Berlin: Mouton de Gruyter.

Goddard, Cliff (ed.) 2006. *Ethnopragmatics: Understanding Discourse in Cultural Context*. Berlin: Mouton de Gruyter.

Goddard, Cliff. 2014. Interjections and emotion (with special reference to "surprise" and "disgust"). *Emotion Review* 6. 53–63.

Goddard, Cliff. 2015. "Swear words" and "curse words" in Australian (and American) English. At the crossroads of pragmatics, semantics and sociolinguistics. *Intercultural Pragmatics* 12 (2). 189–218.

Goddard, Cliff. 2018. *Ten Lectures on Natural Semantic MetaLanguage: Exploring Language, Thought and Culture Using Simple, Translatable Words.* Leiden: Brill.

Goddard, Cliff & Anna Wierzbicka (eds.). 2002. *Meaning and Universal Grammar: Theory and Empirical Findings. 2 Vols*. Amsterdam: John Benjamins.

Goddard, Cliff & Anna Wierzbicka. 2014. *Words and Meanings: Lexical Semantics across Domains, Languages and Cultures.* Oxford: Oxford University Press.

Google Books Ngram Viewer https://books.google.com/ngrams/ (accessed 15 September 2022).

Humboldt, Wilhelm von. 1971[1836]. *Linguistic Variability and Intellectual Development*. Coral Gables, Florida: University of Miami Press.

Humboldt, Wilhelm von. 1988[1836]. *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*. Cambridge: Cambridge University Press.

Humboldt, Wilhelm von. 1997. *Essays on Language*. In Theo Harden & Daniel Farrelly (eds.). Frankfurt am Main: Peter Lang.

Krongauz, Maksim. 2008. *Russkii yazyk na grani nervnogo sryva* [The Russian language at the edge of a nervous breakdown]. Moscow: Yazyki slavyanskoi kul'tury.

Krongauz, Maksim. 2013. *Samouchitel' Olbanskogo* [Teach yourself Olbanian]. Moscow: ACT.

Krylov, Grigory A. 2005. *Etimologicheskii slovar' russkogo yazyka* [Etymological Dictionary of Russian]. https://gufo.me/dict/krylov (accessed 19 November 2022).

Krysin, Leonid P. (ed.). 2014. *Tolkovyi slovar' russkoii razgovornoi rechi* [Explanatory Dictionary of the Russian colloquial speech]. Vol. 1 А-И. Moscow: Yazyki slavyanskoi kul'tury.

Larina, Tatiana, Vladimir Ozyumenko & Svetlana Kurtesh. 2020. Deconstructing the linguacultural underpinnings of tolerance: Anglo-Slavonic perspectives. *Lodz Papers in Pragmatics* 16 (2). 203–234. Special issue on (In)tolerance and (in)civility in public discourse from multidisciplinary perspectives https://doi.org/10.1515/lpp-2020-0010

Lerner, Julia & Michele Rivkin-Fish. 2021. On emotionalisation of public domains. *Emotions and Society* 3 (1). 1–11. https://doi.org/10.1332/263169021X16149420145743

Lerner, Julia & Claudia Zbenovich. 2013. Adapting the Therapeutic Discourse to Post-Soviet Media Culture: The case of Modnyj Prigovor. *Slavic Review* 72 (4). 828–849.

Levontina, Irina. 2015. *O chem rech'?* [What is it about?] Moscow: Corpus.

Levontina, Irina. 2016. *Russkii so slovarem* [Russian with a dictionary]. Moscow: Corpus.

Levisen, Carsten. 2012. *Cultural Semantics and Social Cognition. A case study of the Danish universe of meaning.* Berlin: Mouton de Gruyter.

Levisen, Carsten & Sophia Waters. 2017. How words do things with people. In Carsten Levisen & Sophia Waters (eds.), *Cultural Keywords in Discourse,* 1–23. Amsterdam: John Benjamins Publishing Company.

Mel'čuk, Igor. 2012. *Semantics: From Meaning to Text*. Amsterdam: John Benjamins.

Mel'čuk, Igor A. 2018. Anna Wierzbicka, Semantic decomposition, and the Meaning-Text approach. *Russian Journal of Linguistics* 22 (3). 521–538. http://dx.doi.org/10.22363/2312-9182-2018-22-3-521-538

Ozyumenko, Vladimir & Tatiana Larina. 2018. Understanding Social Values and Attitudes through Cultural Semantics: Tolerance vs. tolerantnost'. *INTCESS18 Proceedings. 5th International conference on education and social sciences*, 5–7 February, 2018, Istanbul, Turkey, 589–595.

Ozyumenko, Vladimir & Tatiana Larina. 2021. Threat and fear: Pragmatic purposes of emotionalisation in media discourse. *Russian Journal of Linguistics* 25 (3). 746–766. https://doi.org/10.22363/2687-0088-2021-25-3-746-766

Peeters, Bert 2015. Tall poppies in the land down under: An applied ethnolinguistic approach. *International Journal of Language and Culture* 2 (2). 219–243. http://dx.doi.org/10.1075/ijolc.2.2.04pee

Richmond, Yale. 2009. *From nyet to da: Understanding the new Russians* (4th ed.). Boston, MA: Intercultural Press.

Russian National Corpus. https://ruscorpora.ru/ (accessed 15 September 2022).

Sapir, Edward. 1949. *Selected Writings of Edward Sapir in Language, Culture and Personality.* In David Mandelbaum (ed.). Berkeley: University of California Press.

Whorf, Benjamin. 1956. *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf.* In Carroll J. (ed). Massachusetts: The M.I.T. Press.
Wierzbicka, Anna. 1992. The semantics of interjection. *Journal of Pragmatics* 18. 159–192.
Wierzbicka, Anna. 1996. *Semantics: Primes and Universals.* Oxford: Oxford University Press.
Wierzbicka, Anna. 1997. *Understanding Cultures through their Key Words: English, Russian, Polish, German, and Japanese.* Oxford: Oxford University Press.
Wierzbicka, Anna. 2021. "Semantic Primitives", fifty years later. *Russian Journal of Linguistics 25* (2). 317–342. http://dx.doi.org/10.22363/2687-0088-2021-25-2-317-342
Zalizniak, Anna, Irina Levontina & Alexey Shmelev. 2005. *Kluchevye idei russkoi yazykovoi kartiny mira* [Key ideas of the Russian linguistic picture of the world]. Moskva: Yazyki slavyanskoi kul'tury.
Zappettini, Franco, Douglas M. Ponton & Tatiana V. Larina. 2021. Emotionalisation of contemporary media discourse: A research agenda. *Russian Journal of Linguistics* 25 (3). 586–610. https://doi.org/10.22363/2687-0088-2021-25-3-586-610

Bionote:
Anna GLADKOVA received her PhD in Linguistics from the Australian National University. She is an Honorary Lecturer in Linguistics at the Australian National University and a Digital Professor at the Higher School of Economics. Her research interests include semantics, pragmatics, language and culture interface, cognitive linguistics and Natural Semantic Metalanguage. She has taught linguistics and applied linguistics at universities in Australia, Great Britain, and Russia. She is member of the Editorial Board of Corpus Pragmatics and Yearbook of Corpus Linguistics and Pragmatics (Springer).
e-mail: angladkova@gmail.com
https://orcid.org/0000-0002-4225-4805

Сведения об авторе:
Анна ГЛАДКОВА получила степень PhD по лингвистике в Австралийском национальном университете, где является почетным лектором. Также она – профессор Высшей Школы Экономики (Нижний Новгород). Ее научные интересы включают семантику, прагматику, взаимодействие языка и культуры, когнитивную лингвистику и Естественный Семантический Метаязык. Она преподавала лингвистику и прикладную лингвистику в университетах Российской Федерации, Австралии и Великобритании.
e-mail: angladkova@gmail.com
https://orcid.org/0000-0002-4225-4805

Research article / Научная статья

Pushing the boundaries:
Marginal phonemes and dialogic interaction

Nicholas EVANS ✉

*Australian National University, Australian Research Council Centre of Excellence
for the Dynamics of Language (CoEDL), Canberra, Australia*
✉nicholas.evans@anu.edu.au

Abstract
Phonemes with restricted distribution represent an interesting analytic challenge. Well-known sources include the adoption of certain phonemes from other languages in borrowed words, emerging phonemic splits, and special phonological subsystems (e.g. ideophones). This paper aims to widen our conception of such marginal phonemes, by incorporating another source: specific vocal gestures called into play in interactional settings. Our initial puzzle involves a restricted phoneme set in the Papuan language Nen: two classes of sounds are restricted to interactive contexts, namely interjections and deictics. These sounds are the nasal vowels *ã*, *ẽ*, and the glottal fricative *h*. Several questions arise here. Should these restricted sounds be considered part of the phoneme system? How did they evolve? How does their presence interact with seemingly equivalent sounds in neighbouring languages, in contexts of possible loanwords? We then pass to two other languages where sounds that are unquestionably phonemes have, in at least some phonotactic positions, clear correlations with interactive uses: initial /ð/ in English, essentially restricted to words of person (thou), space (that), time (then), or discourse deixis (the, though), and glottal stops with morphemic function in Bininj Kunwok, restricted to immediate aspect[1], addressee-engaged demonstratives, and kinship vocatives. It is already known that non-phonemic speech sounds (e.g. what is written mhm in English) are used in interaction. This paper proposes that the special phonetics of interaction can integrate further into the sound system and, in such cases as those presented here, either expand the phonological system in absolute terms by adding new phonemes, or expand the phonotactic possibilities of phonemes already occurring in other phonotactic positions.
Keywords: *marginal phonemes, Nen, Bininj Kunwok, voiced dental fricatives, restricted phonemes, phonologisation*

---

[1] 'Immediate aspect' indicates that the event is unfolding in the here and now. Bininj Kunwok lacks a present tense inflection (using a general 'non-past'), so in some circumstances this would be translated by an English present tense, but the semantics is more precise, drawing attention to the immediacy of the situation.

Раздвигая границы:
Фонемы с ограниченной дистрибуцией
и диалогическое взаимодействие

Николас ЭВАНС ✉

Австралийский национальный университет, Австралийский исследовательский консультативный центр передового опыта в области динамики языка (CoEDL), Канберра, Австралия
✉nicholas.evans@anu.edu.au

Аннотация
Исследование фонем с ограниченной дистрибуцией – это интересная аналитическая задача. Хорошо известные источники включают усвоение определенных фонем из других языков в заимствованных словах, возникающее расщепление фонем и специальные фонологические подсистемы (например, идеофоны). Цель этой статьи – расширить наши представления о такого рода фонемах с ограниченной дистрибуцией посредством привлечения других источников: специфических вокальных жестов, возникающих в условиях интеракции. Наша первая проблема связана с набором ограниченных фонем в папуасском языке нен: два вида звуков ограничены интерактивными контекстами, а именно междометиями и дейктическими словами. К ним относятся назальные гласные ã, ẽ, а также глоттальный фрикативный согласный *h*. Здесь возникают некоторые вопросы. Следует ли считать эти фонемы с ограниченной дистрибуцией частью фонематической системы? Как они возникли? Как они взаимодействуют в составе заимствований со звуками соседних языков, кажущимися эквивалентными? Затем мы переходим к двум другим языкам, где звуки, несомненно являющиеся фонемами, имеют, по меньшей мере, в некоторых фонотактических позициях явные взаимосвязи с интерактивным использованием: начальное /ð/ в английском языке, которое существенно ограничено словами, выражающими лицо (thou), пространство (that), время (then) или дискурсивным дейксисом (the, though), и морфемный твердый приступ в австралийских языках гунвиньгу, позиционное употребление которого ограничено формами глагола «непосредственного аспекта»[2], также указывающими на адресата демонстративами и терминами родства в функции вокативов. Уже известно, что в коммуникации используются нефонематические звуки речи (например, обозначаемые как mhm в английском языке). В статье высказывается мысль о том, что интерактивная фонетика может в дальнейшем интегрироваться в звуковую систему и в случаях, подобных описанным, либо расширять фонологическую систему за счет включения в нее новых фонем, либо расширять фонотактические возможности фонем, встречающихся в других фонотактических позициях.
Ключевые слова: *фонемы с ограниченной дистирибуцией, нен, яхыки гунвиньгу, звонкие зубные фрикативы, фонологизация*

---

[2] «Непосредственный аспект» указывает на то, что событие разворачивается здесь и сейчас. В языках гунвиньгу отсутствует форма настоящего времени (используется общее «непрошедшее»). Хотя в некоторых случаях можно было бы говорить о настоящем времени, термин «непосредственный аспект» является более точным, так как привлекает внимание к непосредственности ситуации.

Dedication

With this modest offering I pay tribute to Igor Aleksandrovič as a mentor, friend, inspiration and mensch. Igor was asked to visit the Australian National University in the early 1980s when I was just beginning my linguistic studies, taking up an invitation by Anna Wierzbicka. He taught a course on Meaning Text Linguistics which I had the good fortune to attend, and which offered a completely different perspective on how to integrate meaning into grammatical description, particularly at the interface between lexicon, meaning and syntax. He made a deep impression on a group of us who were just starting out on our careers as linguists. Quite apart from his linguistic work, including a crucial article (Melchuk 1979) that led Australianist linguists like Cliff Goddard (1982) to argue for a significantly different analysis of Australian case systems he mesmerised us as a larger-than-life figure. A memorable example was his lunchtime talk to the Linguistics Students Association on 'Why I am a linguist in Montréal and not in Moscow', whose Tolstoyan dimensions left the succeeding speaker (the late Michael Silverstein) standing waiting for the seminar room to become free until in desperation he began his own seminar in the corridor. Equally memorable were his inexhaustible repertoire of jokes, his complete independence of thought on all topics, and his love of cross-country skiing, a passion he shared with David Wilkins and myself in the Snowy Mountains. These deep impressions led to him being invited back to Australia a number of times, greatly enriching the linguistic scene here.

In these mad times, when his beloved Odessa lives in the shadow of a senseless war, one of his replies during a seminar on Meaning-Text Theory sticks in my mind. He had spent the whole seminar tracing the whole derivation of a sentence (I forget which) from the level of a meaning network to its linear surface form. Bob Dixon, though willing to concede that this example had worked, felt the urge to ask him something along the lines of: 'that's all very well, but not all speakers would accept this form. How do you deal with speakers who might disagree?' With a twinkle in his eye, Igor replied 'My solution is simple, and characteristically Soviet: I shoot them!' In another episode he told us about how during his time doing Soviet military service he managed to earn the right for him and his company to keep hidden the mufti they needed to go out at nights incognito. In competition with three swearing army sergeants, renowned for their mastery of swearing (мат), he compiled and memorised a vast combinatoric table of obscene linguistic elements and used it to outlast his rival in a competition to see who could swear the longest without repeating themselves: 'My soulless, inhuman technique vanquished these inspired intuitive masters' (email from IM to author, 18/1/2018).

Igor, I could never associate the words 'soulless' or 'inhuman' with you, and however much you try to remain in the realm of technique you cannot escape being inspired and intuitive!

So happy birthday, dear friend and teacher!

1. Introduction

Igor Melchuk chose, as an opening quote to one of his many masterpieces, the Explanatory Combinatorial Dictionary of Modern Russian (Melchuk & Zholkovsky 1984: 34, 38), the following lines from Vladimir Nabokov:

«…для ума	*'And no obstruction for the sage*
Внимательного нет границы	*Exists where I have put The End;*
Там, где поставил точку я:	*The shadows of my world extend*
Продленный призрак бытия	*Beyond the skyline of the page,*
Синеет за чертой страницы,	*Blue as tomorrow's morning haze –*
Как завтрашние облака,	*Nor does this terminate the phrase'.*
И не кончается строка»	
Владимир Набоков «Дар»	Vladimir Nabokov 'The Gift'

This awareness of what lies beyond the boundaries is typical Mel'čukian paradox, since one of the hallmarks of Igor's work has always been to neatly delimit a set of phenomena and then to investigate them with ruthless rigor. The present offering explores one such delimitation, and also pushes the boundaries in terms of how this current special issue is conceived. It is not about Meaning Text Theory, though it does explore one corner of the Linguistic Universe and at least touches on a phenomenon close to what Iordanskaja and Melchuk (2017) call 'pragmatèmes'. And, vast as the range of topics which Igor has tackled in his lifetime, it does not to my knowledge treat a topic he has written about. Nonetheless, at least from the point of view of this author it reflects two ways in which my own work has been inspired by his.

Firstly, his interest in the application of 'calculi of possibilities' to exploring the design space of possible linguistic phenomena, something he has advocated and employed widely (see e.g. Melchuk 2006). It was through my interactions with Igor that I realised that while many linguists with intellectual roots in the English-speaking world unconsciously take Darwin's induction from the sprawling jungle of natural organisms as the relevant scientific paradigm, for many from the Russian-speaking tradition it is rather Mendeleev who is the 'type scientist', with his demonstration of how apparent gaps in the 'calculus of possibilities' (in his case, the periodic table of elements) can be filled if we look in the right places.

Secondly, I was always struck by the following puzzle. While Igor's professional interests in linguistics rarely if ever focused on interaction, conversation or pragmatics, he is in fact one of the most interactive people I have ever met, drawing an enchanted and intense circle around all sorts of people who come into contact with him. I hope that by the end of this article the connection between these two points and the topic I examine here will become apparent.

Here, then, is the problem we pose in this contribution: how do we deal with putative phonemes which appear to be confined to words found only in interactional settings? And what are the means by which they arise?

Determining the number of phonemes in a language is a fundamental and heuristically initial problem in the study of any language. But this is often a less simple question than it appears, because of the existence of what are often called 'marginal' phonemes. And phonemes may be marginal for many reasons. They may be confined to loanwords (like English /æ/ in French *scanner* /skæne/ 'to scan', or /x/ in the pronunciation of Arabic loanwords in the Indonesian of some (predominantly Muslim) speakers, like /xabar/ <*kabar ~ khabar*> 'news'). Or they may be limited (either absolutely, or in terms of particular phonotactic positions) to onomatopoeic or ideophonic words. In Kisi, for example (Childs 1988: 172–173) the particular properties of ideophone phoneme inventories include allowing word-initial labial-velar stops (/gb/), as in /gbólúng-gbólúng/ 'ringing, switching', a specially raised and lengthened nasal vowel /ã/, word-final voiceless vowels, and the presence of a schwa phoneme. And in Nungon (Sarvasy 2016) word-initial consonant clusters like *kr* and *br* only occur in 'warblish' – ideophonic imitations of birdsong. The case that interests us here, though, constitutes a third type, which to my knowledge has not been examined in the literature:[3] the existence of phonemes that are confined, either absolutely or in particular phonotactic environments, to interactional contexts.

We begin by examining the phenomenon in Nen (§2), the language where I was most clearly forced to confront the phenomenon. I then pass to some other languages – Bininj Kunwok and Dalabon in northern Australia (§3), but also, less exotically, English (§4), before concluding in §5. It is not my goal to explore the phenomenon across the world's languages – a vast undertaking that would burst the bounds of my allocated space – but to draw attention to the phenomenon as a first step in confronting it properly.

2. Marginal phonemes in Nen and their interactional setting

Nen is a Papuan language of the Yam family – see Evans (2014, 2015a,b) for basic grammatical information. Here we focus on its phonological system, for which more detailed information can be found in Evans & Miller (2016). In

---

[3] I distinguish cases like those we will be discussing, where interactionally-derived sounds behave like phonemes, in the sense of combining with other phonemes to build morphemes and words, from the use of non-combining sounds for interactional purposes. Dingemanse et al. (2013), in their discussion of the sound approximated in their article as 'huh?', is one such case; others are e.g. the use of the reduplicated dental click to express disapproval in English (variously rendered tut-tut or tsk-tsk in English orthography), or the use of ʃ (only in the word *ʃa*) to shoo away dogs in a number of languages of Anhem Land (e.g. Bininj Kunwok) which lack fricatives in their normal phoneme inventory. While broadly relevant to the argument advanced here, in the sense of showing how interaction calls forth a wider palette of sounds than those on the regular phonemic inventory, they differ because of their lack of combinatoric options.

particular, I focus on the status of three marginal phonemes: the nasalised vowels /ã/ and /ẽ/, and the glottal fricative /h/.

The phoneme inventory for Nen is given in Figures 1 (consonants) and 2 (vowels). Most phonemes in this inventory are richly attested. However, each of the three phonemes at issue here are marginal, in terms both of the number of words they occur in and their contexts of occurrence, and so are placed in rounded brackets (as distinct from angle brackets, used for graphemes where these depart from the standard IPA values; this is the practical orthography to be used here).

	Bilabial	Dental/ Alveolar	Palatal	Velar	Labial Velar	Glottal
Plosive	p b	t̪ d		k g	k͡p \<q\> g͡b \<ḡ\>	
Prenasalised Plosive	ᵐb	ⁿd		ⁿg	ᴺg͡b (nḡ)	
Affricate			dʒ \<z\>			
Prenasalised affricate			ⁿdʒ \<nz\>			
Nasal	m	n	ɲ (ñ)			
Trill		r				
Fricative		s				(h)
Approximant			j (y)		w	
Lateral Approximant		l				

Figure 1. Nen consonant phonemes

	Front	Mid	Back
High	i	ɪ \<í\>	u
Mid	e (ẽ)	ə[4]	o
Low	æ \<ä\>	a (ã)	

Figure 2. Nen vowel phonemes

Let us now exemplify these marginal phonemes one at a time. Since the number of words they occur in is small and finite – in contrast with all other phonemes of the language – we will give complete listings.

First, /ẽ/. This is confined to two interactional words. The first of these is the word for 'yes', *ẽ*, as in (1).

1 *Bm*   *be-gre*   *n-ng-m?*   *Ẽ*   *ta-gre*   *w-ng-m*
  2ABS   2SG-alone   2SG-AWAY-be:nd   Yes   1SG-alone   1SG-AWAY-be:nd
  'Are you going on your own?'   'Yes, I'm going on my own.'

---

[4] Schwas in Nen are almost entirely predictable epenthetically, with the exception of a couple of words where they occur word-initially. When predictable, they are simply not written, e.g. \<kənəm\> [knm] 'come!' Word-initially they are written as \<á\>. /ə/ and /ɪ/ are both short and in their phonology are rather reminiscent of the jers in early Slavic.

If this were only our example, it would be easy to relegate it to the realm of once-off sounds comparable to what is represented orthographically in English as *mhm*. However, it also occurs in one of the demonstratives, *gehẽ ~ gẽhẽ*, roughly translatable as 'that one there – you should be able to find it easily, following my point'. In this demonstrative the last vowel is always /ẽ/; nasality may or may not creep back to the preceding vowel.

2 *Bä* *y-m*   *gehẽ*
 3ABS 3SG- be:nd DEM
 'There he is; here, this one is him (with pointing accompaniment)'

3 *Tande*   *nne*   *bermber*   *gehẽ*
 1SG.POSS food(ABS) portion(ABS)   DEM
 'Here's my portion (accompanying presentation of something).'

Each of these two words – *ẽ* 'yes' and *ehẽ* 'this/that one here/there' – are 'interactional' in the sense that they only make sense in closely-coupled dialogic contexts, the first as a response by one party to a question by the other, and the second when one party guides the attention of the other through pointing or presentation.

Now consider the other nasal vowel, /ã/. This is a slightly different case to /ẽ/. One of the words it occurs in is clearly interactional, while the other is an onomatopoeic bird name (and recall that onomatopoeic words and ideophones were mentioned above as another place where marginal phonemes are found).

Giving the interactional example first, there is a word /ahã/ or /ãhã/, whose primary use is in handing something to another person – something like 'here you are' in English, though there are other languages with special interjections to signal this, such as the word *nja!* In Bininj Gunwok / Mayali (Evans 1992). Examples are given in (4) and (5). There is also a secondary use, not exemplified here, with a meaning close to English 'sprung!' (but also one sense of 'aha!'), uttered when you have come across your interlocutor at the moment of doing something they shouldn't be doing.

4 *Ãhã*   *ämbs* *bm*   *t-parma-ø*
 here.you.are one 2SG(ERG)   3SG.O-break.banana.off.bunch-IMP.SG
 'Here you are (offering a bunch of bananas), break one off!'

5 *Ahã* *Gbae*   *ynd*   *begta* *tande*   *yép*
 2ABS [name]   1SG(ERG)   2SG.OBL 1SG.POSS   bag(ABS)
    räm-s-t   *n-ng-a-w-apap-nd-n*
    give-INF-AL   2SG.O-AWAY-BEN-TR-begin-ND-1SG.S
 'Here, Gbae, I'm about to give you my bag.'

The other place this phoneme occurs (for some speakers only) is in the name of the 'whistling kite', a type of bird. According to the speaker, this word is pronounced in one of four ways: *sikãka ~ sĩkãka ~ siŋkaŋka ~ sinkanka*. Each of these is phonologically anomalous in some respect. In the first two there are nasal vowels (in non-interactional words). In the third version we have the abnormal

sequence *ŋk* (/ŋg/, with voiced /g/ preceded by /ŋ/, would be alright, since prenasalised voiced stops occur, but there is no other case of a voiceless stop preceded by a homorganic velar nasal). And in the fourth version we have a heterorganic nasal + stop sequence /VnkV/; this is permitted at the phonemic level in Nen but the sequence *nk* would normally be broken up by an epenthetic vowel (see footnote 3 above), to give *[sinəkanəka], whereas the attested pronunciation among the speakers is [sinkanka].

In any case, what is important here is that for some speakers, at least, the name of the bird[5] includes a nasal /ã/ (with further optional backwards propagation of nasality, as we saw with *ahã ~ ãhã* 'here you are' and *gehẽ ~ gẽhẽ* 'this/that one I'm indicating').

Let us now pass from the nasal vowels /ã/ and /ẽ/ to the other marginal phoneme in Nen: the glottal fricative /h/. This is an interesting case, because Nen speakers are impressively multilingual, due to rules of clan exogamy that typically constitute bilingual households in which husband and wife speak different languages (Evans 2012a) and on top of that there is substantial exposure to English as a modern lingua franca. As a result, Nen speakers also knowing the neighbouring and closely related language Nmbu are familiar with, and use when speaking Nmbu, the phoneme /h/, which is in fact the regular reflex of Nen /s/, as can be seen from cognate pairs like Nen /sakr/ Nmbo /hakr/ 'brother, boy'; Nen /samba/ Nmbo /hamba/ 'village'; Nen /suri/ Nmbo /huri/ 'true'.[6] However, I do not know of any loanword from Nmbo into Nen that preserves the /h/ phoneme from Nmbo. The situation with English loanwords is rather different: according to their degree of familiarity with English, Nen speakers retain or drop English /h/ from loanwords like *headmaster* (> Nen /hedmasta/ ~ /edmasta/) or *horse* (> Nen /hos/ ~ /os/). In these cases the /h/ is word-initial and not connected with any unusual behaviour in adjoining segments. What we see with English loans, then, is a highly variable treatment of initial /h/, reflecting the complexities of how much English the speakers have in their repertoire.

In contrast to all of this, the situation with /h/ in interactional words is quite stable. All Nen speakers, in all contexts, reliably pronounce the /h/ in the words we have already seen, namely the two words /ahã/ and /gehẽ/. As far as I know (based on an initial collection of around 4,000 vocabulary items, given in Evans 2019) these are the only words with this phoneme. It is striking that in both cases the /h/ precedes a nasal phoneme (and of course that these are in turn restricted phonemes). This appears to be an instance of what Matisoff (1975) called rhinoglottophilia – the connection between laryngeal (glottal) and nasal articulations. He proposed that the effect was due to the acoustic similarity between glottal and nasal segments: both produce antiformants, due to their branched resonators, namely both nasal and

---

[5] For another bird-related Papuan example with a nasal vowel absent from the general phonemic system, see Mian, though here it is a matter of bird calls rather than bird names: 'Nasal vowels are not phonemic in Mian but the cry of a crow is consistently emulated as *hẽẽ*.' (Fedden 2011: 580).

[6] See Evans et al. (2018) for some regular sound correspondences. However, for Nmbo speakers the status of /h/ is tenuous, since younger speakers tend to drop it word initially (Kashima 2020).

oral cavities for nasal vowels, and both for glottals and laryngeals in the case of /h/, since the space below the glottis acts as a second resonator.[7]

What we have in the case of /h/, then, is a double distribution: a sociolinguistically-fluctuating distribution, word-initially, without other phonetic consequences, in the case of /h/ in loanwords, coupled with a sociolinguistically-fixed usage, by all speakers, in just a couple of interactional words, and in each case clearly associated with following nasal vowels.

More generally, to draw together our examination of marginal Nen phonemes in this section:

(a) The three relevant Nen phonemes – /ã/, /ẽ/ and /h/ – are all confined to interactional contexts, namely the interjections /ahã/ 'here you are', /ẽ/ 'yes' and /gehẽ/ 'over here/there, look!'.

(b) Each of these marginal phonemes is invariant across the speaker population, in contrast to what appear phonetically to be comparable marginal phonemes either in onomatopoeic words (*sikãka ~ sĩkãka* 'whistling kite') or loanwords (*hedmasta ~ edmasta* 'headmaster')

(c) The nasal-vowel and glottal-fricative marginal phonemes are closely interlinked, exhibiting an interesting form of rhinoglottophilia: interactive /h/ only occurs in intervocalic position before a nasal vowel, and all interactive words of more than one syllable that contain a nasal vowel also contain an /h/.

Should these be included as regular phonemes or not? There is no straightforward answer to the question. /ẽ/ forms a minimal pair with /e/ ('cry', a preverb [Evans 2019]), but for the others it is impossible to find an exact minimal pair because of the presence of two, mutually conditioned, restrictive phonemes (i.e. since both a nasalised vowel, and an /h/, are co-present, it is not possible to find a word differing on just one of these). In other words we can contrast $eC_{<-nasal>}e$ with $ẽhẽ$ (as in *gege* 'son' vs *gẽhẽ* 'over here/there, look!'), but we can't independently vary the vowel nasality and the glottal articulation to construct contrasts on just one of those.

If we appeal to number of words the case for including them is weak; on the other hand they exhibit regular phonotactics apart from the rhinoglottophilia effects just mentioned. The most convincing answer is probably to say that they are phonemes, but not regular ones: that the phonology is structured around a core of fully regular phonemes augmented by an outer layer of more restricted ones, to which /ã/, /ẽ/, and /h/ belong. And unlike all other phonemes, for which it is not possible to find a unifying semantic or pragmatic characteristic, these three restricted phonemes all have a strong 'interactional' flavour – in each case they occur in contexts where the dialogic element of language is particularly salient: the interaction between speaker and hearer across question-and-answer (*ẽ* 'yes'),

---

[7] For further examples of rhinoglottophilia see e.g. Krim, which lacks contrastive nasal vowels but in which are strongly nasalised after /h/, Pirahã which exhibits similar nasalisation effects after /h/ and /ʔ/, and Inor (Gurage, Ethiopia) which has developed nasal vowels where there were etymological laryngeals/pharyngeal consonants. See Ahland (2006) and Boivin (1996).

request-and-fulfilment (*ahã* 'here you are') or training of mutual attention (*gẽhẽ* 'here/there, look!').

3. An English puzzle: initial /θ/ vs /ð/

After our initial discussion of a little-known Papuan language, Nen, the reader may ask how relevant such obscure languages are to general questions of linguistics. It therefore behoves us to return to a well-studied language, English, to show that comparable effects are to be found there, in a different guise, namely the distribution between /θ/ and /ð/ in word-initial position.

As is well-known, English contrasts fricatives, by voicing, at several points of articulation, and in all major phonotactic positions – word-initial (*feel* vs *veal*), word-medial (*elfish* vs *elvish*) and final (*life* vs *live*[8]). Curiously, however, the contrast between /θ/ and /ð/ (unhelpfully represented by the same digraph <*th*>) is not found in all three positions. While occasional minimal pairs can be found word-finally, particularly in association with noun-verb heterosemy: *wreath* [θ] (n.), *wreath* [ð] (v.); *teeth* [θ] (n.), *teeth* [ð] (v.), *mouth* [θ] (n.), *mouth* [ð] (v.).

However, there are basically[9] NO minimal pairs word-initially: both phonemes are found in word-initial position, but they are assorted by lexeme, as in Table 1.

Table 1. **English /θ/ vs /ð/ in word-initial position**

θ	ð
thick, think	this, these
thin	that. those
through	then
thresh	there, thither, thence
thew, three, thigh	thou, thee, thy
thought	though
theft	they, them, their
thalidomide	the
thank	than

Different scholars have commented on the semantics of this division. Minkova (2011: 39) refers to it as 'initial voicing [of [θ] ~ [ð] – NE] in … function and pronominal words'.[10] She points out that this distinction was already present in Old

---

[8] The deficiencies of English orthography make it necessary to specify that the pronunciation of *live* intended here is /laiv/, as in 'live wire' or 'live show'.

[9] Depending a bit on where we draw the boundaries. If we include the archaic *thy*, then *thy* vs *thigh* is a minimal pair. If we allow multi-word sequences, with *this'll* [ðɪsel] (< *this will*) contrasts minimally with *thistle* [θɪsəl]. Both these pairs contrast a closed-class, interactional word with an open-class noun.

[10] Bickel & Nichols (2007) introduce the notion of 'eidemic resonance' to account for how 'forms of a paradigm often resonate with each other through alliteration, rhyme, or other paronomasia, but without entailing any consistent semantics. Rather, the resonances serve to structure paradigms, compartmentalize the lexicon, and provide psycholinguistic processing cues.' They see it as 'probably best attested in small closed lexical paradigms such as personal pronouns, basic kin terms …, essential deictics, and the like, but also … in inflectional paradigms'. While there

English, and calls it a 'transparent case of prosodically induced change', without discussing why prosody should operate differently in precisely this set of words; elsewhere (Minkova 2014: 94ff) she argues that it operates in words that usually appeared in prosodically weak positions. Further, she advances a scenario (synthesising analyses by Bennett (1955) and Lass (1992: 41) in which '/θ-/ was categorically voiced in initial position in Old English (West Saxon, Kentish, or West Mercian)' (ibid: 40), with 'a later reversal of the initial voicing in major class words, presumably under dialectal influence from the northern areas where the voicing did not occur' (ibid: 40). This raises the question, though, of why all and only the 'major class' words should be subject to the devoicing influence of other dialects.

Lass (1992: 59) characterises the group of modern forms with initial /ð/ as 'deictics like *the, this, that, these, there, then, thou* and a few conjunctions like *though*'. He states that these normally occur 'under low sentence stress', but while this is certainly true[11] for words like *the* and perhaps *though* it is far from true when deictics are used as one-word answers in ignorative-deictic sequences (Karcevski 1941, Wierzbicka 1980, Evans 2012b) like *Which? This!*, *When? Then!* Or *Where? There*.

This characterisation by Lass can be pushed further. Of the words beginning with /ð/ in Table 2, it is not unreasonable to say that are all are in fact deictics: of space in *this/these*, *that/those*, and *there/thither/thence*, of person in *thou/thee/thy*, of time in *then*, of presupposed identifiability (i.e. discourse deixis) in *the*, and of speaker beliefs about the expected compatibility of two events in *though*. The only apparent exception is *than*, but the etymology of this goes back to Old English þanne, a variant of þonne ('then, since, because'), in turn from Proto-Germanic *þan ('at that, at that time, then'), so that at least etymologically it is also a deictic. Aside from *than*, all these words are fundamentally interactional in synchronic terms, although they take in a wider range of interaction types than the Nen words examined in §2. On the other hand, none of the words with /θ/ are deictic or interactional in this sense. We therefore have a very clear semantic partition of the two realisations: interactional for /ð/, non-interactional for /θ/.

Before going on it is necessary to address another possible explanation: frequency. Could it be the case that is simply the higher-frequency words that exhibit the voiced forms? It turns out that this is only partially true. If we take the listing of English word frequencies at (ENA, 13 October 2022)[12], and look at the top 200 words by frequency, we find that 17 begin with /ð/ or /θ/, and that there is a correlation, but not an absolute one. In terms of frequency, we find the ranking in Table 2.

---

are similarities to the present case, they do not specifically mention that eidemic resonance can condition allophonic differences, nor do the cases discussed here form a paradigm in the strict sense.

[11] And of course 'normally' is not categorical: in a sentence like *Igor is THE expert on this* the article receives full stress but retains its initial voicing.

[12] https://frequencylist.com

Table 2. **Frequency rankings of words beginning with interdental fricatives, top 200 words of English**[13]

ð		θ	
Freq #		Freq #	
4	the		
9	that		
18	this		
		38	think
51	they		
70	there		
		73	thing
83	there		
91	them		
93	then		
		119	thank
130	these		
140	those		
163	than		
178	their		
		188	three
		191	thanks

Source: ENA, 13 October, 2022.

The data in Table 2 suggest that although frequency is a correlation with voicing of dental fricatives, the correlation is not perfect. It is therefore unlikely that frequency is, in itself, the reason for the split distribution we find in English. We have also rejected, above, the idea that these simply reflect low sentence stress. The issue thus remains an unresolved puzzle. It is useful, then, to look at another language in which a particular phoneme has – at least in some positions – a highly interaction-dependent distribution, and it is to this language – the Australian language Bininj Kunwok – that we turn in the next section.

4. A clue from Bininj Kunwok

Bininj Kunwok[14], like many other languages of Arnhem Land, Northern Australia (Harvey 1991, Evans 1995), has a phonemic glottal stop with a very limited distribution.

In morpheme-final, syllable-final position it contrasts with zero in a few words, e.g. *kunwor* 'leaf' vs *kunworʔ* 'satiation', *lar* 'sandpaper fig' vs *anlarʔ* 'callitris pine'. In none of these cases does it form a minimal semantic contrast, in other

---

[13] https://frequencylist.com

[14] Orthography for the language has now stabilised on this spelling; at the time of publishing my pan-dialectal grammar (Evans 2003) the spelling *Gun-wok* for the second part of the name was still prevalent.

words it is not the glottal stop itself but the assemblage of sounds which gives the morpheme its meaning.[15]

On the other hand, there are three sets of words in which the glottal stop does constitute a morpheme in itself:

(a) 'immediate aspect' on verbs (Evans 2003: 524–5) – something happening in the here and now, forming minimal pairs with verbs lacking the glottal stop, which typically have a generic sense. Cf *ŋaŋun* 'I eat', *ŋaʔŋun* 'I am eating right now'. In this use, the glottal stop directly follows the subject pronominal prefix (here *ŋa-* '1SG.SUBJ') and precedes the verb stem (here *ŋu* 'eat'; *-n* is the non-past suffix).

(b) immediacy in some demonstrative contrasts (Evans 2003: 290–302). Cf *nabeŋu* 'the one over there' vs *nabeʔŋu* 'the one which you wanted to know about, which is right here (presenting it at the moment of utterance)', or *nani* 'there (in a series, e.g. a series of stops on a journey)' but *naʔni* 'this one right here (presenting an object to the addressee'

(c) vocatives of some kin terms, e.g. *beḷu* 'auntie (referential)' *beḷuʔ* 'hey auntie!'

Each of these uses is clearly interactional in the sense of relating the statement to the here and now: in (a) it locates the unfolding action to the moment of speech, in (b) it draws the addressee's attention to an entity being presented for their attention, and in (c) it summons the addressee by calling their attention through the use of a kinship term.

While this generalisation is true of these basic uses, there are two constructions in Bininj Kunwok which have taken these basic interactional uses and extended them to non-interactional meanings:

(a) the immediate aspect also gets used in complements of perception verbs, as in (6). This is best treated as a type of 'displaced immediacy', comparable to the displaced deixis in direct speech ('He thought: 'There's someone out there'').

6 *Ba-na-ŋ*   *ka-ʔ-baŋḍi*
   3SG.SUBJ:PST-see-P.PFV   3SG.SUBJ:PRES-IMM-hang.up(NPST)
   'He saw it hanging up'

(b) the vocative use gets used with address-based kinship verbs, e.g. in *ŋaŋalkurŋʔme* 'I call her *ŋalkurŋ* 'mother-in-law'. Here the /ʔ/ is added to the kin term *ŋalkurŋ* 'mother-in-law' to form a displaced vocative, which is then incorporated into the verb in a special 'call OBJ kin' construction, literally 'I 'mother-in-law!' call her'. See Evans (2000) for other examples of languages which form 'call Obj Kin' verbs by incorporating kinship expressions into a verbal stem;

What is interesting about both these constructions is that they show how semantics that originate in interactional settings can be adapted into less contextual

---

[15] In at least one case, one word of such a pair is interactional in the sense used here: cf *kun-waŋ* 'armpit', *waŋʔ* 'well then; now; let's move on to the next thing' (Evans 1992). However, as with the other words exemplified here, the ʔ is not in itself a morpheme in this case.

uses through displaced deictic use. Concomitantly, they show how phonemic possibilities that are initially constrained by interactional settings may have their uses widened as interactional meanings are adapted into other construction types.

5. Conclusion

We have explored three languages in which phonemes, marginal in one way or another, are all linked to what we have broadly labelled 'interactional uses'. In Nen the relevant phonemes are the glottal fricative /h/ and the two nasal vowels /ã/ and /ẽ/, between them restricted to a handful of words only found in face-to-face interaction: 'here you are', 'over there look!', 'yes'. In English the marginality – concerning the voicing of the dental fricatives /ð/ and /θ/ – is limited to word-initial position. Elsewhere in the word the contrast serves other roles (e.g. limited possibilities of converting nouns into verbs). But initially there is a clean cleavage of /ð/ into words that are synchronically or diachronically deictic, and /θ/ into words whose meaning is simply referential, and does not need to take interactional context into account. In Bininj Kunwok the relevant phoneme is the glottal stop: while this sound can occur intra-morphemically as part of referential lexical contrasts (e.g. 'leaf' vs 'satiation') its primary use as a morpheme comprising a sole phoneme is restricted to the three interactional senses of 'immediate aspect (in the here and now)', 'engaged attention' (in demonstratives) and 'calling the attention of designated kin' (in kinship vocatives).

What is common to all these examples is that particular vocal gestures – vowel nasalisation (Nen), initial voicing of fricatives (English), glottal fricatives (Nen) and glottal stops (Bininj Kunwok) – seem to arise, whether entirely (Nen) or in some phonological or morphological positions (English, Bininj Kunwok) – specifically in contexts of interaction.

We can hypothesise that what has happened, in such contexts, is that suprasegmental prosodic signalling, aimed at attracting or directing the addressee's attention, has entered the speech system as part of the suite of paralinguistic and prosodic sounds used to modulate conversational interaction (agreeing, pointing, presenting...). In doing so, it has become associated with particular words or lexical sets to the point where it has ceased to be simply 'extraphonemic', and begun to be integrated into the core phonological system. Sometimes, as in Nen, this process is at a very early stage, and the phoneme is only attested in a handful of words. At other times, as in English, the process has advanced much further and is only detectable in particular phonotactic positions.

If correct, the importance of this mechanism for our understanding of how phonological systems evolve is that it removes the impermeable barrier between 'extralinguistic' speech sounds found just in interaction (but in a non-combining way), and the set of phonemes which a language uses to construct new words and morphemes. As examples of 'extralinguistic' speech sounds, we may consider the discussion by Dingemanse et al. (2013) of the sound approximated in their article as 'huh?', the use of the reduplicated dental click to express disapproval in English (variously rendered *tut-tut* or *tsk-tsk* in English orthography), or the use of ʃ (only

in the word *ʃa*) to shoo away dogs in a number of languages of Arnhem Land (e.g. Bininj Kunwok) which lack fricatives in their normal phoneme inventory. In each of these cases, a full study of the communicative use of speech sounds in interaction calls forth a wider palette of sounds than those on the regular phonemic inventory. But, in contrast to the cases discussed here, the relevant sounds lack combinatoric options.

In another of his important works, Igor and his long-term collaborator Lydia Iordanskaja introduce the notion of 'pragmatemes' (pragmatèmes in French), defining them as follows:

'Un formulème est un pragmatème si et seulement si il est contraint par rapport à la situation extralinguistique de son utilisation.' (Iordanskaja & Melchuk 2017: 102)

This characterisation relates closely to the phenomena discussed in this article – with the exception that pragmatemes are lexical or even phrasal items rather than the phonological building blocks from which they are constructed. A bit further on in the same book (p. 104), they point out that

'toute unité linguistique, y compris un syntagme compositionnel, peut être contrainte par la situation extralinguistique de son emploi'

The phenomena discussed here show that this observation can be applied to the very building blocks of a language: its phonemes. The pragmatic demands of interaction feed, over time, into the organised heart of the linguistic system, through the impact of prosody on such features as nasalisation and glottal gestures for signalling aspects of the here-and-now. And in this way the intense human interaction which makes us all think of Igor with such vivid warmth, also has its part to play in the never-ending creation and re-creation of our phonological systems.

REFERENCES

Ahland, Michael. 2006. Nasal spreading, rhinoglottophilia and the genesis of a non-etymological nasal consonant in Mesmes. *Proceedings of the annual meeting of the Berkeley Linguistics Society.* 13–24.

Bennett, William H. 1955. The Southern English development of Germanic initial [f s]. *Language* 31 (3). 367–371. https://doi.org/10.2307/410804

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description,* 169–240. Cambridge: Cambridge University Press.

Boivin, Robert. 1996. Spontaneous nasalization in Inor. In Grover Hudson (ed.), *Essays on Gurage language and Culture*, 21–33. Wiesbaden: Harrassowitz Verlag.

Childs, George Tucker. 1988. The phonology of Kisi ideophones. *Journal of African Languages and Linguistics* 10 (2). 165–190. https://doi.org/10.1515/jall.1988.10.2.165

Dingemanse, Mark, Francisco Torreira & Nick J. Enfield. 2013. Is "Huh?" a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. *PLOS ONE* 8 (11). e78273. https://doi.org/10.1371/journal.pone.0078273

Evans, Nicholas. 1992. Wanjh, bonj, nja: Sequential organization and social deixis in Mayali interjections. *Journal of Pragmatics* 18 (2/3). 71–89. https://doi.org/10.1016/0378-2166(92)90053-E

Evans, Nicholas. 1995. Current issues in Australian phonology. In John Goldsmith (ed.), *Handbook of Phonological theory,* 723–761. Oxford: Blackwells.

Evans, Nicholas. 2000. Kinship verbs. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the Typology of Word Classes,* 103–172. Berlin: Mouton de Gruyter.

Evans, Nicholas. 2003. *Bininj Gun-wok: A Pan-Dialectal Grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.

Evans, Nicholas. 2012a. Even more diverse than we thought: The multiplicity of Trans-Fly languages. In Nicholas Evans & Marian Klamer (eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st century. Language documentation and conservation special publication No. 5*, 109–149. Honolulu: University of Hawai'i Press.

Evans, Nicholas. 2012b. Nen assentives and the problem of dyadic parallelisms. In Andrea C. Schalley (ed.), *Practical theories and empirical practice. Facets of a complex interaction*, 159–183. Amsterdam: John Benjamins.

Evans, Nicholas. 2014. Positional verbs in Nen. *Oceanic Linguistics* 53 (2). 225–255.

Evans, Nicholas. 2015a. Chapter 26. Valency in Nen. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages,* 1069–1116. Berlin: Mouton de Gruyter.

Evans, Nicholas. 2015b. Inflection in Nen. In Matthew Baerman (ed.), *The Oxford handbook of inflection*, 543–575. Oxford: Oxford University Press.

Evans, Nicholas & Julia Colleen Miller. 2016. Nen. *Journal of the International Phonetic Association* 46 (3). 331–349. https://doi.org/10.1017/S0025100315000365.

Evans, Nicholas, Wayan Arka, Matthew Carroll, Yun Jung Choi, Christian Döhler, Volker Gast, Eri Kashima, Emil Mittag, Bruno Olsson, Kyla Quinn, Dineke Schokkin, Philip Tama, Charlotte van Tongeren & Jeff Siegel. 2018. The languages of Southern New Guinea. In Bill Palmer (ed.), *The languages and linguistics of New Guinea: A comprehensive guide*, 641–774. Berlin: Walter de Gruyter.

Fedden, Sebastian. 2011. *A Grammar of Mian*. Berlin: Mouton de Gruyter.

Goddard, Cliff. 1982. Case systems and case marking in Australian languages: A new interpretation. *Australian Journal of Linguistics* 2. 167–196. https://doi.org/10.1080/07268608208599290

Harvey, Mark. 1991. Glottal stop, underspecification and syllable structure among the Top End Languages. *Australian Journal of Linguistics* 11. 67–105.

Iordanskaja, Lydia & Igor Melchuk. 2017. *Le mot français dans le lexique et dans la phrase.* Paris: Hermann.

Karcevski, Serge. 1941. Introduction à l'étude de l'interjection. *Cahiers Ferdinand de Saussure* 1. 57–75.

Kashima, Eri. 2020. *Language in My Mouth: Linguistic Variation in the Nmbo Speech Community of Southern New Guinea* [Unpublished PhD Dissertation]. The Australian National University.

Lass, Roger. 1992. Phonology and morphology. In Norman Blake (ed.), *The Cambridge history of the English language. Volume II: 1066–1476,* 23–155. Cambridge: Cambridge University Press.

Matisoff, James, A. 1975. Rhinoglottophilia: The Mysterious connection between Nasality and Glottality. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, 265–287. Universals Language Project, Stanford University, Stanford.

Melchuk, Igor. 1979. *Studies in Dependency Syntax*. Ann Arbor: Karoma.

Melchuk, Igor. 2006. Calculus of possibilities as a technique in linguistic typology. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The art and craft of grammar writing,* 171–205. Berlin: Mouton de Gruyter.

Minkova, Donka. 2011. Phonemically contrastive fricatives in Old English? *English Language and Linguistics* 15 (1). 31–59. doi:10.1017/S1360674310000274

Minkova, Donka. 2014. *A Historical Phonology of English*. Edinburgh: Edinburgh University Press.

Sarvasy, Hannah. 2016. Warblish: Verbal mimicry of birdsong. *Journal of Ethnobiology* 36 (4). 765–782. https://doi.org/10.2993/0278-0771-36.4.765.

Wierzbicka, Anna. 1980. *Lingua Mentalis: The Semantics of Natural Language.* Sydney: Academic Press.

Dictionaries

Evans, Nicholas. 2019. *Nen dictionary. Dictionaria 8*. 1–4997. (Available online at https://dictionaria.clld.org/contributions/nen) (accessed 12 October 2022).

Melchuk, Igor & Aleksander Zholkovsky. 1984. Tolkovo-kombinatornyi slovar' sovremennogo russkogo yazyka. [Explanatory Combinatorial Dictionary of Modern Russian]. Vienna: Wiener Slawistischer Almanach.

Bionote:
Nicholas EVANS is Distinguished Professor of Linguistics at the Australian National University. He has carried out wide-ranging fieldwork on indigenous languages of Australia (Kayardild, Bininj Kunwok, Dalabon) and Papua New Guinea (Nen). His driving interests are the interplay between the diversity of the world's languages and the many scientific and humanistic questions they can help us answer about human history, culture, mind and society.
e-mail: nicholas.evans@anu.edu.au
https://orcid.org/0000-0003-0893-3713

Сведения об авторе:
Николас ЭВАНС – Заслуженный профессор лингвистики Австралийского национального университета. Он проводит обширные полевые исследования языков коренных народов Австралии (каярдильд, гунвингу, далабон) и Папуа – Новой Гвинеи (нен). Основная сфера его научных интересов – взаимосвязь между многообразием языков мира и многочисленными научными и гуманитарными вопросами о человеческой истории, культуре, мышлении и обществе, на которые они могут помочь нам найти ответы.
e-mail: nicholas.evans@anu.edu.au
https://orcid.org/0000-0003-0893-3713

Research article / Научная статья

Perfective, performative and present: Some non-standard combinations in Slavic and beyond

Vladimir PLUNGIAN[1,2,3] ✉, Ekaterina RAKHILINA[1,4] and Tatiana REZNIKOVA[4]

[1] *Vinogradov Institute of Russian Language / Russian Academy of Sciences, Moscow, Russia*
[2] *Lomonosov Moscow State University, Moscow, Russia*
[3] *Institute of Linguistics / Russian Academy of Sciences, Moscow, Russia*
[4] *HSE University, Moscow, Russia*
✉plungian@gmail.com

Abstract

The combination of perfective aspect and present tense is frequently considered as an example of semantically incompatible grams. If verbal forms including markers of both perfective aspect and present tense do exist in a language, they tend not to express present resp. perfective in the strict sense. Thus, in Russian such forms usually convey the future, as in *napishu* 'I will write'. The article discusses a specific type of contexts where these forms develop a less trivial meaning of what can be called "prospective present". Obligatory components of these contexts are first person of the verb and negation. We focus on three instances of this kind*: ne skazhu* (lit. 'I won't tell'), *ne dam* (lit. 'I won't give') and *ne pushchu* (lit. 'I won't let')'. With the data of Russian National Corpus (RNC) and notably of the parallel corpora within RNC, we demonstrate that in certain uses, these constructions correspond to speech acts of refusal or prohibition and can be viewed, accordingly, as expressing a kind of performative meaning. As performatives, these verbs refer to a present situation: the speaker's refusal or prohibition comes into operation at the moment of utterance, and not at some point in the future. The present-tense reference is corroborated by the translational counterparts of *ne skazhu* / *ne dam* / *ne pushchu* from parallel corpora, as other languages regularly use present forms in these contexts. Thus, performative-like constructions provide new data on potential non-future meanings of perfective present forms.

Keywords: *perfective, prospective, performative, negation, present, future*

Перфектив, перформатив и настоящее время: некоторые нестандартные комбинации в славянских и других языках

В.А. ПЛУНГЯН[1,2,3] ✉, Е.В. РАХИЛИНА[1,4], Т.И. РЕЗНИКОВА[4]

[1] *Институт русского языка им. В.В. Виноградова РАН, Москва, Россия*
[2] *МГУ им. М.В. Ломоносова, Москва, Россия*
[3] *Институт языкознания РАН, Москва, Россия*
[4] *Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*
✉plungian@gmail.com

Аннотация

Перфектив и настоящее время часто называют в качестве примера семантически несовместимых граммем. Если в языке существуют глагольные формы, включающие одновременно показатели и перфектива, и настоящего времени, то, как правило, они либо не имеют семантики настоящего, либо не выражают перфективность в строгом смысле. Так, в русском языке такие формы обычно передают значение будущего, например, *напишу*. В статье рассматривается особый тип контекстов, в которых эти формы получают менее тривиальное значение – его можно было бы назвать «проспективным настоящим». Обязательными компонентами этих контекстов являются первое лицо глагола и отрицание. Подробнее мы обсудим три примера такого рода: *не скажу*, *не дам* и *не пущу*. На материале Национального корпуса русского языка (НКРЯ) и, в частности, параллельных корпусов в составе НКРЯ мы покажем, что в некоторых случаях эти конструкции соответствуют речевым актам отказа или запрета и тем самым могут считаться своего рода перформативами. В качестве перформативов эти глаголы относятся к плану настоящего: отказ или запрет говорящего вступает в силу в момент высказывания, а не когда-то в будущем. Отсылка к плану настоящего подтверждается переводными аналогами сочетаний *не скажу*, *не дам* и *не пущу* по данным параллельных корпусов: другие языки регулярно используют в этих контекстах формы настоящего времени. Таким образом, квази-перформативные конструкции расширяют наши представления о значениях, которые могут приобретать перфективные презентные формы.

Ключевые слова: *перфектив, проспектив, перформатив, отрицание, настоящее время, будущее время*

Instead of a foreword

Once, Igor' Aleksandrovich happened to write: "My very first words in life were said in Russian. And my last words <…> will most certainly be in Russian" (Mel'chuk 1995: xvii). That's why we dare preface our paper with another piece of Russian which would significantly fade when translated.

Никому из авторов этой статьи не повезло учиться у Игоря Александровича: кто-то из нас входил в лингвистику, когда он уже уезжал, а кто-то – когда он уже давно уехал. Тогда все были уверены, что уехать – это навсегда, и русский язык отражал это грамматически: о живых и здравствующих эмигрантах обычно говорили, используя формы прошедшего времени. Обрублено было не только личное общение, но строго преследовалось и научное: запрещено было преподавать модель «Смысл <=> Текст», опираться на ее достижения и даже просто ссылаться на ее существование и упоминать ее автора. За этим следили, нарушители преследовались: В.М. Андрющенко, уже в 80-е осмелившийся вставить в свой обзор по машинному переводу ссылку на англоязычную работу Игоря Александровича и его коллег, был вызван на ковер и заработал сердечный приступ, от которого так и не оправился.

Но как раз в этом отношении мы были гораздо счастливее прочих: благодаря нашим учителям и старшим лингвистическим друзьям (огромное спасибо им всем за это!) мы не только знали имя Мельчука, но и, как лингвисты, «варились» в его теории. Включать ее в программу официально было, конечно, нельзя, но каким-то образом мы «проходили» ее основные положения, причем на разных курсах. Застрельщиком в этом отношении была, безусловно, Анна Константиновна Поливанова, которая успела поработать с Мельчуком в Ин'язе, в знаменитой Лаборатории машинного перевода (об этом много написано в недавней книге Бурас 2022). Первой лингвистической статьей, которая запомнилась на всю жизнь, была фундаментальная работа об определении основных понятий морфологии Мельчук 1975 (из которой впоследствии вырос пятитомный «Курс общей морфологии»): она каким-то чудом вышла перед самым отъездом автора в знаменитом тогда журнале под ред. В.А. Успенского с загадочным названием «Семиотика и информатика». Статью велено было прочесть и… предложить свои альтернативные решения. Мы были первокурсники. Более увлекательного задания трудно было себе представить. Спорили сутками. Жалко только, что без Игоря Александровича – уж он бы оценил. Дальше были другие статьи: о грамматических значениях (Мельчук 1961), о парадоксе пары *катать~кататься* (Мельчук 1968), о супплетивизме (Мельчук 1972). Мы знали их почти наизусть. По сравнению с традиционной русистикой они завораживали свободой, и не только свободой мысли, но и свободой изложения. Личность автора (сплошная невиданная свобода!) просвечивала в них в каждой строке. Сам автор был далеко в Канаде, но идеями можно было и восхищаться, и, как мы привыкли, – спорить. «В России надо жить долго»: эти споры и обсуждения всё же продолжились потом и напрямую, в переписке по поводу перевода «Курса общей морфологии» или большой публикации работы о валентностях (Mel'chuk 2004).

Так что мы тоже, конечно, учились у Мельчука…

1. Preliminaries

It is a well-known fact that grammatical markers can be mutually incompatible: in this case, they are not allowed to modify one and the same word-form even if they belong to different grammatical categories and in principle can be thought of as coexisting within one lexical unit. Usually, the incompatibility of this type is explained in semantic terms: the combination of two corresponding grams appears semantically impossible, because they include elements which are semantically contradictory. Examples of such contradictory combinations can be 'imperative' + 'past' (imperatives normally apply to a future event), 'irreal' + 'assertive' (the first, unlike the second, normally applies to a non-realized event), etc. Beyond these plain observations, few researchers have ventured so far to elaborate on this issue. However, a useful discussion, with some important findings and generalizations, can be found, for example, in (Aikhenvald & Dixon 1998/2011, Malchukov 2009, 2011, 2019, Khrakovskii 1990, 1996, Khrakovskii & Malchukov 2016).

Among the frequently mentioned examples of semantically incompatible grams the case of perfective and present seems to be one of the best studied. Indeed, what can be characterized in formal terms as 'perfective present' appears to be a semantically awkward combination: verbal forms including markers of both perfective aspect and present tense reference either do not exist in the world's languages or do not express present resp. perfective in the strict sense. More specifically, if they do not express present, they usually convey either the future (as in East and West Slavic) or the (resultative or recent) past values (as in many Creoles or in Samoyedic and Tungusic). On the other hand, if they do not express perfective, they usually express habitual (as in Modern English or, in some contexts, in Bulgarian). The phenomenon is basically known as "present perfective paradox" (as proposed in Malchukov 2009); cf. also an in-depth cross-linguistic overview in De Wit 2017, with special reference to French, English, Slavic, and Sranan Tongo Creole systems, as well as some interesting additional data from West African and North Siberian languages in Shluinskii 2012. As De Wit (2017) puts it, the straightforward semantic explanation of this type of incompatibility lies in the fact that "<…> there is an epistemic alignment constraint preventing the identification and reporting of events in their entirety at the time of speaking".

It should be noted that, if we consider this problem from a diachronic point of view, we can see that perfective presents often go back to non-perfective presents which gradually develop a perfective meaning and then impose a non-present construal via coercion (see Michaelis 2004, among others). Cf. English Present Simple (as in *plays* or *runs*) which could be considered, according to De Wit 2017, as an instance of what is a perfective present *morphologically*, an aspectually neutral present *diachronically*, and a habitual present *semantically*.

Accordingly, what we are going to discuss in the main part of this paper is related mostly to less trivial diachronic semantic changes of the perfective meaning. Especially interesting are those cases where what are now genuine perfective forms

seem to maintain (at least to some degree) the present tense reference without any prominent signs of coercion-driven effects.

One of such cases lies at the intersection of **prospective** and **performative** semantics. Russian will serve us as the main source of examples here. However, what we find in Russian is more or less typical for other East and West Slavic languages, South Slavic representing a slightly different configuration.

2. Prospective, performative, and present perfective paradox

Recall that **prospective** is an aspectual gram (in a broad sense of the term) which, roughly, describes the state corresponding to a preliminary phase of some imminent situation (as in English *be going to V*). The grammaticalization of prospective is not uncommon: it is attested in English and some Romance languages, very frequent in Turkic and elsewhere in Eurasia, Africa etc.; cf. Korn & Nevskaya 2017 for a recent cross-linguistic overview, and Kozlov 2021 for a more fine-grained theoretical account. On the other hand, **performative** (the notion goes back to J. L. Austin's and J. R. Searle's work) is a common term for first person forms V_1 such that saying V_1 is equivalent to performing the event 'V': thus, saying *I promise P*, the speaker just makes a promise to do P; saying *I declare P*, the speaker performs the act of declaring P, etc. (for the analysis of performatives in Russian cf. primarily Apresyan 1988/1995 and Voeikova 1996).

For the issues considered here it is particularly important that cross-linguistically performatives naturally tend to a present tense reference, though other configurations (for example, present perfect, perfective past or future) are also attested. On the other hand, their aspectual behaviour (if applicable) is somewhat more complicated: performative semantics can effectively draw on both perfective and imperfective – cf. notably Wiemer 2014, Dickey 2016 and Biasio 2021a, 2021b for Slavic, and de Wit et al. 2018 and Fortuin 2019 in a wider cross-linguistic perspective.

Before we proceed to a more detailed discussion, let us briefly recapitulate some well-known features of Russian verbal inflection. Modern Russian distinguishes two structurally different tense-marking systems ("past" and "present", with a certain amount of conditionality) as well as two morphologically different aspectual stems ("imperfective" and "perfective"). The past is obtained by the suffix *-l-* followed by a set of gender/number-marking flections, whereas the present is obtained by a different set of person/number-marking flections alone. Imperfective stems, for their part, are either underived, or "simplex" (i.e. without any overt morphological marking) or contain a number of imperfective suffixes like *-yva-*, *-va-*, *-a-* and some other. Perfective stems can be also underived (very few) or – in most cases – prefixal (with a large set of verbal prefixes expressing also spatial and related values) or suffixal (with semelfactive *-nu-*). The combination of two tense forms and two aspectual stems yields four possible grammatical classes, namely imperfective and perfective past and imperfective and perfective present:

of all of them, the last one is the most problematic semantically (in line with what has already been discussed above).

Indeed, what is morphologically a perfective present in Russian (i.e., a combination of a perfective stem and suffixal person/number markers) is usually described as a form with a default (perfective) **future** reference, as in *my* [*nepremenno*] **sdelaem** *eto* 'we'll [certainly] **do** it'. In a number of specific constructions, however, this verbal form can have a (diachronically older) **present** meaning – usually, in combination with iterative, habitual, or potential readings, as in contexts like *to syadet, to vstanet* 'always sitting down and standing up' (iterative), *vsyakii skazhet* 'anybody would tell you' (potential or habitual), *otkroet lyuboi zamok za 45 sekund* 'could open any lock in 45 seconds' (potential). As a specific type of potential reading (primarily, in the context of negation) the so-called "present of vain expectation" can also be considered, originally described by Zaliznyak (1990); the stock example here is *deneg vse ne soberem* ≈ 'we still haven't been able to collect money', a famous line by Bulat Okudzhava.

However, there exist other uses of perfective present in Russian (mostly neglected in previous studies) which are not fully consistent with the list above, i.e. perfective presents which are neither future nor habitual. We believe that they can be best labelled semantically as **prospective**, because they describe a situation that starts immediately at the moment of uttering the verbal form and then continues into the future. At the same time, they are semantically performative (or quasi-performative), because they usually correspond to speech acts of permission or prohibition.

3. Data and methodology

In what follows, we will focus on three typical instances of such performative-like prospectives (or, maybe, prospective-like performatives). These are *ne skazhu* (lit. 'I won't tell'), *ne dam* (lit. 'I won't give') and *ne pushchu* (lit. 'I won't let'). Of course, this list is not intended to be exhaustive; however, these three expressions (i) are the most frequent in the corpus, (ii) form a rather homogeneous set and (iii) can easily be identified and found in the parallel corpora which were at our disposal within Russian National Corpus. Therefore, we prefer to concentrate on their semantic properties and leave aside the question about other possible representatives of this group (if any).

Recall that all the three verbal forms are morphologically perfective presents, and their default out-of-context construal is undoubtedly futural, as befits the modern Russian verbs. This standard interpretation is indeed widespread and can be illustrated by (1–2)[1] and many other examples like it:

---

[1] Hereinafter, all the examples are from Russian National Corpus (www.ruscorpora.ru), unless otherwise specified.

(1) Ru > Engl
   a. *Esli hochesh', ya dazhe nikomu **ne skazhu**, chto byl zdes' i videl vas* [A. Kuprin. Olesya (1898)]
   b. *I **shan't tell** anybody I was here and saw you, if you don't want me to* [A. Kuprin. Olesya, transl. by Stepan Apresyan (1982)]

(2) Ru > Engl
   a. *Konechno, ya nichego ei **ne skazhu**, eto ee ub'et* [A. Rybakov. Deti Arbata (1966–1983)]
   b. *Of course I **won't say** anything, it would kill her* [A. Rybakov. Children of the Arbat, transl. by Harold Shukman (1989)]

(3) Ru > Engl
   a. *– Nu ladno, Toropyzhka! – obidelsya Neznayka. – Poprosish' ty u menya chto-nibud', ya tozhe tebe **ne dam**.* [N. Nosov. Priklyucheniya Neznayki i ego druzei (1953–1954)]
   b. *"All right, Swifty," said Dunno sulkily, "the next time you ask me for something I **won't give** it to you either."* [N. Nosov. The Adventures of Dunno and his Friends, transl. by Margaret Wettlin (1980)]

Here, in all the cases, the event of not-saying or not-giving P is clearly postponed to some moment in the future: the basic idea behind the contexts like (1–3) is, roughly, 'whenever, at some moment M, the speaker is eventually asked about P, the reaction will not follow'. That being the case, there is an apparent temporal break between M and the moment of utterance TU: at TU, the dilemma of doing / not doing P is not (yet) at issue.

Unlike these standard contexts with future-tense reference, the contexts we are concerned with here refer to a present situation. The refusal to perform P relates to the very moment of utterance, and, more than that, the action of refusal (according to what is expected from a performative expression) is equivalent to the utterance of the type 'I won't [tell/give/let]'. Indeed, uttering something like *ne skazhu* or *ne dam*, the speaker – in doing so – refuses to tell or give P **from now on**, and not at some moment in the future, as standard contexts like (1–3) suggest.

Obviously, a special context is needed for this "present performative" interpretation. Usually, it is obtained when a negated first-person verb functions as a short "conversational turn" of the speaker triggered by a previous request from the addressee. It is thus an immediate (negative) reaction to some proposal; its meaning approaches closely what can be labelled "discourse formulae", i.e. a language-specific set of largely non-compositional positive or negative reactions used in conversation (see (Rakhilina et al. 2021) and (Bychkova, Rakhilina forth.) for more detail).

The present tense reference, in principle, is typical for performative expressions, though cross-linguistically the situation can be more heterogeneous, and various verbal forms (not necessarily morphologically present) can occur in this function: sometimes past or perfect, sometimes future (cf. (De Wit et al. 2018), (Fortuin 2019) for a more detailed overview). In Russian (and Slavic in general)

imperfective present tense (like *klyanus'* 'I swear' or *obeshchayu* 'I promise') remaining the main grammatical device for performatives, perfective present is also possible; the distribution is not quite clear and seems to be largely lexical. The most frequent case is the verb *prosit'* 'to ask, to beg' which expresses performative semantics equally well in the imperfective (*proshu*) and perfective (*poproshu*) forms; the latter can be exemplified by (4); nota bene the present progressive form in the English counterpart.

(4)
    a. *Gospodin prem'er-ministr, ya **poproshu** vas tut ostanovit'sya.*
    b. *Prime Minister, I'm **going to ask** you to stop right there.*

In (4), *poproshu* (lit., 'I will ask') is both perfective and performative (it represents the act of asking as such); *proshu* is also possible here, with some subtle semantic or pragmatic difference.[2]

4. Perfective performatives

Our three constructions (*ne skazhu*, *ne dam* and *ne pushchu*) can be seen, in a sense, as complementing the stock of perfective performatives in Russian. As performatives, they have a present-tense reference (witness many of their translational equivalents, which will be considered below in more detail); however, they do not correspond exactly to most frequently discussed types of non-standard perfective presents, because they don't belong to any type of habitual uses (mentioned earlier). As we suggested previously, the most plausible aspectual construal in these cases would be the prospective one.

Let us now consider more examples of these constructions, with the translations from RNC parallel corpora.

4.1. Telling

Recall that the essential meaning of *ne skazhu* can be rendered as 'in reply to your request to tell you P, I let you know that I refuse to tell you P'. Cf. (5) –(6):

(5) Ru > Engl
    – *Irina, a gde vy hranite svou poslednuiu medal' iz nastoyashchego zolota?*
    – ***Ne skazhu!*** [regional press, 2002]
    '– Irina, where do you keep you medal of real gold? – **I won't tell you!**'

(6) Ru > Engl
    a. – *U menya est' eshche odna ideya. – Kakaya? – **Ne skazhu**. – Pochemu?*
    – *Pob'ete. – My tebya i tak pob'em.* [A. Strugackii, B. Strugackii. Ponedel'nik nachinaetsa v subbotu (1965)]

---

[2] Aspectual and illocutive properties of this verb in Russian and other Slavic languages (some of them display the similar type of aspectual variation) have been widely discussed: cf., for example, (Slavkova 2014) and (Biasio 2019), as well as, in a more general perspective, (Wiemer 2014, Dickey 2016, Biasio 2021a, 2021b).

b. "I have one more idea." "What?" "**I won't say**." "Why?" "You'll beat me." "We'll beat you if you don't." [A. Strugatsky, B. Strugatsky. Monday Begins on Saturday, transl. by Leonid Renen (1977)]

In (5)–(6), *I won't tell* actually means 'I *tell* you right now that I refuse to tell what you require'. Other examples of this kind are as follows (let us pay special attention to the different translational strategies observed in different languages and in different translators).

(7) Ru > Engl
a. – *A vy mne ne skazhete, otkuda vy uznali pro listki i pro moi mysli? –* ***Ne skazhu****, – suho otvetil Azazello. – No vy chto-nibud' znaete o nem?* [M. Bulgakov. Master i Margarita (1929–1940)]
b. – *And you won't tell me how you found out about the pages and about my thoughts? –* ***No, I won't****, – Azazello replied drily. – But do you know anything about him?* [M. Bulgakov. Master and Margarita, transl by Richard Pevear and Larissa Volokhonsky (1979)]

The English translation uses here a form with future-tense reference (though mitigated by a strong modal overtone of English *will*); a more transparent strategy is observed in the Italian translation of the same passage (8), where a genuine present *dico* appears.

(8) Ru > It
b. – *E non vuol dirmi com'è venuto a sapere dei foglietti e di quello che penso? – No,* ***non lo dico****, – rispose asciutto Azazello. – Lei, però, sa qualcosa di lui?* [transl. by Vera Dridso (1967)]

Other Italian examples demonstrate that this is not an accidental correspondence: Italians seem to regularly use the present tense of the speech verbs in such contexts, both in the translations and in the original texts; (9) is taken from an Italian novel and its translation into Russian.

(9) It > Ru
a. *Qual è? –* ***Non te lo dico****. – Ora me lo devi dire.* [Niccolò Ammaniti. Io non ho paura (2001)]
b. *I kto èto? –* ***Ne skazhu****. – Pochemu? Tak nechestno.* [transl. by Valerij Nikolaev (2005)]

In fact, English can also use here constructions with a more direct present-tense reference, as, for example, *I am not telling you* attested in many similar passages in original English texts, as in (10):

(10) *"Did he break your heart and send you running?" <...> She smiled. "Well,* ***I'm not telling you*** *anything."* [Michael Connelly. City of bones (2002)]

In (10), the present progressive is used as exactly the same conversation turn as in previous examples: the speaker informs (at the moment of utterance) that she refuses to give any clarifications asked.

A SHORT NOTE ABOUT OTHER USES OF RUSSIAN *NE SKAZHU*.

A further complication here is related to the fact that Russian *ne skazhu* can (and frequently does) represent another type of constructions with a present-tense reference, where perfective present is used in a more widespread (and, in a sense, more common) potential meaning 'I cannot tell'. The whole expression, however, is somewhat less compositional and amounts to a marker of subjective epistemic evaluation (≈ 'I'm not that sure; I don't know for certain'), as in (11) or (12):

(11) Ru

Ne skazhu, *chto ya horoshii fizionomist, no lyudei chuvstvuyu* [regional press, 2003]
'**I wouldn't say** I am a good physiognomist, but I get a feel for people'

(12) Ru >En

a. *Ya **ne skazhu**, chtoby francuzskie knizhki byli i talantlivy, i umny, i blagorodny* [A. Chehov. Skuchnaya istoriya (1889)]
b. *I **don't say** the French books have talent, cleverness, and a good tone* [A. Chekhov. A Dreary Story, transl. by Constance Garnett (1930)].

This type of (present) uses is not to be confused with what we consider as performative-like perfective presents, even when they occur – like performative expressions usually do – as isolated lines in a dialogue. However, the translation usually reveals their modal construal ('I cannot tell because I don't know / I'm not sure'), as in (13):

(13) Ru > En

a. *Skazala pered samym ukhodom, chto proshchaetsya nadolgo, potomu chto uezzhaet s muzhem za granicu. Olya sprosila dovol'no ravnodushno, kuda. Galya usmehnulas: – Predstav', na Blizhnii Vostok. Konkretno **ne skazhu**.* [L. Ulickaia. Zelenyi shater (2011)]
b. *Right before she left, she said she was saying good-bye for a long time, because she and her husband were going abroad. Olga, with seeming indifference, asked her where. Galya grinned. "Just imagine, we're going to the Middle East. **I can't say** where exactly."* [L. Ulitskaia. The Big Green Tent, transl. by Bela Shayevich (2014)]

4.2. Giving

A similar behaviour is observed with *ne dam* (lit. 'I won't give). In fact, its more precise meaning is something like 'in reply to your request to give you Z, I let you know that I refuse to give you Z'. A typical example is (14).

(14) Ru

*– Dai mne eshche drovishek – u nas zhe ih mnogo. – Net, – skazal Ezhik, – **ne dam**. V dome i tak teplo.* [Sergei Kozlov (1981)]
'– Give me some more firewood, we've got it a lot, right? – No, – said the Hedgehog, – **I won't**. It is warm enough in the house.'

Again, we can provide two translations from parallel corpora (one from Russian into German, another, into Italian), where the inherent present-tense reference of *ne dam* situation is rendered by grammatical means. Let it be a passage from the famous novel "Oblomov", where the main protagonist discusses with his servant the possibility of lending his tail-coat: the servant, with the help of *ne dam*, refuses to part with it.

(15) Ru > Ge, Ru > It
a. – *Day, Zahar, frak, ne upryam'sya!* – ***Ne dam****! ― kholodno otvechal Zahar. – Pust' prezhde oni prinesut nazad zhilet da nashu rubashku: pyatyi mesyac gostit tam.* [I. Goncharov. Oblomov (1849–1858)]
b. – *Sachar, gib den Frack her, sei nicht eigensinnig!* – ***Ich gebe ihn nicht her****! – sagte Sachar kühl, – er soll uns zuerst unsere Weste und unser Hemd zurückgeben, die sind jetzt schon fünf Monate bei ihm auf Besuch.* [transl. by Clara Brauner (1960)]
c. – *Su, Zachar, porta il frac, non essere cocciuto!* – ***Non glielo do****! – rispose freddamente Zachar. – Prima deve riportarci il nostro panciotto e la nostra camicia: sono cinque mesi che se li tiene.* [transl. by Argia Michettoni]

The translations show that English retains future *will*-forms more frequently (probably due to their modal side-values), whereas other languages (not only Italian and German, but also French) prefer other strategies:

(16) Ru > En, Ru > Fr
a. – *O net!* – *voskliknula Margarita, porazhaya prohodyashchih, – soglasna na vse, soglasna prodelat' etu komediyu s natiraniem maz'yu, soglasna idti k chertu na kulichki.* ***Ne otdam!*** [M. Bulgakov. Master i Margarita (1929–1940)]
b. – *Oh, no!* – *exclaimed Margarita, shocking the passers-by. – I agree to everything, I agree to perform this comedy of rubbing in the ointment, agree to go to the devil and beyond!* ***I won't give it back!*** [transl. by Richard Pevear and Larissa Volokhonsky (1979)]
c. – *Oh non! s'exclama Marguerite, d'une voix qui fit se retourner les passants. – Je suis d'accord pour tout, je suis d'accord pour me barbouiller de crème et toute cette comédie, je suis d'accord pour aller à tous les diables!* ***Je garde la boîte****!* [transl. by Claude Ligny (1968)]

4.3. Letting

Finally, our third example of non-trivial performative is *ne pushchu* (lit. 'I won't let <you go>') meaning roughly ≈ 'in reply to your request to let you leave the place L, I let you know that I do not authorize you to leave L and obstruct it physically [e.g. closing the way out, retaining physically, etc.].' The translated examples from "Master and Margarita" (here, again, English and Italian display two different strategies, with futurate *will*-form and genuine present tense) reproduce what we have seen previously, cf. (17).

(17) Ru > En, Ru > It

a. *Ruchku dveri snaruzhi v eto vremya krutili i dergali, i slyshno bylo, kak kur'ersha za dveryami otchayanno krichala: – Nel'zya!* ***Ne pushchu!*** *Khot' zarezh'te! Zasedanie!* [M. Bulgakov. Master I Margarita (1929–1940)]

b. *The door handle was all the while being turned and pulled from outside, and the messenger girl could be heard through the door crying desperately: – Impossible!* ***I won't let you!*** *Cut me to pieces! It's a meeting!* [transl. by Richard Pevear and Larissa Volokhonsky (1979)]

c. *Nel frattempo la maniglia della porta veniva girata e scrollata dall'esterno, e si sentiva l'inserviente gridare istericamente dietro la porta: – Non si può!* ***Non vi lascio passare!*** *Anche se mi ammazzate! Sono in riunione!* [transl. by Vera Dridso (1967)]

However, even English translations can also be sensitive to this inherent present semantics, witness the following passage from Gogol's famous novel describing an authoritarian host, cf. (18), where all the available translations (viz., English, German and Italian ones) concur:

(18) Ru > En, Ru > Ge, Ru > It

a. *— Net, ty uzh, pozhaluysta, menya-to otpusti, — govoril belokuryy, — mne nuzhno domoy. — Pustyaki, pustyaki, brat,* ***ne pushchu*** [N. Gogol'. Mertvye dushi (1835–1852)]

b. *"No, no, put in the flaxen-haired man," you must excuse me, for I must be off home." – "Rubbish, rubbish,* ***I am not going to excuse you****".* [N. Gogol. Dead Souls, transl. by D. J. Hogarth (1931)]

c. *„Nein, verzichte bitte auf mich", sagte der Blonde, „ich muß nach Hause." – „Unsinn, Unsinn, Bruder,* ***ich laß dich nicht fort****."* [N. Gogol. Die toten Seelen, transl. by Michael Pfeiffer (1978)]

d. *«No, a questo punto, per favore, lasciami andare» disse il biondo, «devo tornare a casa.» «Sciocchezze, sciocchezze, fratello,* ***non ti lascio****.»* [N. Gogol. Anime morte, transl. by Paolo Nori]

Here, clearly, the utterance expressing the intention of non-allowing the guest to go is accompanied by some actions performed at the moment of utterance, and the whole situation is undoubtedly present. It should be noted the repeated use of progressive form in English, which is also in keeping with the aspectual semantics of the situation.

5. Discussion

We have demonstrated that in Russian the verbal form of what is morphologically a perfective present is capable to develop, for the avoidance of "PP paradox", a less trivial (and probably still unnoticed) meaning of **performative-like prospective present.** Its main semantics relates to an immediate start of the action described (somewhat in between of present and future). Obligatory components of this type of context are first person and negation. Diachronically,

these uses rather seem (like habitual ones) to be remnants of some more archaic state when perfectivity was not fully grammaticalized and perfective present was more present than perfective (cf. [Dickey 2016, Biasio 2021b]).

The view of performative-like contexts as archaic is corroborated by two essential features of these context: negation and first person. It is well known that negation is typical for "relict contexts" preserving the most frequent old uses [Bybee et al. 1994]; in our case, however, these relict contexts could be preserved only with first person forms responsible for a performative-like effect of the construction. Still, what we have here is probably not a prototypical performative speech act, but the moment of saying *ne dam* 'I won't give' or *ne pushchu* 'I won't let' is similar to the moment of saying 'I promise' (especially as concerns the pragmatic effects) – and is not exactly like "ordinary" uses of negated perfective present with a clear future tense reference.

The comparison of the same constructions in more ordinary contexts (where they have a future time reference) would be compelling. Cf., for example, the following passage with *ne skazhu*, where it is not an isolated conversational turn, but a part of the longer narrative:

(19) Ru > Ge
a. *Ya nikogda nikomu **ne skazhu** etogo, no, Bozhe moy, chto zhe mne delat', ezheli ya nichego ne lyublyu, kak tol'ko slavu, lyubov' lyudskuyu?* [L. Tolstoi, Voyna i mir (1867–1869)]
b. *Ich **werde** das niemals einem Menschen **sagen**, aber, mein Gott, was soll ich nur tun, wenn ich nun einmal nichts so sehr liebe wie den Ruhm und die Anerkennung der Menschen?* [L. Tolstoi. Krieg und Frieden, transl. by Hermann Röhl (1922)]

The action of (not) saying is clearly detached from the moment of speech and rendered as future even in German, where *werden*-future is used relatively rarely. Cf. a performative-like context (not detached from the moment of utterance) in (20):

(20) Ru > Ge
a. *Dolg khotel otdat', dolg chesti, a komu – **ne skazhu**.* [F. Dostoevskii, Brat'ya Karamazovy (1878)]
b. *Ich wollte eine Schuld zurückzahlen, eine Ehrenschuld wollte ich zurückzahlen. An wen, das **sage** ich **nicht**.* [F. Dostojewski. Die Brüder Karamasow, transl. by Hermann Röhl (1923)]

The specific type of use considered in our paper is characteristic for several predicates which seem to occur in appropriate contexts with a particularly high frequency. Semantically, these predicates represent a scenario with two participants where the one requires from the other some (profitable or necessary) action, and the other resists and refuses to perform it. This type of interpersonal interaction with a prominent component of affectedness / concernment is found primarily among predicates subcategorized for recipients and addressees (as 'give' or 'tell'), but the predicates involving the semantic component of permission ('let') also play an

important role. Notice that the last group includes the core performative (*ne*) *razreshat'* '(not) to allow' as well, which, in Russian, behaves in a fully canonical way in the sense that its performative uses require a present imperfective (and not perfective) form: – *razreshi!* 'allow!' – *ne razreshayu!* 'I don't / won't allow!'. Cf. a less trivial (though obsolete in Modern Russian) *pozvolyat'* 'allow' which permits the aspectual variation: – *pozvol'!* 'allow!' – *ne pozvolyayu!*[IMPF] / *ne pozvolyu!*[PF] 'I don't / won't allow!'. Interestingly, the present perfective affirmative form is still standard: *pozvolyu* can have only a future-time reference and cannot describe an ongoing present. But instead, the negative form *ne pozvolyu!* behaves exactly like *ne pushchu!* 'I won't let' considered in the previous sections, blurring the distinction between an actual and a future eventuality.

Another fact which deserves special attention is that the one-place predicate *idti* 'go' demonstrates some affinities with the two-place predicates subcategorized for beneficiary: – *idi!* 'go!' – *ne poydu!*[PF] 'I won't go / I'm not going!'. Like the latter example, it demonstrates a kind of performativity: its semantics includes a declaration (the refusal to move) – and at the same time the action of non-moving. Some translational equivalents (from Italian or English) found in the parallel corpus corroborate this construal, because the corresponding predicates are used in the form of present tense there, as in (21) and (22).

> (21) En > Ru
> a. *"Mommy says you have to go to school,"* Izzy says, head-butting my shoulder. **"I'm not going** to school.' That's it: that's how it starts*. [Lauren Oliver. Before I fall (2010)]
> b. – *Mama govorit, tebe pora v shkolu. Ona tolkaet menya golovoi v plecho. –* **Ya ne poydu** *v shkolu. Vot kak vse nachinaetsya*. [L. Oliver. Prezhde chem ya upadu, transl. by A. Kilanova (2017)]

> (22) It > Ru
> a. *Maria ha incrociato le braccia. "Io* **non ci vado**.*" Perché?* [Niccolò Ammaniti. Io non ho paura (2001)]
> b. *Maria zakinula nogu na nogu*[3]: *– Ya* **ne poydu**. *– Pochemu?* [N. Ammaniti. Ya ne boyus', transl. by Valerij Nikolaev (2005)]

So, the phenomenon described above is somewhat larger than just three verbal units, and definitely needs further study.

6. Concluding remarks

To sum up, as we hope to have shown so far, "performative-like prospective presents" in Russian is another way of resolving the widely discussed present perfective paradox: the corresponding verbal form remains (still) present, but is not

---

[3] Strange enough, the Russian translator has rendered *ha incrociato le braccia* 'crossed her arms' as *zakinula nogu na nogu* 'crossed her legs'.

(fully) perfective, which can be considered as an aspectual archaism for East Slavic as a whole.[4]

A terminological note needs to be made at this point – namely, regarding the term 'performative', which is used here in a somewhat noncanonical way. Indeed, performativity has always been considered to be a property exclusive to verbs of speech (cf., for example, Wierzbicka 1987) and is sometimes viewed as their non-trivial semantic feature (cf. Apresyan 1988). Here, instead, we deal with predicates of completely different semantics.

The theoretical framework which could provide a suitable ground for our approach is Construction Grammar (see Fillmore et al. 1988, Goldberg 1995). Following this theory, one could regard performativity as a property of constructions featuring first-person verb forms and exhibiting special semantic-pragmatic characteristics. Then, like any other construction, the performative one should have core instances that combine all the features of the construction, and peripherical cases that are coerced into a performative interpretation due to the fact that they share several features with the prototype. From this perspective, utterances containing verbs of speech (such as 'promise', 'declare', etc., usually in non-negative forms) should be regarded as the core of performativity. As for the non-trivial marking of tense in negative first-person forms that we discussed here, it can serve as a kind of test for identifying non-central extensions of the performative construction.

Generally, the construction approach has repeatedly been applied to verbal grammatical categories. For performativity, this has not been the case. It is appropriate here, however, to recall an insightful remark by Yu. Apresyan (1988), which anticipated the theory by several years: "It must be emphasized that the performative formula takes precedence over a performative verb."

---

[4] The common Slavic origin of this type of uses finds support in the fact that the most archaic Slavic aspectual systems still allow for perfective present forms in performative contexts of "stiff denial". Notably, this is widely attested for Slovenian (and is characteristic for Slovenian performatives in general, cf. (Derganc 2012)). On the contrary, Bulgarian and Macedonian tend to use here present imperfective forms (which is more consistent semantically and more innovative morphologically), as *ne te puskam* or *ne davam* in Bulgarian, *ne davam* or *ne kažuvam* in Macedonian, etc. (we are grateful to Mladen Uhlik, Elena Ivanova and Eleni Bužarovska for personal consultations). Among South Slavic languages, Macedonian seems to have the most widespread uses of imperfective present in quasi-performative contexts. Eleni Bužarovska (2010) points to a strong morphosyntactic parallelism between Macenonian and Modern Greek in this respect: Macedonian present imperfective constructions like *ne ti kažuvam* 'I am not telling / won't tell you' literally correspond to Greek δεν σου λέω.

REFERENCES

Aikhenvald, Alexandra Y. & Robert M. W. Dixon. 1998. Dependencies between grammatical systems. *Language* 74 (1). 56–80 (also in: Aikhenvald, Alexandra Y. & Robert M.W. Dixon. 2011. Dependencies between Grammatical Systems. *Language at Large: Essays on Syntax and Semantics*. Leiden: Brill. 170–204).

Apresyan, Yurii D. 1988. Glagoly momental'nogo deistviya i performativy v russkom yazyke (Verbs of Instantaneous Action and Performatives in Russian). In Yurii D. Apresyan. 1995. *Izbrannye trudy. T. 2: Integral'noe opisanie yazyka i sistemnaya leksikog*rafiya. Moscow: Yazyki russkoi kul'tury. 199–241. (In Russ.).

Biasio, Marco. 2019. The etiquette of aspect. How and why *prositi* stopped worrying and entered a pair. *Lingue e Linguaggi* 31. 191–218.

Biasio, Marco. 2021a. Dictum erit factum? Vidovremennye svoistva performativov v sopostavitel'nom aspekte (na materiale russkogo i serbskokhorvatskogo yazykov) (Tempo-Aspectual Properties of Performative Verbs from a Crosslinguistic Perspective (based on Russian and Serbo-Croatian Languages)). Università degli Studi di Padova PhD. (In Russ.).

Biasio, Marco. 2021b. Slavic performatives were originally marked for general aspect. In Ettorio Gherbezza, Vesselina Laskova & Anna Maria Perissutti (eds.), *Le lingue slave: Sviluppi teorici e prospettive applicative. Atti del VIII incontro di linguistica slava (Udine, 10–12 settembre 2020)*. Roma: Aracne. 47–62.

Buras, Maria M. 2022. *Lingvisty, prishedshie s kholoda* (Linguists who came in from the cold). Moscow: AST, Redaktsiya Eleny Shubinoi. (In Russ.).

Buzarovska, Eleni. 2010. Izrazuvanie na segashnost vo grchkiot yazik. In Zuzana Topolinjska (ed.), *Segashnost kako lingvistichki poim: Gramatichkite sredstva za izrazuvanie segashnost vo slovenskite i balkanskite zemii*. Skopje: MANU. 137–150. (In Maced.)

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bychkova, Polina & Ekaterina Rakhilina. forthcoming. Towards pragmatic construction typology: The case of discourse formulae. In Alessandra Barotto & Simone Mattiola (eds.), *Discourse Phenomena in Typological Perspective*. Amsterdam: John Benjamins.

De Wit, Astrid, Frank Brisard & Michael Meeuwis. 2018. The epistemic import of aspectual constructions: The case of performatives. *Language and Cognition* 10 (2). 234–265.

De Wit, Astrid. 2017. *The Present Perfective Paradox across Languages*. Oxford: Oxford University Press.

Derganc, Alexandra. 2012. Performativnye glagoly v slovenskom yazyke (Performative verbs in Slovenian). *Vestnik Moskovskogo universiteta. Ser. 9. Filologiya* 6. 228–236. (In Russ.).

Dickey, Stephen M. 2016. The Aspectual Development of Performatives in Slavic. *Zeitschrift für Slavische Philologie* 71 (2). 249–304.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64 (3). 501–538.

Fortuin, Egbert. 2019. Universality and language-dependency of tense and aspect: Performatives from a crosslinguistic perspective. *Linguistic Typology* 23 (1). 1–58.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago & London: The University of Chicago Press.

Khrakovskii, Viktor S. 1990. Vzaimodeistvie grammaticheskikh kategorii glagola: opyt analiza (Interaction of grammatical categories of the verb: A tentative analysis). *Voprosy yazykoznaniya* 5. 18–36. (In Russ.).

Khrakovskii, Viktor S. 1996. Grammaticheskie kategorii glagola (opyt teorii vzaimodeistviya) (Grammatical Categories of the Verb (Towards a Theory of Category Interaction)). In Aleksandr V. Bondarko (ed.), *Mezhkategorial'nye svyazi v grammatike*. Saint Petersburg: D. Bulanin. 22–42. (In Russ.).

Khrakovskii, Viktor S. & Andrei L. Mal'chukov. 2016. Vzaimodeistvie i ierarkhiya grammaticheskikh kategorii glagola: vvedenie v temu i tipologicheskaya anketa (Interaction and hierarchy of verbal categories: Research program and a typological questionnaire). *Voprosy yazykoznaniya* 6. 51–83. (In Russ.).

Korn, Agnes & Irina Nevskaya (eds.). 2017. *Prospective and Proximative in Turkic, Iranian and beyond* (Iran – Turan 18). Wiesbaden: Reichert.

Kozlov, Aleksei A. 2021. K semanticheskoi tipologii prospektiva (Towards a semantic typology of prospective aspect). *Voprosy yazykoznaniya* 2. 28–52. (In Russ.).

Malchukov, Andrej. 2009. Incompatible categories: Resolving the "present perfective paradox." In Lotte Hogeweg, Helen de Hoop & Andrej Malchukov (eds.), *Cross-linguistic semantics of tense, aspect, and modality*, 13–32. Amsterdam: John Benjamins.

Malchukov, Andrej. 2011. Interaction of verbal categories: Resolution of infelicitous grammeme combinations. *Linguistics* 49 (1). 229–282.

Malchukov, Andrej L. 2019. Interaction of Verbal categories in a Typological perspective. *Gengo Kenkyū* 156. 1–24.

Mel'chuk, Igor A. 1961. O nekotorykh tipakh yazykovykh znachenii (On some types of linguistic meanings). *O tochnykh metodakh issledovaniya yazyka*. M.: MGU. 33–39. (In Russ.).

Mel'chuk, Igor A. 1968. Stroenie yazykovykh znakov i vozmozhnye formal'no-smyslovye otnosheniya mezhdu nimi (The structure of linguistic signs and possible formal-semantic relations between them). *Izvestiya Akademii nauk SSSR. Seriya literatury i yazyka* 27 (5). 426–438. (In Russ.).

Mel'chuk Igor A. 1972. O suppletivizme (On Suppletion). In Sebastian K. Shaumyan (ed.), *Problemy strukturnoi lingvistiki 1971*. Moscow: Nauka. 396–438. (In Russ.).

Mel'chuk, Igor A. 1975/1997. Opyt razrabotki fragmenta sistemy ponyatii i terminov dlya morfologii (k formalizatsii yazyka lingvistiki) (Experiments on elaborating a fragment of the system of notions and terms for morphology (towards the formalization of the language of linguistics)). *Semiotika i informatika* 35. 15–58. (In Russ.).

Mel'chuk, Igor A. 1995. *Russkii yazyk v modeli «Smysl ↔ Tekst» (The Russian Language in the Meaning ↔ Text Perspective)*. Moscow-Vienna: Shkola "Yazyki russkoi kul'tury". (In Russ.).

Mel'chuk, Igor A. 1997–2006. *Kurs obshchei morfologii (Course in General Morphology)*. Vol. I–V. Moscow: Yazyki russkoi kul'tury. (In Russ.).

Mel'chuk, Igor A. 2004. Actants in semantics and syntax. *Linguistics* 42 (1), 42 (2).

Michaelis, Laura A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15 (1). 1–67.

Rakhilina, Ekaterina V., Polina A. Bychkova & Svetlana Yu. Zhukova. 2021. Rechevye akty kak lingvisticheskaya kategoriya: diskursivnye formuly (Speech acts as a linguistic category: The case of discourse formulae). *Voprosy yazykoznaniya* 2. 7–27. (In Russ.).

Shluinskii, Andrei B. 2012. Faktativ i smezhnye kategorii: opyt tipologii (The factative and related categories: A tentative typology). *Acta Linguistica Petropolitana* 8 (2). 950–996. (In Russ.).

Slavkova, Svetlana. 2014. Performativnoe upotreblenie glagolov *prosit' / poprosit'* i *molya / pomolya* v russkom i bolgarskom yazykakh: pragmaticheskaya rol' vida i vremeni (The performative use of the verbs *prosit' / poprosit'* and *molya / pomolya* in Russian and

Bulgarian: The pragmatic role of aspect and tense). *Scando-Slavica* 60 (2). 231–252. (In Russ.).

Voeikova, Maria D. 1996. Kategorial'nye priznaki performativnykh vyskazyvanii v russkom yazyke (The categorial features of performative utterances in Russian). In Aleksandr V. Bondarko (ed.), *Mezhkategorial'nye svyazi v grammatike,* 153–167. Saint Petersburg: D. Bulanin. (In Russ.).

Wiemer, Bjorn. 2014. Upotreblenie sovershennogo vida v performativnom nastoyashchem (The use of the perfect form in the performative present). *Acta Linguistica Petropolitana* X (3). 90–113. (In Russ.).

Wierzbicka, Anna. 1987. *English Speech Act Verbs*: *A Semantic Dictionary*. Sydney: Academic Press.

Zaliznyak, Andrei A. 1990. Ob odnom upotreblenii prezensa sovershennogo vida («prezens naprasnogo ozhidaniya») (On one type of present perfect use ("present of vain expectation")). In Zygmunt Saloni (ed.), *Metody formalne w opisie języków słowiańskich,* 109–114. Białystok: W-wo Univ. Warszawskiego. (In Russ.).

Bionotes:
Vladimir PLUNGIAN is Doctor Habil., full member of the Russian Academy of Sciences, Head of the Typology department at the Institute of Linguistics of the Russian Academy of Sciences, and Head of the Department of corpus linguistics and linguistic poetics of Vinogradov Russian Language Institute. He is also Professor of the Department of theoretical and applied linguistics at the Philological faculty of Moscow State University. His research interests embrace linguistic typology, theory of grammar, corpus linguistics, Russian morphology in synchrony and diachrony, African studies, and poetics.
e-mail: plungian@gmail.com
https://orcid.org/0000-0002-2393-1399

Ekaterina RAKHILINA is Doctor Habil., Professor and Head of the School of Linguistics at the Faculty of Humanities at the National Research University "Higher School of Economics". She is also a leading researcher at Vinogradov Russian Language Institute. Her research interests include semantics, lexical typology, corpus linguistics, construction grammar, history of Russian, lexicology, and discourse analysis.
e-mail: rakhilina@gmail.com
https://orcid.org/0000-0002-7126-0905

Tatiana REZNIKOVA holds a Ph.D. in Philology and is Associate Professor of the School of Linguistics at the Faculty of Humanities at the National Research University "Higher School of Economics". Her research interests include lexical typology, semantics, grammaticalization and corpus linguistics.
e-mail: tanja.reznikova@gmail.com
https://orcid.org/0000-0003-0976-224X

Сведения об авторах:

Владимир Александрович ПЛУНГЯН – доктор филологических наук, академик РАН, заведующий сектором типологии Института языкознания РАН и отделом корпусной лингвистики и лингвистической поэтики Института русского языка им. В.В. Виноградова РАН, профессор Отделения теоретической и прикладной лингвистики филологического факультета МГУ им. М.В. Ломоносова. В сферу его научных интересов входят лингвистическая типология, теория грамматики, корпусная лингвистика, грамматическая система русского языка в синхронии и диахронии, африканистика, поэтика.
e-mail: plungian@gmail.com
https://orcid.org/0000-0002-2393-1399

Екатерина Владимировна РАХИЛИНА – доктор филологических наук, профессор, руководитель Школы лингвистики факультета гуманитарных наук Научно-исследовательского университета «Высшая школа экономики», ведущий научный сотрудник Института русского языка им. В.В. Виноградова РАН. Ее научные интересы включают семантику, лексическую типологию, корпусную лингвистику, грамматику конструкций, историю русского языка, лексикологию, анализ дискурса.
e-mail: rakhilina@gmail.com
https://orcid.org/0000-0002-7126-0905

Татьяна Исидоровна РЕЗНИКОВА – кандидат филологических наук, доцент Школы лингвистики факультета гуманитарных наук Научно-исследовательского университета «Высшая школа экономики». В сферу ее научных интересов входят лексическая типология, семантика, теория грамматикализации, корпусная лингвистика.
e-mail: tanja.reznikova@gmail.com
https://orcid.org/0000-0003-0976-224X

Research article / Научная статья

Meaning and function of Spanish formulemes and pragmatemes vs. illocutionary verbs

María Auxiliadora BARRIOS RODRÍGUEZ ✉

Universidad Complutense de Madrid
✉auxibarrios@filol.ucm.es

Abstract
From the perspective of Pragmatics, some scholars claim that the taxonomy of illocutionary acts should be revised. The aim of this paper is to propose such a review by means of a research field in which Lexicography and Pragmatics overlap. As we attempt to prove, formulemes offer the advantage of being a narrower field of study than free utterances. Formulemes (*Have a nice day!*) have been defined within the Meaning Text-Theory as a type of cliché and Pragmatemes (*Happy birthday!*) as a type of formuleme more restricted by the extralinguistic situation (someone's birthday). E-dictionaries require a formal method to express both the meaning and the function of formulemes, yet this lexicographic development may well elicit problems. Within Meaning-Text Theory pragmatemes have been formalized to date by Lexical Functions. However, we have observed that this tool is unsatisfactory for didactic purposes. Therefore, in the Spanish e-dictionary *Diretes*, we have attached each formuleme to one illocutionary verb that we call "Pragmatic Function" (such as *to wish* and *to congratulate* for the examples above). In order to identify whether a formalization by means of Pragmatic Functions could be both possible and successful, we have formalized more than two hundred formulemes (sixty of them pragmatemes). Although the project is in progress, up to now any kind of formuleme (being or not pragmateme) was successfully analyzed by means of thirty Pragmatic Functions created *ad hoc*. Pragmatic Functions could be useful not only for the formalization of formulemes and pragmatemes when teaching Spanish, but also to revise the list of illocutionary verbs from the perspective of Phraseology and Lexicography.
Keywords: *E-lexicography, formulemes, illocutionary verbs, lexical functions, pragmatemes, pragmatic functions*

Значение и функции испанских формулем и прагматем в сравнении с иллокутивными глаголами

Мария Ауксилиадора БАРРИОС РОДРИГЕС ✉

Мадридский университет Комплутенсе
✉auxibarrios@filol.ucm.es

Аннотация

С точки зрения прагматики некоторые ученые утверждают, что таксономию иллокутивных актов следует пересмотреть. Цель статьи – определить исследовательскую область пересечения лексикографии и прагматики, что может способствовать уточнению списка иллокутивных глаголов. Мы пытаемся доказать, что преимущество формулем как объектов исследования в том, что они являются более узкой областью исследования, нежели свободные высказывания. В теории «Смысл – Текст» формулемы (*Have a nice day!*) определяются как разновидность клише, а прагматемы (*Happy birthday!*) – как тип формулем, более ограниченных экстралингвистической ситуацией (чей-л. день рождения). Для электронных словарей требуется формальный метод выражения значения и функций формулем, однако их лексикографическая обработка может вызвать проблемы. На сегодняшний день в рамках теории «Смысл – Текст» прагматемы подвергались формализации на основе лексических функций. Однако мы полагаем, что этот инструмент не соответствует нашим дидактическим целям. В связи с этим в испанском электронном словаре *Diretes* мы присоединили к каждой формулеме один иллокутивный глагол, который мы называем «прагматической функцией» (например, *to wish* и *to congratulate* по отношению к вышеприведенным примерам). Чтобы определить, может ли формализация с помощью прагматической функции быть возможной и успешной, мы формализовали более двухсот формулем (их них шестьдесят прагматем). Хотя проект еще не завершен, каждая формулема (независимо от того, относится ли они к прагматемам) была успешно проанализирована с помощью тридцати прагматических функций, созданных *ad hoc*. Как мы пытаемся доказать, прагматические функции могут быть полезны не только для формализации формулем и прагматем при преподавании испанского языка, но и для пересмотра списка иллокутивных глаголов с точки зрения фразеологии и лексикографии.
Ключевые слова: *электронная лексикография, формулемы, иллокутивные глаголы, лексические функции, прагматемы, прагматические функции*

Dedication

Before going into the subject of this paper, I would like to dedicate a few brief words to the person who inspired this research. When I met Igor Mel'čuk for the first time in my life in 2004, he was a famous researcher and professor and I was starting as a PhD student. For nine months I attended the four courses he gave at the University of Montreal, those being Morphology, Lexicology, Semantics and the PhD course.

I found each one of them extremely interesting and truly inspiring with the last one being the most fruitful. This was because each day Professor Mel'čuk started

by asking if we had any questions. Any of the six students would sometimes express doubt and ask questions accordingly, and I usually chose one or two of the many questions I had, and he spent most of the time talking about those questions. At the finalization of the course I felt extremely grateful for those personalized classes from Igor Mel'čuk about Lexical Functions applied to the Spanish Language.

However, probably the most important subject I learned during my stay in Montreal was in regards to being a good teacher. Professor Mel'čuk would start each course giving us a sequenced syllabus, and much to my surprise he constantly delivered, explaining each theme in three hours! He was always available for his students and also answered our emails in less than one hour. He was also free for tutorials at any given time, even if it was for one tutorial each week, as it was in my case.

He was an expert in more than thirty languages and would ask the students questions related to their mother tongue. He proved to know the Spanish grammar better than me. Commonly he accompanied his lessons with frequent jokes demonstrating not only a vast knowledge of many subjects but also a great sense of humour.

Having said all that, if I had to choose only one feature of his academic repertoire, I would probably choose his generosity. To cite two simple examples, during the Meaning-Text Theory Conference in 2009 he drove people from the airport to their residence more than ten times in two days, acting as a taxi driver instead of the creator of the Meaning-Text Theory (MTT) model. On one of those days I saw him in the hall of our residence talking for more than two hours with the young researcher Dina El Kassas. I met her some days later at the airport and asked her about this conversation. She told me that Professor Mel'čuk had attended her presentation at the conference, felt that she was somewhat lost in her research and very generously chose to discuss this with her. Dina shared with me her gratitude for his spontaneous assistance. Sadly, she passed away seven years later. Our memory is with her.

I could continue writing of more fond memories and moving anecdotes of Professor Mel'čuk, in lieu of a research paper dedicated to him, as I think I learned not only by hearing or reading him, but very fundamentally by seeing him. However, thinking about him as a potential reader of this paper, I will change the subject and focus on pragmateme which is a concept I learned from him.

1. Introduction

Formulemes have been defined within the Meaning Text-Theory as a type of cliché with a specific abstract referent, such as a wish, and/or a specific event associated with them (Mel'čuk 2015a): for instance, *Happy birthday!* and *Have a nice day!* are attached to a birthday and to a farewell respectively. Pragmatemes are expressions restricted by the extralinguistic situation (Mel'čuk 1995, 1996, 2008, 2013, 2015b) and can be expressed by means of lexemes, idioms, collocations and clichés: meaning that a collocation such as *wet paint* (on a sign), an idiom such as

hold the line (in a phone call), a lexeme such as *rest!* (a military command) or a cliché such as *no parking* (on a sign) all function as a pragmateme (Mel'čuk 2020: 16–18).

Consequently, pragmatemes are a crosscategorical concept which corresponds to phrasemes restricted by the extralinguistic situation with an utterance value (Ovejas 2021). As Ovejas proposes, a pragmateme is a speech act for which there is, at least, one of the following parameters: medium, space, time or event (Blanco Escoda & Mejri 2018) and the relation between the speakers (Barrios 2020a). Analyzing the four examples in the previous paragraph, we find at least one of those parameters for each pragmateme: the medium (such as a sign), the space (such as a place where it is forbidden to park), the time (such as the end of military standing to attention), the event (such as a telephone call), and the relation between the speakers (such as military hierarchy). According to Barrios & Ovejas (2019a), the concept of pragmateme is a *continuum*. Ovejas (2021) claims that some pragmatemes are essential or more typical and some others are outlying.

As Mel'čuk summarizes through various ideas in a number of his preliminary papers, from the semantic point of view a pragmateme is "a linguistic expression that does not represent logical propositions and therefore cannot be negated or questioned"; and from the syntactic point of view, a pragmateme is a "full utterance equivalent to an independent clause" (2020: 19).

Whilst pragmateme is a crosscategorical concept, formuleme corresponds to only one concept: a subtype of cliché, which is in turn a compositional semantic-lexemic phraseme, and corresponds to daily speech formulas (*see you later*), technical formulas (*to sum up*), commands (*all hands on deck*), and proverbs (*all good things come to an end*) (Mel'čuk 2020: 14). The concept of formuleme is close to that of pragmateme, to the point that *Happy birthday!* and *Have a nice day!* are simultaneously formulemes and pragmatemes: *Happy birthday!* is a formula used on the day of someone's birthday (it contains at least the parameter of time), and *Have a nice day!* is a formula used in a farewell (it contains at least the parameter of event).

The concept of pragmateme can be understood in a narrow sense (Mel'čuk 1995), which includes expressions such as *for rent, drive slow(ly), do not enter, beware of the dog*, etc. (Mel'čuk 2015a: 29); or in a broad sense: "a formuleme is a pragmateme if it is pragmatically constrained" (Mel'čuk 2015a: 29). Most researchers follow the broad sense (Blanco Escoda 2013, 2014, Gader, Olliger & Polguère 2014, Polguère 2016, Barrios 2017, 2020, Barrios & Ovejas 2019). Ovejas (2021) includes under this concept phraseological schemes, such as *fabricado en (X lugar)* (made in (X place)); *tiene la palabra (alguien)* (to give (someone) the floor)[1]; *la paz sea con vosotros* (peace be with you, in a religious ceremony).

---

[1] As far as we know, this expression is used in a slightly different sense in Spanish than in English: in Spanish it is used to introduce someone, such as a speaker, just one second before this person starts to present at a conference. It is always used as a formula in 3rd person: for instance, *el profesor Polguère tiene la palabra*, Professor Polguère has the floor.

We structured this paper in the following way: Section 2 summarizes how we formalize pragmatemes and formulemes; Section 3 focuses on illocutionary acts; Section 4 presents the hypothesis, the data we use and the methodology; Section 5 presents the results obtained; Section 6 exhibits a number of problems that arose in the course of research; finally Section 7 summarizes the conclusions.

2. Formalizing pragmatemes and formulemes

E-dictionaries require a formal method to express both the meaning and the function of pragmatemes and formulemes, however this lexicographic development can elicit problems (Cowie 2011). Meaning-Text Theory (MTT) is a model applied to lexicographic projects in several languages (see, among others, Mel'čuk & Zholkovsky 1984, Iordanskaja & Paperno 1996, Apresjan et al. 2003, Alonso Ramos 2004, Mel'čuk & Polguère 2007, Polguère 2014, 2018, Alipour, Robichaud & L'Homme 2015, Apresjan 2018, Mel'čuk 2018, Barrios & Boguslavsky 2019b, Barrios 2020b).

Within MTT pragmatemes have been formalized to date by Lexical Functions (Polguère 2007, 2016, Fréchon et al. 2012, Blanco 2013). A Lexical Function (LF) is a formal tool which associates a given lexical expression L (such as *to sleep*), which is called the argument or keyword, with a set of lexical expressions, which are called values (such as *deeply, profoundly, like a baby, like a log*). The formalization of this concrete lexical relation is created via a function called Magn (in Latin Magnus) which means 'intense', 'big', as shown in (1.a). A similar technique has been applied to formalize pragmatemes, with the particularity that in this case LFs are created *ad hoc* for each pragmateme, as shown in (1.b) by an example taken from Mel'čuk (2008):

(1) a. Magn(to sleep) = *deeply, profoundly, like a baby, like a log*
     b. [This object was] recently painted: *Fresh paint* [on a sign, to avoid someone touching it]

In Barrios (2020a), we claim that LFs seems to be unsatisfactory for the formalization of a large set of pragmatemes and formulemes. We strive to summarize the arguments we presented in this paper by means of the examples in (1): (1.a) shows a very productive LF, Magn, which covers hundreds of collocations, such as *torrential rain, heavy drinker, confirmed bachelor, crass mistake*, etc.; whilst (1.b) is a compositional expression (*recently painted*) created *ad hoc* for the pragmateme *fresh paint* which demands extra-linguistic information (on a sign, to avoid someone touching it).

Specifically due to the great importance of extra-linguistic features in understanding the function of each pragmateme, we propose the concept of Pragmatic Function, a tool relevant in the case both of LFs and illocutionary verbs. We define this concept as shown in (2):

(2) A Pragmatic Function (PF) is a function that expresses a speech act (such as to thank, to order, to greet, to congratulate, to evaluate, to warn, etc.)

and is associated with a given extra-linguistic situation (such as an encounter, or something freshly painted) which is called the argument, with a set of expressions (pragmatemes, formulemes) which are called values (such as *how do you do, how are you going; wet paint*, *do not touch*) (Barrios 2020a: 24–25).

In such a way we propose formalizing pragmatemes and formulemes by means of PFs, adding some semantic features when necessary via some adverbs, such as *formally* or *colloquially*, as shown in (3); or even changing the referential expression (the extra-linguistic situation) written in brackets, as shown in (4) (Barrios 2020a: 25):

(3) a. To greet formally (greeting encounter) = how do you do?
     b. To greet_colloquially (greeting encounter) = how are you going?
(4) a. To warn (something freshly painted) = wet paint; do not touch[2]
     b. To warn (a house/room freshly painted) = fresh paint

In order to apply PFs to the e-dictionary *Diretes* (Barrios 2020b) with a more sophisticated system and more explicative power than in (3) and (4), we created a new table in the relational database of the dictionary. Figure 1 shows various columns via which we describe the meaning and the extra-linguistic information related to pragmatemes and formulemes.

The first column shows the pragmateme or formuleme; columns 2–4 present the lexical anchorage (words to which this pragmateme or formuleme would be attached in the dictionary); in 5 we note if it is a pragmateme (because it is a crosscategorial concept) and in 6 if it is an answer to any preliminary question (for a similar reason); 7 describes the PF and 8 – the scenario; 9 offers an example of a typical situation; 10 gathers the names of the typical places where it occurs; and 11 details the feeling of the speaker when using this pragmateme or formuleme. There are additional columns in the database, such as the attitude of the speaker and/or potential attitude of the listener after hearing this expression; lack of space does not allow the presentation of all the features we are working with.

In *Diretes* we work with the concepts of phraseological schemes and proverbs functioning as pragmatemes, formules, etc. We also utilize the concept of pragmateme (as defended in Barrios & Ovejas 2019 and Barrios 2020a), and that of stereotyped speech act (SSA) (Kauffer 2013). An SSA shares almost all the characteristics of the pragmateme described in Section 1 but the extra-linguistic situation is less restricted than that of the pragmateme. Subsequently, if someone says *what a pity!*, the extra-linguistic situation associated with this expression only demands an event classified as something bad, and a feeling of compassion or pity on the part of the speaker, but there exists no scenario or other lexical anchorage different from *pity* or *compassion* (observe that they are abstract words whereas

---

[2] We acquired these examples of the use of *wet paint* and *do not touch* in several images on the Internet attached to objects recently painted, while the images of *fresh paint* were of houses and rooms.

ship, drown or *rescue* for *Man overboard!* are concrete words). For this reason, it is possible to identify hundreds of typical situations attached to any SSA.

1	2	3	4	5	6	7	8	9	10	11
id-L	Lexical Anchorage 1	Lexical Anchorage 2	Lexical Anchorage 3	Pragmatema	Answer	Pragmatic Function	Scenario	Typical Situation	Place	Feeling
¡hombre al agua!	barco	ahogarse	rescate	Sí	No	**Alertar_ para que se rescate a alguien**	Alguien ha caído de un barco y puede ahogarse.	Un marinero cae en una maniobra con mar picada, y otro compañero lo ve, y grita: ¡hombre al agua!	Barco	Alarmado
Man overboard!	ship	drown	rescue	Yes	No	**To warn to rescue someone**	Someone has fallen off the ship and could drown	Navigating in a rough sea during some manoeuvers, a sailor falls off the ship and someone else cries: ¡man overboard!	Ship	Alarmed

Figure 1. Description of the pragmateme "Man overboard!" in the Spanish e-dictionary *Diretes*

As *Diretes* is a project in progress, we rely on the web page[3] where it is possible to show, for the present time, only some of the data from our database. For instance, if someone writes *hombre* (man) in the search engine of the web page, the response will be a set of dictionary entries containing the word *man*, among them some idioms, the pragmateme *hombre al agua* (Man overboard!) and the phraseological scheme *pobre hombre* (lit. poor man), such as shown in Figure 2.

hombre	sustantivo
hombre anuncio	locución nominal
hombre bala	locución nominal
hombre de negocios	locución nominal
hombre de paja	locución nominal
hombre del saco	locución nominal
hombre lobo	locución nominal
hombre orquesta	locución nominal
hombre rana	locución nominal
gentilhombre de placer	locución adjetiva
¡hombre al agua!	fórmula oral o escrita
¡pobre hombre!	estructura productiva

Figure 2. Results after writing *hombre* (man) in the search engine of the Spanish e-dictionary *Diretes*

---

[3] Available in: http://diretes.es/.

Currently, due to the development status of the program, we cannot automatically recover the lexical anchorage from database to web page, and therefore we cannot arrive at *Man overboard!* from the entries of *ship*, *drown* and *rescue* as was hoped (we will endeavour to develop better software in the coming years). For this reason, if we click on *¡Hombre al agua!* (see the next to final line in Figure 2) the dictionary shows only some of the data collected in the database, as Figure 3 shows, especially when compared to 1:

¡hombre al agua!

Categoría gramatical: fórmula oral o escrita
Definición mínima: Se dice cuando alguien cae al agua desde un barco.
Definición expandida: Normalmente se grita en barcos de marina mercante o militar para avisar de que hay que rescatar a qu
Etiqueta semántica: Formulema
Función pragmática: Avisar de que alguien ha caído de un barco y puede ahogarse.
lo de situación típica 1: Un marinero cae en una maniobra con mar picada, y otro compañero lo ve, y grita: ¡hombre al algua!
Escenario: Barco

Figure 3. Entry of the dictionary for the pragmateme *¡Hombre al agua!* (Man overboard!) in *Diretes*

In order to ascertain whether the formalization we are working with would be both possible and successful, we have, as a pilot test, formalized more than two hundred formulemes (sixty of them pragmatemes) of the Spanish e-dictionary *Diretes* by means of Pragmatic Functions created *ad hoc*. At the conclusion of our task we had collected thirty PFs similar to *to warn*, specified by different paraphrases, in like manner to *to rescue someone* (see column 7 from Figure 1).

In Section 5, we will present some of the results of the analyses. Our data proves that PFs were more appropriate for our didactic purpose than LFs created *ad hoc*, and that any kind of formuleme (whether pragmateme or not) can be analysed by this method.

3. Illocutionary acts

Regarding the Pragmatic perspective, few researchers have addressed the issue of how to compile a complete and coherent inventory of illocutionary verbs. Searle's taxonomy was proposed (1968) and revised by the author (1975, 1979, 1990). It was also presented in a coherent proposal for Illocutionary Logic (Searle & Vanderveken 1985). Vanparys analyzed the valence potential of 120 illocutionary verbs in written and written-to-be-spoken language from a cognitive perspective, concluding that "a that-clause portrays the content as a more or less independent entity, while an infinitival complement presents it as being more dependent on the speech act" (1996: 221). More recently Weigand offered a dialogic taxonomy of minimal games based on Searle's monologic speech act taxonomy (2010: 154).

After his revisions, Searle (1990) recognized that there should be few changes implemented:

> There are five, and only five basic things we can do with propositions: we tell people how things are (assertives), we try to get them to do things (directives), we commit ourselves to doing things (commissives), we express our feelings and attitudes (expressives), and we bring about changes in the world so that the world matches the proposition simply by virtue of the utterance (declaratives) (Searle 1990: 410).

As the author claimed, his proposal was not to offer a list of illocutionary verbs but to revise Austin's taxonomy by means of a richer theory, and to offer a taxonomy of illocutionary acts. Responding to some critics, particularly related to certain verbs, he comments: "*remark* and *comment* are not names of types of illocutionary acts, they are illocutionary verbs" (Searle 1990: 417).

From our point of view, as outsiders to Searle's theory, this is one of the most interesting points of the controversy. As Austin stressed, some words, such as *hereby*, are a "useful criterion that the utterance is performative" (Austin 1962: 57). The equivalent to *hereby* in Spanish is *por la presente*: words and idioms play a crucial role by marking the situational context, in this case, an official document. In MTT, *hereby* is considered a pragmateme. In fact, some of the most well-known examples of Austin's proposal (*I declare you married*[4], *I name this ship, I give and bequeath*) are also pragmatemes (they are fixed and attached to particular situations in the real world).

On the other hand, there are performative words, such as *promise*, that can be used both as performative (*I promise you*) and non performative (*you promised*) (Austin 1963: 59). Thus, the item can be labeled as an illocutionary verb depending on the grammar. We can conclude that illocutionary force is expressed both by vocabulary and by grammar, and the combination of both conditions offers a wide variety of possibilities. Searle accepted this idea of Austin's, and questioned whether there are some "kernel elements in illocutionary force on which these various operations are performed", as well as if "there is a finite list of these elements" (1990: 410).

Analyzing some of the data of Searle's work, we realized that we can identify idioms, formulemes and pragmatemes attached to different types of illocutionary acts. Table 1 shows certain examples we created for different types of assertives, the first group in Searle's taxonomy (we take the subtypes from Searle 1975: 347):

We attempted to classify our data and found illocutionary verbs, such as *mention* and *comentar* (to comment) present in compositional (semantically clear) discourse markers, which in turn are formulemes (see lines 4 and 5). We also found formulemes (*Elementary, my dear Watson; What else does he want?*) and pragmatemes (*I declare you married*) which are *per se* performative utterances (see lines 7, 9 and 11).

---

[4] We consider this expression here instead of the original answer written by Austin (*I do*) because he realized it was a mistake, although he could not change it in the original book, as explained in note 2 (Austin 1968: 5).

Table 1. **Types of assertive from Searle's taxonomy with some examples and our comments**

TYPE OF ASSERTIVE	EXAMPLE (IN ENGLISH)	EXAMPLE (IN SPANISH)	COMMENTS
Statements *Afirmación*	Aristotle **postulates** that man is a political animal	*Aristóteles **afirma** que el hombre es un animal político*	Performative-looking word potentially used as non performative[5]
Assertions *Aseveración*	He's not living there, **that's for sure**	*Él no vive ahí, **eso seguro***	Discourse marker: concluding an assertion.
	I **assure** you **I will do my best**	*Te **aseguro** que **haré todo lo que pueda***	Illocutionary verb + false assertion (promise) & formuleme (I'll do my best)
Remarks *Comentario*	**Not to mention**	***Sin mencionar que (...)***	Illocutionary verb *mention* as part of a compositional discourse marker: before adding information
Explanations *Explicación*	(lit. I'll commentate on)	***Te comento, (...)*** *(oral use)*	Illocutionary verb *comentar* as part of a compositional discourse marker: before adding information and explaining something
Declarations *Declaración*	He made a **statement** to the press	*Hizo una **declaración** a la prensa*	Performative-looking word used as non performative[6]
	I pronounce/declare you married/ man and wife/ You may seal your vows with a kiss	***Yo os declaro marido y mujer***	Pragmateme. Performative utterance. Function: marry two people
Deductions *Deducción*	Atistotle **deduced** that Earth is spherical	*Aristóteles **dedujo** que la Tierra es redonda*	Performative-looking word used as non performative
	Elementary, my dear Watson	***Elemental, querido Watson***	Formuleme. Performative utterance. Function: emphasise the deduction
Arguments *Argumento*	He has good reasons for leaving this job	*Tiene motivos justificados para dejar ese trabajo*	Non performative-looking sentence used as performative (the speaker believes that what he says is true)
	What more does he want?	***Qué más quiere?***[7]	Formuleme. Performative utterance. Function: to complain (someone demands some logical arguments speaking with some other person about a third person, maybe a boss, who is not satisfied with the speaker)

---

[5] We claim that in this example it could be a non performative use because the speaker could say: Aristotles **postulates** that man is a political animal but I claim it is not true. Here we use Austin's terminology, although it corresponds to Searle's words, to an illocutionary verb used in a non-illocutionary act, as we will see later.

[6] Again, from here on we use Austin's terminology (1963) within the Table 1 . We believe it to be clearer for the didactic intention of this Table.

[7] Apparently this expression is a question when someone is searching for arguments in order to understand the attitude of another person, whilst concurrently this person is complaining.

The last examples (lines 7, 9 and 11) were the catalyst for us to design and develop the research we aim to showcase in the next sections, specifically because these kinds of formulemes and pragmatemes are idiomatic (fixed), hence the grammar does not relate to their function.

4. Hypothesis, data and method

Our hypothesis is that formulemes and pragmatemes constitute a set sufficiently bound to be properly analyzed as speech acts and, at the same time, a set large enough to allow us to create a representative list of illocutionary verbs.

One datum that would initially support our hypothesis is that there should be at least one pragmateme or one formuleme for each illocutionary act in the didactic material focused on Pragmatics. We will attempt to explain in a few words the argument that supports this idea. From our perspective, we use idioms and proverbs because of linguistic economy to express productive concepts in a particular language, avoiding the need to look for a novel expression each time. Languages have expressions for frequently used meanings because of "the possibility of an internal economy of speech" (Zipf 1949: 20). We could "therefore expect that fundamental speech act types have economically short grammatical expressions at their disposal" (Weigand 2010: 155).

Assuming that, as pragmatemes and formulemes express frequent utterances in a fixed way, linguistic economy reasons would equally compel speakers to select a well-known expression (well-known implies easier to be understood by the listener) to assure the felicity (which means well understood by the listener) of the illocutionary act.

We will try to verify if our hypothesis is valid working with two main sources of data:

a) the didactic material created by the *Instituto Cervantes* for the teachers of Spanish as a Second Language, called *Plan Curricular del Instituto Cervantes* (PCIC), available from the Internet[8];

b) the list of pragmatemes, formulemes, SSA, phraseological schemes, and proverbs functioning as Speech Acts, collected and formalized to date in the Spanish dictionary *Diretes* (see Section 2).

As the second source was presented in Section 2, we will summarize here some of the characteristics of the first. The relevant part of the PCIC dedicated to Pragmatics is a repertoire of one hundred and twenty functions classified in six groups:

a) to give and ask for information;
b) to express opinions, attitudes and knowledge;
c) to express pleasure, desires and feelings;
d) influencing the interlocutor;

---

[8] https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/niveles/05_funciones_intro-duccion.htm

e) to relate socially with other people;

f) to structure the discourse.

In turn, all of these groups can be classified in several groups. For instance, in the third group (c), the PCIC offers the Spanish versions of *to toast*, *to offer condolences*, *to congratulate*, *to welcome*, etc.; whilst in the sixth group (f) there are expressions for greetings, telling stories, etc. All of the material is also classified by levels according to the Common European Framework of Reference for Languages (A1, A2, B1, B2, C1, C2).

For each sub-group and for each level, the PCIC collects several examples. Most of them correspond to utterances freely written by the creators of the material, but for some functions there are many examples of pragmatemes and formulemes, such as in (a) to ask for information: *what's your name?*, *how old are you?*, *where are you from?*; in (b) to express disagreement: *I don't think so*; in (c) to ask about feelings: *how are you?*; and many more[9]. The list of one hundred and twenty functions of the PCIC likely constitutes so far the most complete list of Spanish Speech Acts (see Section 3).

We followed a very simple research methodology: initially we checked the complete list of examples of the PCIC and confirmed that there is at least one pragmateme or formuleme for each function it proposes[10]. Assuming that it partially validates our hypothesis, we then decided to analyze the PCIC's examples and to compare their function with the list of our Pragmatic Functions created *ad hoc* for the Spanish e-dictionary Diretes, as we have seen in Section 2.

5. Results

As we mentioned in Section 4, as a result of the analysis in *Diretes* we obtained a list of thirty Pragmatic Functions created *ad hoc* to express the function of more than two hundred Spanish pragmatemes. In Table 2 (second column) we portray eighteen of these PFs (lack of space prevents us from illustrating more results), and we endeavour to relate these PFs to Searle's taxonomy (see Section 4). The third column shows a number we will use to comment on some of these pragmatemes in a straightforward manner. The fourth shows Spanish pragmatemes and the fifth the English equivalents. Regarding assertives and declaratives, until now these types of speech acts involve fewer pragmatemes than the other types. However, this seems to be not relevant for now because hundreds of pragmatemes and formulemes remain to be analysed.

As Table 2 shows, there are English equivalents for some of the Spanish pragmatemes and formulemes (as we cannot deal with varieties of Spanish, we only work with the Spanish of Spain). However, this concept should be applied cautiously. For instance, in (13) *es así de claro* seems to be slightly different to *it's*

---

[9] In the interest of reading clarity, we have written the English equivalents to the Spanish expressions of the PCIC.

[10] We will present some of the problems we identified in Section 5.

that clear: apparently the English expression can be used assertively, while the Spanish expression is used to express attitude, usually after evaluating something or someone.

Table 2. **Pragmatemes and Formulemes, as well as its Pragmatic Functions in *Diretes***

Searles's taxonomy	Pragmatic Functions in *Diretes*	N.	Spanish pragmatemes and formulemes in *Diretes*	English equivalents or explanation of use
Assertives	Deducting	1	*Elemental, querido Watson*	*Elementary, my dear Watson*
Directives	Encouraging people to play like a child or enjoy themselves with jokes	2	*Ábrete Sésamo*	*Open Sesame*
		3	*Frío, frío*	*Cold, cold*
		4	*Pide un deseo*	*Make a wish!*
		5	*Cierra los ojos*	*Close your eyes!*
	Manipulating	6	*No hay por dónde cogerlo*	*There is no way to take it*
	Ordering	7	*Pasemos página*	*Let's draw a line on the past*
	Ordering with arrogance	8	*Porque lo digo yo y punto*	*Lit. Because I say it and that's it*
Commissives	Pledging sth	9	*¡Cuenta conmigo!*	*Count on me!*
	Relinquishing responsibility	10	*Eso es cosa tuya*	*It's up to you*
Expressives	Expressing an excuse	11	*No pude hacer otra cosa*	Lit. *I cannot do another thing*
		12	*Ha sido sin querer*	*It was unintentional*
	Emphasising	13	*Así de claro*	Lit. *That clear*
	Emphasising and expressing refusal	14	*Por encima de mi cadaver*	Lit. *Above my cadaver* (used to refuse)
		15	*Ni por todo el oro del mundo*	*Not even for all the gold in the world*
	Expressing evaluation	16	*Ni fu ni fa*	*Not good, not bad*
		17	*Esto es pan comido*	*That's a piece of cake*
		18	*A la vejez viruela*	Lit. *To old age, smallpox*
		19	*Has hecho muy bien*	*You did the right thing*
		20	*Estás muy equivocado*	*You're very wrong*
	Expressing negative reproach	21	*A buenas horas, mangas verdes*	Lit. *At good hours, green sleeves*
	Expressing positive reproach	22	*Qué calladito te lo tenías*	Lit. *How quiet you had it*
	Expressing positive feeling	23	*Dichosos los ojos*	Lit. *Blissful the eyes*
	Expressing approval	24	*Así está bien*	Lit. *Like this is good*
	Expressing indifference	25	*A mí qué me cuentas*	*Why do I care about that?*
	Expressing relief	26	*Hogar, dulce hogar*	*Home, sweet home*
Declaratives	Getting married	27	*Sí, quiero*	*Yes, I do*

In (19) there is an SSA which is quite frequently used in Spanish, *has hecho muy bien* (lit., you did very well). Because of the influence of dubbing actors in English or American films, we frequently hear *has hecho lo correcto* which is literally *you did the right thing*. However, whereas the English expression seems to be strongly attached to US culture, its literal translation does not correspond very

well to Spanish culture. Something similar occurs with *lamento su pérdida* (I mourn your loss): in Spain we say le *acompaño en el sentimiento* (lit. I accompany you in your feelings), but again, the film's translators translate the English pragmateme word for word.

In (24) the Spanish expression is apparently equivalent to the English one, but it is a false equivalence. In fact the equivalent to *that's right* is *de acuerdo, está bien*. The formuleme *así está bien* is used in Spain in more restricted extra-linguistic situations, for instance, when someone is adding milk to your coffee and you want her to stop.

Finally, as far as we know, in (26) the pragmateme *hogar dulce hogar* is slightly different from the English one: In Spain it is not associated with homeland, it is said when some arrives home, for instance, after several hours out in the cold, or after a number of days living in some other place, the speaker expresses the joy of returning home. We do not consider possible ironic uses in this case.

At this juncture we should add that there are some pragmatemes and formulemes from *Diretes* without a clear link to one of the five types of Searle's illocutionary acts (see Section 3). For instance, while we have labeled as directives intimidating (*ándate con ojo*, watch your back) and warnings (*ojo al parche*, keep an eye on), there is a need to revise pragmatemes labeled as manipulating (see the one in 6, Table 2): when someone says *no hay por donde cogerlo* (there is no way to take it) this person is evaluating and, at the same time, trying to influence the opinion of the listener.

6. Problems which arose

Several problems arose while analysing our data, and we will briefly present some of them.

The first relates to the richness of linguistic information of the illocutionary verbs. If we return to Figure 1, initially in our database we only used the verb *Warn* to express the function of the pragmateme (what we call PF, see Section 2). The meaning of *to warn* demands that someone (X) warns someone else (Y) about something (Z): we realized that, because of its argument structure, we can use this verb as an illocutionary verb by focusing the meaning on one or another argument. For instance, as we see in Section 2, if someone cries *Man overboard!* to some other people on a ship, someone usually tries to rescue the person who fell from the ship. The speaker (X) wants someone (Y) to save this person (Z).

Similarly, if someone cries *Mayday! Mayday!* on a ship (or in some other place) to some other people, those others would attempt, to whatever possible extent, to help them to avoid imminent danger. The speaker (X) wants some other people (Y) to save the crew and passengers on board from serious danger (Z).

However, if someone cries *Fire!*, other people who hear this cry would run far away from the place where it is taking place; likewise if someone cries *A bomb!*, the speaker (X) wants other people (Y) to be safe from fire/bomb (Z). For this reason, we chose to specify where the focus of *warning* was directed, adding some semantic features to the PF, as (5) shown:

(5) a. To warn_ someone to run: *A bomb!*; *Fire!*
b. To warn_to rescue someone: *Man overboard!*; *Mayday! Mayday!*

The reader is potentially thinking that someone could cry *Man overboard!* before leaping to help this person. In this case, the speaker (X) wants someone else (Y) to know he is going to rescue this person (Z). Observe that the question then arises as to whether the illocutionary verb should be *to inform* instead of *to warn*. Should we introduce two lexical entries for these two different speakers' intentions? Should we write *to inform that the speaker is going to rescue someone*? As yet we have no answer. Most probably we need to work on the terminological field in order to know how many pragmatemes would be adequate for the potentialities of this last PF. We wonder if this situation is frequent in everyday life. Our picture of illocutionary words is thus still incomplete.

The second problem relates to the examples of the PCIC's lists we were consulting. We present some of them in (6)[11]:

(6) a. *No seas miedica y tírate ya, hombre* (Don't be a chicken and just throw yourself off, man!)
b. *¡Lánzate!* (Jump!)

(7) a. *Te echo una mano/ un cable/ un capote* (Lend a hand/ a cable/ a cap)
b. *¿De cuánto estás? vs. ¿De cuánto está?* (Literally How many (months) are you? vs. How much is she?)

(8) a. *Toda la vida cuidando de vosotros para que luego me tratéis así* (Literally All your life taking care of you and later you treat me like this)
b. *Toda la vida cuidando de ellos para que luego me traten así* (Literally All your life taking care of them and later they treat me like this)

In (6) we show an example of PCIC (6.a) which is written without any explication of the extra-linguistic situation. As the Spanish verb *tirarse* (to jump) is attached to a physical movement, this utterance would usually be said close to a swimming pool, when someone (typically a child) is trying to jump off a diving board without success. The PCIC does not say anything about any context or about the expression in (6.b): the verb *lanzarse* (to launch) is used not in a physical but in a figurative sense, as in *come on and enrol in the master degree*!).

In (7) we show the problem of idioms and phraseological schemes used as pragmatemes: (7.a) reflects three PCIC utterances based on idioms and (7.b) shows a phraseological scheme without an equivalent in English, *¿De cuánto estás?* (lit. how much are you?). This is a pragmateme used by the speaker when asking a pregnant woman how many weeks or months the baby is.

In such cases, regular expressions are relied on: the idioms *echar una mano/ un cable/ un capote,* lit. lend a hand/ a cable/ a cap, and the productive structure *estar + de + X + semanas/meses* (lit. to be + of + X + weeks/months). The

---

[11] All the examples preceded by the letter (a) come from the PCIC, whilst the ones preceded by the letter (b) are written by us to show the kind of problems we encountered working with this material.

phraseological scheme explains the frequent use of this verb in indirect style, such as speaking about a pregnant woman (the speaker asks someone how many weeks or months along is (the baby) (of) a pregnant woman which is not present at this moment). Should we teach the pragmatemes without any reference to the regular expressions through which they are formed?

In (8) we show two sentences with minimal grammatical difference: (8.a) uses the second person and (8.b) the third person. However, as in other cases, minimal grammatical differences involve relevant pragmatical deviation: (8.a) is a reproach whilst (8.b) is a grumble. We do not find any relevant explication regarding these circumstances in the PCIC.

Consequently, we claim that any didactic material, whatever it may be, a dictionary such as *Diretes* or an on-line resource like the PCIC, should be enriched with explications that can solve the problems we mention in this Section, even if the solutions are not evident.

7. Conclusions

The set of Pragmatic Functions could be useful not only for the formalization of formulemes and pragmatemes in other languages but also as a catalogue of illocutionary verbs based on phraseology.

However, given the small sample, vigilance must be observed: size limitations of our present research give rise to many questions which will most likely be challenging for years to come. Detailing any small feature related to the function and the extra-linguistic situation for each pragmateme is a huge, necessary and demanding task.

Returning to our hypotheses, and recognising that our study is a simple pilot test, data indicate that formulemes and pragmatemes constitute a set sufficiently bound and, concurrently, large enough to create a representative list of illocutionarys verbs. However, their analysis will take more time than was anticipated when we commenced this study.

Further studies, which take the number of parameters into account will need to be undertaken in order to clarify if there is a relation between the parameters and the position of the pragmateme related to the continuum of this concept: could it be possible that the larger the number of parameters, the more central the pragmateme is? Does the presence of more parameters hold for the more essential pragmatemes, while fewer parameters imply more outlying pragmatemes? In other words, are the number of parameters directly attached to the degree of typicality of the pragmateme?

Another important issue to resolve for future studies is the relation between the idiomaticity of one expression and its categorization: are outlying pragmatemes responding to grammar rules or phraseological tendencies? Should the phraseological schemes, such as *made in X*, be analyzed as phrasemes or as productive constructions?

We would like to conclude this paper with the two first pragmatemes we mentioned in the Introduction, using both now in direct speech, saying:

Dear Igor Mel'čuk, *Happy birthday* and *have a nice* day and a lovely new decade! Pragmatemes exist in this world thanks to you. We wish you ninety more years to discover new concepts in the immense world of words. *Thank you for everything!*

Acknowledgments

REFERENCES

Apresjan, Jury. 2018. *New Explanatory Dictionary of Russian Synonyms*. Dublin: PubMix.com.

Apresjan, Jury, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov & Leonid Tsinman. 2003. ETAP-3 Linguistic Processor: A Full-Fledged NLP Implementation of the Meaning-Text Theory. Institute for Information Transmissions Problems. Russian Academy of Sciences. http://proling.iitp.ru/bibitems/ETAP_and_MTT.pdf [15/11/2021].

Alonso Ramos, Margarita. 2004. Elaboración del Diccionario de colocaciones del español y sus aplicaciones. In Paz Battaner & Janet A. De Cesaris (eds.), *De lexicografía: Actes del I Symposium Internacional de Lexicografía*, 169–162. Barcelona: Institut Universitari de Lingüística Aplicada.

Austin, William James. 1962. *How to do Things With Words*. Oxford: Clarendon Press.

Barrios, María Auxiliadora. 2017. Hacia un concepto amplio de pragmatema y sus aplicaciones en ELE: El caso de ¡qué + sust./adj.! In Belén Almeida Cabrejas, Ana Blanco Canales., Jairo Javier García Sánchez & María Dolores Jiménez López (eds.), *Investigaciones actuales en Lingüística* (II). *Semántica, Morfología y Lexicología*, 19–35. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

Barrios, María Auxiliadora. 2020a. Lexical Functions and Pragmatic Functions: A proposal for the formalization of pragmatemes within the Meaning-Text Theory. In Joanna Szerszunowicz & Martna Awier (eds.), *Intercontinental Dialogue on Phraseology 9: Reproducible Multiword Expressions from a Theoretical and Empirical Perspective*, 15–32. Bialystok: The University of Bialystok Publishing House.

Barrios, María Auxiliadora. 2020b. The making of the Diretes dictionary. In Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras (eds.), *Lexicography for Inclusion. EURALEX Proceedings. XIX Congress of the European Association for Lexicography*, 13–22. Alexandroupolis: Euralex and Democritus University of Thrace

Barrios, María Auxiliadora & Vanesa Ovejas. 2019a. Pragmatèmes : Concept, limites et formalisation. *Cahiers de Lexicologie* 115 (2). 77–102. https://doi.org/10.15122/isbn.978-2-406-10060-7.p.0077

Barrios, María Auxiliadora & Igor Boguslavsky. 2019b. A Spanish e-dictionary of collocations. In Kim Gerdes & Sylvain Kahane (eds.), *Proceedings of the Fifth International Conference on Dependency Linguistics. ACL Anthology*. https://www.aclweb.org/anthology/events/ws-2019/#w19-77

Blanco, Xavier. 2013. Les pragmatèmes : Définition, typologie et traitement lexicographique. *Verbum* 4. 17–25. https://doi.org/10.15388/Verb.2013.4.4977

Blanco, Xavier. 2014. Microestructura lexicográfica para unidades frásticas: Los pragmatemas. *Revista Káñina* XXXVIII. 13–18.

Cowie, Anthony Paul. 2011. Speech formulae in English: Problems of analysis and dictionary treatment. In Geart van der Meer & Alice G. B. ter Meulen (eds.), *Making sense: From lexeme to discourse, in honor of Werner Abraham on the occasion of his retirement*, GAGL 44, 1–12. Groningen: University of Groningen.

Frechon, Genevieve, Paolo Frassi & Alain Polguere. 2013. Les pragmatèmes ont-ils un charme indéfinissable ? In Pierluigi Ligas & Paolo Frassi (ed.), *Lexiques. Identites. Cultures*, 81–104. QuiEdit.

García Page, Mario. 2007. Los pragmatemas: Algunas consideraciones. In Otal J. Cuartero & M. Emsel (eds.), *Festschrift für Gerd Wotjak zum 65. Geburtstag. Band 1: Vernetzungen Bedeutung in Wort, Satz und Text*, 161–173. Frankfurt: Peter Lang.

Iordanskaja, Lidija & Slava Paperno. 1996. *A Russian-English Collocational Dictionary of the Human Body*, In Richard L. Leed (ed.). Bloomington: Slavica Publishers, INc. [trad. Lesli LaRocco & Jean MacKenzie] http:// russian.cornell.edu/body/ [15/09/2022]

Kauffer, Maurice. 2013. Le figement des « actes de langage stéréotypés » en français et en allemand. Practiques 159/160. 42–54.

Kauffer, Maurice. 2017. De la notion de pragmatème en phraséologie. In Cosimo De Giovanni (ed.), *Fraseologia e Paremiologia Passato, presente, future,* 70–81. Milan: Francoangeli.

Mel'čuk, Igor. 1995. Phrasemes in language and phraseology in linguistics. In Martin Everaert et al. (eds.), *Idioms*: *Structural and psychological perspectives*, 167–232. Nueva Jersey: Erlbaum Associates.

Mel'čuk, Igor. 1996. Lexical functions and lexical inheritance for emotion lexemes in German. In Leo Wanner (ed.), *Lexical functions in lexicography and natural language processing*, 209–278. Amsterdam/Philadelphia. John Benjamin.

Mel'čuk, Igor. 2008. Phraséologie dans la langue et dans le dictionnaire. *Repères & Applications* VI. 187–200. Barcelona: Universidad Autonoma de Barcelona.

Mel'čuk, Igor. 2013. Tout ce que nous voulions savoir sur les phrasèmes, mais... *Cahiers de Lexicologie* 102. 129–149.

Mel'čuk, Igor. 2015a. Clichés, and Understudied Subclass of Phrasemes. *Yearbook of Phraseology* 6 (1). 55–86.

Mel'čuk, Igor. 2015b. *Semantics. From meaning to text III.* Amsterdam/Philadelphia, John Benjamins.

Mel'čuk, Igor. 2018. Anna Wierzbicka, semantic decomposition, and the Meaning-Text Approach. *Russian Journal of Linguistics* 22 (3). 521–538. https://doi.org/10.22363/2312-9182-2018-22-3-521-538

Mel'čuk, Igor. 2020. Clichés and pragmatemes. *Neophilologica* 32. 9–20.

Mel'čuk, Igor & Alexander Zholkovsky. 1984. *Tolkovo-kombinatornyy slovar' sovremennogo russkogo yazyka: Opyty semantiko-sintaksicheskogo opisaniya russkoy leksiki* Wien: Wiener Slawistischer Almanach. http://olst.ling.umontreal.ca/pdf/ Melcuk_Zholkovsky_ 1984.pdf [01/12/2021]

Mel'čuk, Igor & Alain Polguere. 2007. *Lexique Actif du Français*. Bruselas: De Boeck.

Polguère, Alain. 2007. Lexical function standardness. In Leo Wanner (ed.), *Selected Topics in Meaning Text Theory, In Honour of Igor Mel'čuk*, 43–92. Amsterdam: John Benjamins.

Polguère, Alain. 2014. From writing dictionaries to weaving lexical networks. *International Journal of Lexicography* 27 (4). 396–418.

Polguère, Alain 2015. Non-compositionnalité : Ce sont toujours les locutions faibles qui trinquent. *Verbum* XXXVII (2). 257–280.

Polguère, Alain. 2016. Il y a un traître par minou : Le statut lexical des clichés linguistiques, Corela, HS-19, http://corela.revues.org/4486; https://doi.org/10.4000/corela.4486

Searle, John R. 1977. *Speech Acts. An essay in the Philosophy of Language*. 2nd edn. Cambridge: Cambridge University Press.

Searle, John R. 1975. A taxonomy of Illocutionary acts. In Keith Gurdenson (ed.), *Language, Mind, and Knowledge*. Minessota Studies in the Philosophy of Science, vol. 7, 1–29. Minneapolis: University of Minnesota Press.

Searle, John R. 1976. A classification of Illocutionary acts. *Language in Society* 5 (1). 1–23.

Searle, John R. 1979. *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Searle, John R. 1990. A Classification of Illocutionary acts. In Donald Carbaugh (ed.), *Cultural Communication and Intercultural Contact*, 349–372. New York: Psychology Press.

Searle, John R. & Daniel Vanderveken. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Vanparys, Johan. 1996. *Categories and Complements of Illocutionary Verbs in a Cognitive Perspective*. Frankfurt am Main: Peter Lang.

Weigand, Edda. 2010. *Dialogue: The Mixed Game*. Amsterdam: John Benjamins Publishing Company.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge (Massachussetts): Addison-Weslwy Press. Available at the web page https://archive.org/details/in.ernet.dli.2015.90211/page/n23/mode/2up [15/09/2022]

Bionote:
María Auxiliadora BARRIOS RODRÍGUEZ is Associate Professor of the Spanish Language Department at Universidad Complutense de Madrid. Her research interests are at the intersection between Lexicology and Lexicography. Her scientific production includes over sixty publications in international conferences and journals. She is the head of the project DiRetEs (http://diretes.es/), a Spanish dictionary that reflects collocations and semantic relations by means of Lexical Functions. She also manages a Spanish as Second Language project (www.masalladelespanol.com) and some educational innovation projects (https://masespanol.es/).
e-mail: auxibarrios@filol.ucm.es
https://orcid.org/0000-0001-6830-4797

Сведения об авторе:
Мария Ауксилиадора БАРРИОС РОДРИГЕС – доцент кафедры испанского языка Мадридского университета Компутенсе. Ее научные интересы находятся на пересечении лексикологии и лексикографии. Она автор более 60 публикаций в научных журналах и материалах международных конференций. Она руководит проектом DiRetEs (http://diretes.es/) по созданию словаря испанского языка, отражающего сочетаемость и семантические отношения с помощью семантических функций. Она также осуществляет руководство проектом преподавания испанского языка как иностранного (www.masalladelespanol.com) и некоторыми инновационными образовательными проектами (https://masespanol.es/).
e-mail: auxibarrios@filol.ucm.es
https://orcid.org/0000-0001-6830-4797

Research article / Научная статья

An explanatory combinatorial dictionary of English conflict lexis: A case study of modern political discourse

Olga SOLOPOVA ✉ and Tamara KHOMUTOVA

South Ural State University (National Research University), Chelyabinsk, Russia
✉o-solopova@bk.ru

Abstract
Though political discourse is in the mainstream of modern studies, scholars haven't so far paid much attention to compiling political discourse-oriented dictionaries. The need to further develop lexicographic theory and practice for specific purposes and advance new methods to dictionary making is a challenge that linguists are facing today. The aim of the case study is twofold: to work out the principles for making an Explanatory Combinatorial Dictionary (ECD) of English political conflict lexis and the microstructure of an ECD entry. The source of the data is the NOW corpus; the material is current American political discourse (2022). The ECD is generally consistent with Mel'čuk's Meaning-Text theory (MTT). The authors describe a process of collecting and processing the data: corpus search and analysis, automatic and manual text processing, glossary compilation with the use of lexicographic, semasiological, and etymological methods and present an example of an ECD entry consisting of semantic, phonological, and cooccurrence zones. The findings prove that the use of electronic text corpora offers an effective way for compiling a specialized discourse-based dictionary. The research illustrates the validity of MTT: though based on the data of "language in context", the dictionary is synthesis-oriented: it aims at speech production. The paper is the first result of a bigger project sketching the overall framework of the discursive ECD of political conflict lexis, which subsequent studies will hopefully develop with more precision and detail. The dictionary will be helpful for scholars in linguistics, discourse analysis, media and communication, political science, and conflict studies.
Key words: *lexicography, political discourse, explanatory combinatorial dictionary, conflict-provoking lexis, corpus, dictionary entry*

Принципы создания англоязычного толково-комбинаторного словаря конфликтогенной лексики: на материале современного политического дискурса

О.А. СОЛОПОВА⃝⃝, Т.Н. ХОМУТОВА⃝

*Южно-Уральский государственный университет
(национальный исследовательский университет), Челябинск, Россия*
✉o-solopova@bk.ru

Аннотация

Необходимость разработки теории и практики специальной лексикографии и применения новых дискурсивных методов к составлению словарей для специальных целей представляет собой важную задачу, которая стоит перед современной лингвистикой. Целью настоящей работы является: 1) разработка принципов создания англоязычного толково-комбинаторного словаря (ТКС) политической конфликтогенной лексики на материале корпуса текстов политического дискурса; 2) разработка микроструктуры словарной статьи. В качестве методологических принципов построения ТКС политической конфликтогенной лексики использованы принципы теории «Смысл-Текст» И.А. Мельчука. Источником материала является корпус News on the Web. Материал исследования представлен современным американским политическим дискурсом (2022). В статье рассматривается процесс сбора и обработки данных, включающий четыре этапа: формирование и анализ корпуса текстов, автоматическую обработку текстов, ручную обработку текстов, создание глоссария с использованием методов лексикографического, семасиологического и этимологического анализа. Приводится пример словарной статьи, микроструктура которой включает семантическую, фонологическую и сочетаемостную зоны. Результаты проведенного исследования подтверждают, что использование электронных корпусов текстов позволяет эффективно создавать специальные дискурсивные словари, «встраивая» лексикографические данные в социальный, политический и геополитический контекст. В статье представлены первые результаты большого проекта, которые определяют общую методологию создания англоязычного ТКС политической конфликтогенной лексики. Принципы теории «Смысл-Текст», ориентированные, прежде всего, на языковой синтез и производство речи, не исключают возможности его дальнейшего использования для анализа текста и дискурса. Словарь будет полезен специалистам в области лингвистики, дискурс-анализа, массмедиа, политологии и конфликтологии.

Ключевые слова: *лексикография, политический дискурс, толково-комбинаторный словарь, конфликтогенная лексика, корпус, словарная статья*

Dedication

Many people have contributed, both directly and indirectly, to the ideas and analyses of this project. But most of all we would like to thank Igor A. Mel'čuk and Alexander K. Zholkovsky for the inspiration we gained from their Explanatory Combinatorial Dictionary. The insights of Mel'čuk's Meaning-Text theory turned out to be a source of challenge and motivation: his contribution to the project has been invaluable. Using the metalanguage of lexical functions, adopted in the Meaning-Text theory, to characterize Igor Mel'čuk as a lexicographer, we must state that he is genuine (`Ver`), influential (`Magn`), and well-versed (`Bon`) in the system of language.

1. Introduction

Recent years have seen a great surge of interest in lexicography, which is caused by the active development of science, and growing political, cultural and research contacts. Lexicography is generally understood as the theory and practice of compiling dictionaries, which are closely connected and interrelated. The focus of lexicography is the ways of organizing a dictionary entry, the structure of dictionaries and the methods of their compilation. We share the view advanced by Bergenholtz and Gouws (2012) that lexicographic theory is "a discipline not only directed at compiling dictionaries, but in a more general way at producing information tools", which "can ensure enhanced information retrieval" (Bergenholtz & Gouws 2012: 40). Modern lexicography has significantly expanded and upgraded its tools with computer technologies for compiling and using dictionaries. At the same time, studies that organize lexicographic knowledge and introduce new lexicographic theories are rather scarce.

Within the context of novel lexicographic studies those of discourse lexicography should be highlighted, as the classical definition of lexicography is being transformed and modified in modern discourse studies. At the current stage of development, lexicography is interpreted more as the theory and practice of *knowledge representations* about the world, since, as Kozyrev and Chernyak (2015) state, dictionaries "respond to all changes in life and society, demonstrating a picture of dynamic language processes in a static form" (Kozyrev & Chernyak 2015: 6). Lexicographic practice has reached a new level and is becoming more anthropocentric and sociocentric. Lexicography functions within the framework of lexicographic discourse, which is interpreted as a discursive practice, as Issers (2015) puts it, "the dynamic organization of those communication systems within society, which, on the one hand, reflects the speech behavior and thinking mode of a social community, and on the other hand, it creates new forms of communication

in a given socio-cultural reality" (Issers 2015: 65). In the case of considering lexicographic discourse as a discursive practice, it is possible to "refer to the sociocultural, cognitive, communicative and pragmatic features of the dictionary" (Plotnikova 2014: 23). The study of sociocultural, cognitive and communicative-pragmatic influence on dictionary compilation and interpretation determines the appeal to advance new discursive technologies in dictionary making.

Keen interest in the anthropocentric perspective of lexicographic research can be traced in dictionaries of political discourse, since political discourse, to a great extent, sets the linguistic worldview of people. At the same time, political lexis functions both as a unit of language and as a discourse formation that implements not only dictionary but wider meanings due to the existence of a certain system in the social and cognitive-discursive space. However, to date the number of political discourse-oriented dictionaries is relatively limited, one that we could think of is the Dictionary of Modern Political Labels (Skovorodnikov & Kopnina 2021), which is definitely not enough. With this gap in mind, we propose a discourse approach to compiling a dictionary of political conflict lexis. Modern political discourse is abundant in such lexis, which presses for its lexicographic study.

Our aim is to work out the principles for building an explanatory combinatorial dictionary (ECD) of English political conflict lexis based on a large corpus of political discourse texts that will test theoretical and computational methodologies, but, above all, to provide a tool for linguists, discourse analysts, translators, interpreters, and linguistic expertise specialists. As such, our project draws on previous work that shows the inadequacies of current dictionaries to meet the needs of the specialized user.

2.Theoretical framework

2.1. Discourse lexicography

Fesenko (2015) introduces the concept of a discourse dictionary which has the following characteristics: 1) it describes not only the denotative, but also the connotative meaning of the entry (appearing in the context); 2) the dictionary is not prescriptive, but descriptive; 3) the dictionary contains socio-cultural, cognitive and communicative-pragmatic information; 4) the entry is presented in all genre and stylistic diversity (Fesenko 2015: 52). Compiling discourse dictionaries is a labor-intensive process that requires enormous efforts on the part of both theorists and practitioners of lexicographic science. Currently, linguists have done much work on the way of reforming lexicography and turning it from a theoretical construct into such a system that, as Sandomirskaya (2001) states, "refracts the general narrative of the language into the field of practical action, and captures the structure of this general narrative in its own settings" (Sandomirskaya 2001: 219).

Discourse dictionaries include a number of products, one of them is the electronic terminographic dictionary of phraseological units which actively function in Ukranian and Russian linguistics of the last century (Krasnobaeva-

Chernaya 2020). Active use of popular science discourse in mass media accounts for inclusion of terminology in the dictionary of neologisms that reflects the discursive vector of modern lexicography (Gromenko 2020). Another innovative lexicographic product is a discourse dialect dictionary based on data from Internet search engines, where "contexts act as a means of illustrating semantics" (Golev 2019: 114). A new multilingual dictionary of metaphors correlates with the discursive and lingua-cultural vector of development of lexicography, which is aimed at "solving the problem of lexicographic presentation of cultural connotation, assigned to the semantics of figurative words and expressions" (Gerasimova 2020: 95). Obviously, the lexicographic product serves to preserve the cultural heritage of a certain lingua-cultural community. In this regard, dictionaries of concepts act as an effective tool to gain insight into a certain culture, for example, on the basis of Russian vocabulary, phraseology and paroemiology (Con 2010), dialectology (CDDLP 2006–2012), lingua-cultural logic (ML 2005, 2009, 2014), concepts of Russian culture (Radbil & Saigin 2019), etc. The lingua-cultural aspect in dictionary compilation is clearly seen in the study of a dialect linguistic personality and its lexicographic fixation (Zemicheva 2017), gender parameters of dialect speech (Demeshkina & Tolstova 2017, VD 1998–2002, MDSD 2009–2010). A broad overview of discursive dictionaries is provided by Koshkarova and Solopova (2021) (Koshkarova & Solopova 2021).

Our brief analysis shows that present-day lexicography develops in the mainstream of the discursive-anthropocentric paradigm, when the dictionary reflects the functioning of the language not after the fact, but at the moment of its existence in the language community, which brings about the problem of specialized lexicography.

2.2. Specialized lexicography

Relative to discourse lexicography, as well as lexicography in general, is the problem of general and specialized lexicography. According to Bergenholtz and Gouws (2012) there are three branches of lexicography with respect to its object and functions: general lexicography, specialized lexicography, and general and specialized lexicography. The three branches of lexicography describe language for general purposes, language for specific purposes (LSP), and both languages simultaneously (Bergenholtz & Gouws 2012).

As stated above, discourse lexicography is anthropocentric and sociocentric, which means that discourse dictionaries describe the actual functioning of discourse at the moment of its generation in specific settings: social, cultural, temporal, spatial, etc. This brings about the idea that discourse lexicography is specialized per se. Nielsen (2013) points out that the two key fundamentals of specialized lexicography are its object, the dictionary, and its objective (Nielsen 2013). The dictionary is a tool providing specific types of help concerning one or more subject fields and their related LSP to specific types of users in specific types of situations. A specialized dictionary contains the necessary lexicographic data, arranged and

presented using appropriate lexicographic structures. Dictionary functions thus become the point of departure for any discussion of dictionaries and the data they contain must be specifically adapted to user needs and competences. By identifying the key components of specialized lexicography lexicographers can draw up guidelines that can help them design, evaluate, make and use "ideal" and specific types of objects in LSP lexicography (Nielsen 2013: 24).

Our aim, as noted earlier, is the ECD of English political conflict lexis of the XXI century (second decade). It means that problems of political lexicography, as well as relations between lexicography and ideology, power and politics are of great importance for our research.

2.3. Political lexicography

Lexicography and dictionaries are never value-free, apolitical or asocial. Instead, they are subject to ideology, power and politics (Chen 2019: 362). Ideology and power are aspects of a dictionary that a lexicographer and a dictionary user have to encounter in any serious lexicographic enterprise (Kachru 1995: lxv). Chen (2019) states that "a discourse approach to lexicography is needed to unmask the power relations behind, and the ways in which language serves to sustain or disrupt the relation in the global context, before a change in the status quo is possible, as such an approach – with emphasis on social context – enables a practitioner to look at the lexicographic event as a social event and the dictionary as discourse with its own rules and principles" (Chen 2019: 364). Thus, the social context and power relations surrounding the production of a dictionary should not be overlooked.

This approach is taken by Benson (2002) who seeks to reveal the ethnocentric representation in the English dictionaries published in Britain/USA and to show how knowledge in the English dictionaries is filtered through Anglo-American perspectives on English in the world (Benson 2002, Preface). He showcases that the Oxford English Dictionary was more or less explicitly a project of British imperialism concerned with the consolidation of English as the dominant language of the world. It means that the dictionary does not simply replicate its source or just 'transport' meaning; rather, it creates meaning; it rewrites and represents things in new ways (Chen 2015). We share Chen's idea (2019) that "a researcher in lexicographic discourse analysis should embed the lexicographic data in the social context, taking a political stance explicitly and focus on self-reflection as a scholar" (Chen 2019: 368).

Chen states that discourses project meaning; therefore, they evoke ideologies (Chen 2019: 370). Wodak and Meyer (2015) define ideology as a "collectively shared coherent and relatively stable set of beliefs or values" (Wodak & Meyer 2015: 30) while van Dijk (1998) presents a multidisciplinary approach to ideology involving cognitive and social psychology, sociology and discourse analysis (van Dijk 1998). That is, as Chen (2019) puts it, "socially, ideologies sustain the interests of groups; cognitively, they serve to organize the social representations (attitudes,

knowledge) of the group, and thus indirectly monitor the group-related social practices, and hence also the discourse of its members" (Chen 2019: 370).

A dictionary is a representation of the world/society seen through the lexicographer's perspective. Lexicographers are not immune from ideology. Ideologies as ways of representing and constructing society reproduce unequal relations of power, relations of domination and exploitation. For example, we may investigate how lexicographers create a dictionary, within which the historical depth and geographical breadth of the language and the world could be exposed for examination from different perspectives. To take an example, western lexicographers can define "Russian special military operation" as: a war waged by Russia against Ukraine (2022) following Ukraine's attempt to join the EU. This definition attributes the Russian aggression to Ukraine's act of 'national-liberation movement', which contradicts Russia's official view. Russian lexicographers will attribute it to the special military operation of Ukraine's demilitarization and denazification which Russia was forced to engage in. The operation was triggered by US and European elites in search of making Ukraine an enclave of NATO, which threatens Russia and Russian people in Ukraine. Lexicographic discourse studies may also be used to examine how the dictionary influences social understanding of a language and what the world is like. All this brings about the need for a special branch of lexicography, that of political lexicography.

The term "political lexicography" has been introduced by the Russian linguist S. A. Manik (Manik 2019). The novelty of her approach is determined by a broad lingua-cultural and discursive-pragmatic context, including the involvement of users in the process of dictionary compiling. The researcher gives a broad overview of existing political terminology dictionaries (Manik 2019). However, none of them are classified as discourse dictionaries. The author explores the discursive-communicative and cultural properties of English political terminology but fails to attribute the corresponding dictionaries to those of political discourse. Since our aim is to build an ECD of English *political conflict* lexis, we went in search of political conflict dictionaries.

Among the existing dictionaries we managed to find some special dictionaries of conflict, or conflict-provoking terms, which are rather few (Davletchina 2005, Dmitriev 2012, Merriman & Barrach-Yousefi 2021, Miller & King 2003, The Law Dictionary, Shipilov & Antsupov 2020). Conflict-provoking terms are understood as words that cause negative emotions and trigger conflicts. As for special political conflict dictionaries we failed to find any. That is why our research will be based on large corpora of English political conflict discourse and dictionaries of conflict terms available. One more issue worth discussing in connection with compiling the dictionary of political conflict lexis is that of combinatorial lexicography.

2.4. Combinatorial lexicography

Many of the discourse dictionaries mentioned are in fact combinatorial dictionaries, i.e., dictionaries of collocations. DeCessaris (2013) points out that

"most words in the general vocabulary of a language do not have a clearly identifiable meaning out of context, rather they have a meaning potential that becomes specified once the context of use is established" (DeCessaris 2013: 16). Firth (1957) states that meaning only exists in context and through interpretation of that context, so that words can only be understood through the company they keep (Firth 1957: 11), thus, dictionaries can only ever show meaning potentials. Meaning lies in the present, it is about how people use words now: words do not have meanings, meanings have words. Once these premises are accepted, we can look for meaning potentials in corpora (Williams 2013: 26). Hence, we can term combinatorial dictionaries the forerunners of discourse lexicography which paved the way to discourse lexicography proper. Early examples of research into collocations appeared in Shakhmatov (1898), Ushakov (1935–40), Ozhegov (1949), but current understanding of a collocation as a fundamental factor in word meaning started in Russian lexicography only with DMLR (1948–1965), and in foreign studies with "Les mots et les idees. Dictionnaire des termes cadrant avec les idées" (Lacroix 1956), and the BBICDE (1990).

By the end of the XX century, the ideas of lexicographic reflection of lexis combinatorial properties continued to develop and improve. Russian lexicography started to explore new approaches to the interpretation of syntagmatic connections of words, e.g., Apresyan (1974), Zholkovsky & Mel'čuk (1967), Morkovkin (1977), etc. The description of the combinatorial properties of words was provided in the Dictionary of Combinability of the Russian language (DCRL 2002); Dictionary of Russian and English Lexical Intensifiers (DRELI 2007); ECD of the Russian Language (Mel'čuk & Zholkovsky 1984, 2016); New Explanatory Dictionary of Russian Synonyms (NEDRS 2000); Russian Associative Dictionary (Karaulov 2002) and others. A comprehensive overview of different trends in the development of Russian and foreign combinatorial lexicography is offered in (Vlavatskaya 2013).

In this paper we will try to apply the principles of compiling combinatorial dictionaries put forward by Mel'čuk and Žolkovsky (Mel'čuk & Zholkovsky 1984, 2016). We share the view advanced by M. V. Vlavatskaya (2013) that up to date this is the most original and effective dictionary project describing combinatorial-syntagmatic potential of the word (Vlavatskaya 2013). Mel'čuk and Zholkovsky's *ECD* is a theoretical dictionary, anchored in a linguistic theory – namely, Meaning-Text theory (MTT) – and making use of its conceptual tools. The general concept of the dictionary claims: "Natural language is a system that establishes correspondences between any given meaning and all texts expressing it; accordingly, the linguistic description of a certain language should be a set of rules that put all the texts of this language in line with any meaning" (Mel'čuk, 1995: 4). In MTT, meaning is no more than the invariant of synonymous paraphrases and it is thought to be directly accessible to speakers, being a part of their intuitive language competence.

The ECD has three general characteristics: 1) it is active, oriented towards speech production; 2) it is semantic (explanatory); 3) it is combinatorial, describing syntactic and lexical cooccurrence in a systematic way. Mel'čuk and Milićević advance the following principles for compiling an ECD: 1) the formality principle which stands for the formal description of lexical units; 2) descriptive coherence principles which mean that both internal and external coherence should be achieved; 3) uniform treatment principles which demand that all linguistic units of the same class and vocables of the same lexical field must be described in a similar way; 4) the internal exhaustiveness principle which means that the description of a lexical unit must contain all the necessary information to use language correctly and find any other lexical units semantically related to this particular lexeme (Mel'čuk & Milićević 2020: 203–209).

The best-known feature of the ECD and its basic notion is a lexical function that serves "a formal tool used to describe all types of lexical relations in a systematic way" (Mel'čuk & Milićević 2020: 142, 161). This dependency associates with a lexical unit (its argument or keyword) and a set of synonymous lexical units that express a specific meaning associated with the former. Lexical functions can be paradigmatic and syntagmatic: paradigmatic lexical functions are derivative, related to the selection of an adequate value of the keyword, while syntagmatic lexical functions are combinatorial, encoding the collocational potential of the keyword. Any function meets specific requirements: it must embrace a relatively large number of pairs of words; it can be expressed in various ways, which implies the existence of a phraseological relation between the arguments and their values (for detailed treatment see Mel'čuk & Milićević 2020).

The proposed dictionary, as well as ECDs in general, lexicographically provides the performance of the Meaning-Text model (MTT) and is similar to explanatory dictionaries, where lexical combinability and syntactic properties of words are reflected. This concerns general language dictionaries. However, discourse, which is language in real life context, is as multifaceted as life itself, and compiling such a dictionary would imply much time and effort on the part of lexicographers. Here arise the notions of domain-specific knowledge and specialized lexicography which is intended to help users build LSP discourse. As stated above, we refer discourse dictionaries to specialized lexicography. In our project we will try not merely copy Mel'čuk's model but introduce some novel principles to modern lexicographic practice, which will allow us to describe the conflict-provoking system of political discourse as a mobile and dynamic formation that is influenced by extra-linguistic factors.

3. Materials and methods

The case study, presented in the paper, addresses the specific issue of building an ECD of English Political Conflict Lexis by the in-depth analysis of the principles of compiling the dictionary and the microstructure of an ECD entry through both qualitative and quantitative methods. The principles of a research corpus

compilation are of utmost importance for the results to be obtained. Collecting and processing the data for the research was a four-fold process, involving a) corpus search and quantitative analysis, b) automatic text processing, c) manual text processing, and d) glossary compilation with the use of lexicographic, semasiological, and etymological methods.

3.1.   Research corpus compilation and corpus search

The source of the material was the News on the Web Corpus from English-Corpora.org (NOW). Nowadays text corpora are considered the most important tool for research in both computational and other branches of modern linguistics (Soloviev et al. 2022). Currently, the corpus contains web-based periodicals (2010–present), covering newspapers and magazines from more than twenty English-speaking countries. As the ECD of English Political Conflict Lexis is planned to be an experimental dictionary based on a large corpus of political discourse texts, it determines the data the source of the material and the research corpus are supposed to comprise. The central topic of discourse must concern a controversial or divisive political issue that centers on conflicting international priorities: this year one of the highest rated concerns in global political news has been the outbreak of a military clash between Ukraine and Russia, turning the latter into the most sanctioned nation in the world. As US officials frame America's role in the war in ambitious terms and their strategic thinking in relation to Russia and Ukraine is quite transparent, it is the American discourse that was chosen as the material for analysis.

Thus, the search query for compiling Research Corpus 1 was American articles, matching the keyword "Russia" within the date range: 24 February 2022–present, with the result being 1194347 matching strings. Normally, the NOW corpus has a query system that allows for effective searching by word form, lemma, or part of speech, including frequency lists and collocates sorted by time period, creating n-grams lists, generating concordance lines, comparing one section of the corpus to another. As the authors had a restricted license and a limited access to the options of the corpus, including only the search by a keyword and a time period (because of the sanctions context mentioned above), they had to compile Research Corpus 2, cutting it down to 500 texts (580544 words), sorted by relevance (Fig. 1) for the data to be processed with modern technologies using the data processing software, similar to the one the Now corpus is normally equipped with.

LIST NAME ↕	# TEXTS ↕	# WORDS ↕	FIND KEYWORDS ● SPECIFIC ○ FREQ
RUSSIA	500	580,554	NOUN VERB ADJ ADV N+N ADJ+N

Figure 1. Research Corpus 2

Corpus 2 comprised breaking news and editorials from top US news agencies and periodicals (APN, CNBC, CNN, Chicago Tribune, Daily Beast, Forbes, Fox News, Newsday, New York Post, Politico, the Atlantic, the New York Times, the

Washington Examiner, the Wall Street Journal, USA Today, and others) to be further processed by LanA-Key program.

3.2. *LanA-Key automatic processing*

LanA-Key program is an easy-to-use application that requires neither special programming skills nor a preliminary stage of a corpus tagging, which serves as "an instrument for the automatic extraction of multi-component grammatically correct typed lexical phrases, featuring intelligent output and computationally attractive properties" (Sheremetyeva 2017) for various languages (Russian, English, German, French, Spanish, Chinese) and specific subject areas. The program was tuned to the extraction of nominal terminology from the English language political domain.

Processing the input string (Corpus 2), the program automatically lemmatized any word that can be unambiguously linked to a single headword[1] (Fig. 2) and generated a frequency list of lemmas, excluding numbers and functional parts of speech (Fig. 3. NP Frequency).
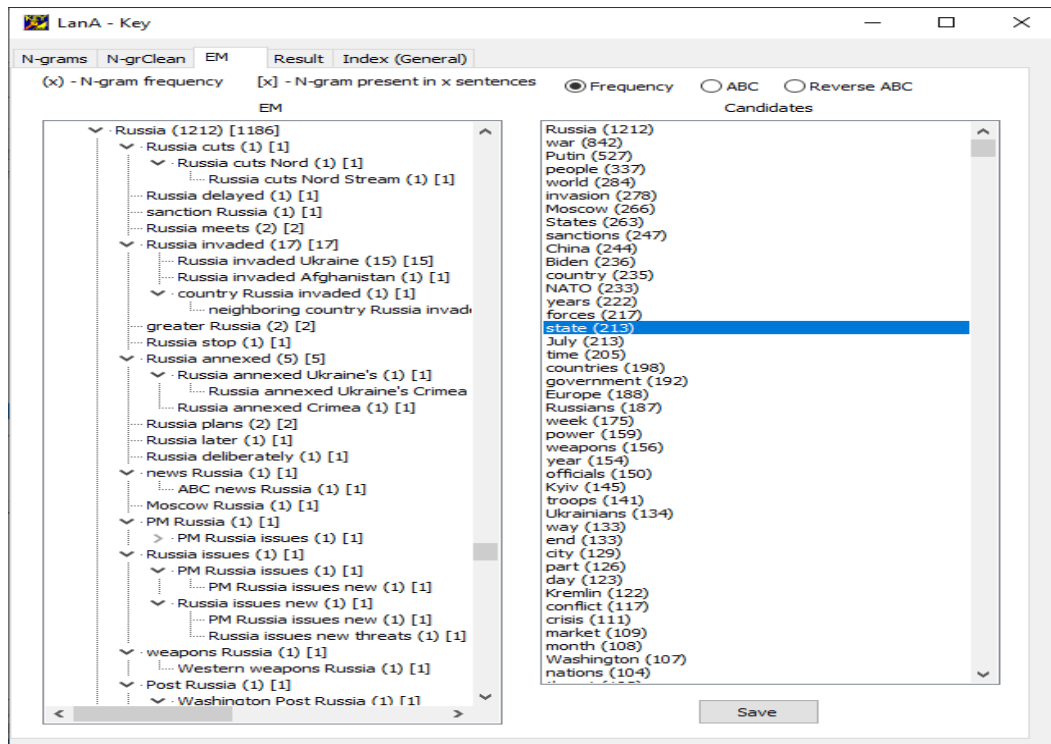
Figure 2. Fragment of LanA-Key processing

At the second stage the program sorted nouns and noun phrases (NPs) with lengths in 1–4 words alphabetically (Fig. 3. NP ABC list) and in reverse order

---

[1] The total number is 23485 lemmas, including numbers and functional parts of speech.

(Fig. 3. NP Reverse ABC list), the latter being relevant for fixing attributive clusters (an attributive cluster is understood here as a group of NPs with a key noun and different attributive components modifying it). The iterative process allowed for classifying meanings within the ontology and revealing new words that will be included in the dictionary because of their environmental relevance rather than their simple frequency.

```
1)              2)                              3)
[1224] Ukraine  [3]  Russian hegemony           [58] regime
[1212] Russia   [1]  Russian historical figures [1]  Assad regime
[857]  war      [4]  Russian history            [1]  Soviet-backed regime
[527]  Putin    [1]  Russian history shows      [1]  Kremlin-sponsored regime
[337]  people   [1]  Russian hostilities        [5]  unacceptable regime
[334]  country  [1]  Russian identity           [1]  end Ukraine's unacceptable regime
[299]  year     [1]  Russian imperial guard     [1]  oppressive regime
[295]  world    [1]  Russian imperialism        [1]  governing regime
[287]  Moscow   [1]  Russian imperialism        [1]  authoritarian regime
[278]  invasion [1]  Russian imperialism        [1]  Russian regime
[263]  State    [1]  Russian imperialist way    [5]  Putin regime
[258]  force    [2]  Russian import ban         [1]  President Vladimir Putin's regime
[247]  sanction [2]  Russian imports            [1]  President Alexander Lukashenko's regime
[247]  time     [3]  Russian influence operations [5] Putins regime
[244]  China    [1]  Russian influence shops    [1]  legitimize Putins regime
[236]  Biden    [2]  Russian information operations [1] entrenchment of Putins regime
[233]  NATO     [1]  Russian institutions       [1]  comprehensive sanctions regime
[213]  July     [1]  Russian intelligence services [3] autonomous sanctions regime
[213]  state    [2]  Russian interests          [1]  current sanctions regime
[211]  government [3] Russian interference      [1]  Soviet regime
[211]  week     [1]  Russian interference effort [4] puppet regime
[188]  Europe   [2]  Russian interior Ministry  [1]  Russian puppet regime
[187]  Russian  [1]  Russian intervention       [1]  Stalinist regime
                [1]  Russian intrusions         [1]  Kyiv regime
                [2]  Russian invaders           [1]  Moscow regime
                [38] Russian invasion           [1]  pro-Moscow regime
```

Figure 3. Fragments of 1) NP Frequency, 2) NP ABC, 3) NP Reverse ABC

3.3. Manual processing

The NP Frequency list with the lemmas 'Russia' (with 1212 entries, ranking second out of the total 23485 NPs) and 'Russian' (with 187 entries, ranking twenty first) shows that Corpus 2 is quite representative for further analysis. The process of compiling an initial glossary[2] for the dictionary is a sequence of several data reduction phases that require manual processing and linguistic analysis with the help of etymological, lexicographic, and semasiological methods. Each procedure takes the input (a list of words) from the previous one and feeds its output (a new list of words) to the next phase.

The procedures used were as follows:

1) List 1, comprising 859 NPs with 'Russia', 'Russian' used as attributes, was extracted from the NP ABC.

2) The data were compared with the NP Reverse ABC; the NPs that had no attributive clusters were excluded from List 2, which was reduced to 423 NPs. In addition to NPs with 'Russia' and 'Russian', used attributively, List 2 contained NPs with nouns modified by 'Kremlin' (134 entries), 'Moscow' (287), 'Putin' (527)

---

[2] The glossary presented in the paper is but the first step in compiling a more or less complete list of terms that will be included in the ECD of English Political Conflict Lexis.

as metonymic names of the country in political discourse, with some dictionaries considering them synonyms of Russia (Power Thesaurus).

3) To check the meaning of each key noun from List 2 six specialized dictionaries (both English and Russian) accessible via the Internet were used (Davletchina 2005, Dmitriev 2013, Merriman & Barrach-Yousefi 2021, Miller & King 2003, The Law Dictionary, Shipilov & Antsupov 2020). The available dictionaries present collections of words related to the concept of *conflict* and provide concise definitions of basic and specific terms, with one of the dictionaries (Dmitriev 2013) having a special mark of a 'conflictogene" (conflict-provoking concept / conflict trigger) for over a hundred of entries. Still, a number of important limitations to this pilot study need to be considered. First, the dictionaries have a different number of conflict terms: two small-sized dictionaries (Merriman & Barrach-Yousefi 2021, Miller & King 2003) contain fewer than 200 entries while, for instance, the Law Dictionary includes over 15,000 words. The second important limitation lies in the fact that, in our opinion, some words, registered in the dictionaries, neither have conflict-provoking connotations nor belong to the concept of conflict; for this reason, further lexicographic data collection is required to determine a final list of terms either with precise 'conflict' meaning in some uses or peculiar to the field of conflict studies. Third, when using Russian dictionaries of conflict studies, we focused mainly on a targeted search to choose between terms suggested by bilingual translation aids with the special subject area 'politics' (Multitran Dictionary, ABBYY Lingvo 12, Cambridge Dictionary). In order to validate the findings further study needs to be carried out through a focus group method that offers a more effective way of assessing the quality of the resulting dictionary. However, as it has already been mentioned in 2.3, few dictionaries have been able to draw on any systematic description of conflict terms, with none of them focusing on *political* conflict studies, which points to an urgent need for a renewed discussion of compiling a dictionary of the kind. Thus, notwithstanding the limitations, the six dictionaries were used for lexicographic analysis.

The criterion for including a term in the initial glossary was its having a particular meaning related to the field of conflict studies, registered in at least two of the dictionaries (Table 1; the asterisk (*) specifies the meaning / form / use of the word).

Table 1. **Fragment of lexicographic analysis**

Word (frequency)	Davletchina	Dmitriev	Merriman & Barrach-Yousefi	Miller & King	Shipilov & Antsupov	The Law Dictionary
Violence (58)	+	+ *conflict-provoking	+	- *non-violence	+	+
Sanction (398)	+	+ *used only in the singular	+	+	+	+

The key nouns of NPs excluded from List 3, which finally formed the glossary, fell into three groups:

a) basic lexemes (183): e.g., *action, behavior, choice, effort, hand, history, hospital, literature, option, people, treatment, version, vision*, etc.:

(1) *The town abuts the key highway that leads from the capital to western Ukraine and Lviv, so keeping it out of* **Russian hands** *is important in the effort to prevent Kyiv from being encircled* (The New York Times, 16.04).

b) discourse-specific lexemes (97), relating to politics and political discussion, registered in the specialized dictionary (Raymond 1930–1992): *ambassador, capital, civilization, constitution, diplomat, economy, leader, media, Minister, official, politics, President, state,* etc.:

(2) To deflect from *his failure to deter* **Russian President** *Vladimir Putin from invading Ukraine, President Biden has praised his anti-Russia coalition as a triumph of diplomacy* (The Hill, 17.03).

NPs with key nouns belonging to these two vocabulary tiers are used frequently over multiple contexts (e.g., the NP 'Russian President' has 527 entries in Corpus 2). Moreover, they often have apparent conflict-provoking negative connotations in the context. Still, following from synthesis-oriented MTT that "aims at speech production rather than speech understanding" (Mel'čuk & Milićević 2020: 8), pragmatic knowledge was not considered. At a later stage of the dictionary compilation lexemes belonging to the group will be analysed and considered for listing as the dictionary entries.

c) 'conflict-provoking' lexemes (54), not registered in the dictionaries: *exclusion, horde, intrusion, isolation, kleptocracy, massacre, nationalism, Putinism, regime, separatist, vassaldom,* etc.:

(3) *The autocratic* **Russian kleptocracy** *does not trust low-ranking and middle-ranking officers, and so cannot allow the imaginative, flexible decision making that NATO air forces rely upon* (The Atlantic, 09.05).

Key nouns of NPs, forming the group, typically or potentially have conflict-provoking connotations in their dictionary meanings. For example, the etymological and lexicographic analyses of the noun *kleptocracy* in (3) shows that the lexeme has a strong negative connotation: Latin *clepere* "to steal, listen secretly to" (cognate with Greek *kleptes* "a thief, a cheater"), used with a noun forming element *-cracy* (cognate with Latin *-cratia* "power, might; rule, sway; power over" and Greek *-kratia* "rule or government by") make up a term, denoting a corrupt political regime, characterized by widespread theft of its nation's wealth and resources, often practiced under an autocratic government that seeks status and personal gain at the expense of the governed (dictionary.com, Online Etymology Dictionary, Webster's, MacMillan). Hence, the noun denotes a form of leadership

universally seen as negative; describing any country as an example of a modern kleptocracy, closely associated with military states, oligarchies, and dictatorships, is a purposeful attempt to insult and humiliate the state, characteristic of conflict discourse. At a more advance stage of the project the lexemes forming the group will be analysed and considered for inclusion into the list of entries.

3.4. Glossary compilation

As a result of the data reduction procedures described above and illustrated in the table (Table 2), the overall amount of data was reduced in a trackable and safe manner that ensured that the entries in the final list are frequent and typical of the current conflict political discourse vocabulary.

Table 2. **Results of the data reduction procedures**

	Procedure / method	Source	Unit	Amount
1	LanA-Key automatic processing, quantitative analysis	NPs Frequency List	NPs with lengths in 1–4 words	23485
2	LanA-Key automatic processing, quantitative	NPs ABC	NPs with 'Russia', 'Russian' used attributively	859
3	analysis, manual processing	NP Reverse ABC	NPs having attributive clusters	423
4	Quantitative analysis, lexicographic, etymological, semasiological analyses	Six dictionaries on Conflict studies	Nouns registered in the dictionaries	88

Thus, List 3 includes 88 entries that remained after using the four reduction procedures. Each noun in List 3 contains its frequency in square brackets [count of its occurrences in Corpus 2], an attributive cluster of NPs with lengths in 2–4 words, and illustrative examples from the discourse:

(4) **invasion** [388]: Russia's invasion, Russia's invasion of Ukraine, Russia's full-scale invasion, Russia's bloody invasion, Russia's botched invasion, Russian despot's reckless invasion, Putin's invasion, Putin's initial invasion, Putin's invasion of Ukraine, Putin's latest invasion, Putin's calamitous invasion, Putin's invasion claims, Moscow's invasion, Moscow's invasion of Ukraine, Moscow's full-scale invasion. *Given the aftershocks of **the Russian invasion** of Ukraine and the relentless US campaign to isolate and punish the Kremlin, there may never have been a worse time to be an American imprisoned in Russia* (CNN, 07.07.2022).

Finally, an alphabetical list of 88 terms, frequent in current conflict political discourse and constituting the initial glossary of the ECD, was formed (Figure 4).

A	crackdown	G	R
abuse	conquest	genocide	regime
accusation	crackdown	H	revolution
aggression	crisis	hegemony	risk
ally	crime	hostility	S
ambition	cruelty	I	sanction
annexation	czar	imperialism	security
anger	D	influence	siege
army	damage	interference	spy
assault	danger	intervention	strategy
atrocity	despot	invasion	strike
attack	dictator	M	T
authority	disaster	menace	tactics
B	disinformation	mobilization	takeover
barrier	defense	O	terror
blockade	dissident	objection	theft
C	E	occupation	threat
campaign	empire	offense	tyranny
catastrophe	enemy	operation	U
challenge	escalation	P	ultimatum
coalition	expansion	policy	V
colonization	F	power	value
competition	fault	pressure	violation
conflict	fear	propaganda	violence
confrontation	force	protestation	W
control	front	provocation	war
			warfare

Figure 4. Glossary of the ECD of English Political Conflict Lexis

4. Results and Discussion

The ECD of English Political Conflict Lexis is generally compiled on the lines of the methodological principles of MTT, but it is much simplified if compared with the ECDs, fully consistent with MTT. As the proposed dictionary is meant not only "to boost the science of language" but also to satisfy practical needs, serving a particular public (linguists, discourse analysts, speechwriters, linguistic expertise specialists, etc.), it is "adapted to a particular level of understanding of its prospective users" (Mel'čuk 2006: 6). The aim of the authors is to test "the theoretical lexicon of a language", successfully applied to compiling the ECDs of Russian and French (Mel'čuk & Žolkovsky 1984, Mel'čuk et al. 1992), on the data of the English language, to be more particular, of American political conflict discourse.

MTT postulates that any act of linguistic communication involves three major entities: meaning (a content to be communicated by linguistic signals), text (a complex signal to be used to communicate the content), natural language (a mapping between meanings and texts) (Mel'čuk & Žolkovsky 1984, Mel'čuk & Wierzbicka 2018). The sets of infinite meanings and infinite texts are contrasted to the finite number of correspondences between them. The postulates of MTT are

applied to compiling the ECD, believed to be "a research tool, providing a standardized framework for the description of the lexical stock of any language" (Mel'čuk 2006, Mel'čuk & Milićević 2020: 203). The ECD deals with strictly linguistic meaning: the literal meaning of lexemes, which can be arrived at solely on the basis of linguistic knowledge, without any reference to the extralinguistic context or common sense (the data for the present study are collected automatically and the information is presented exclusively from the viewpoint of text synthesis, enabling the user to pass from a given meaning to the corresponding text, which is of primary importance when analyzing conflict discourse). Any ECD entry presents a full description of a lexeme, comprising three major divisions that correspond to the triple nature of the linguistic sign, with the signified and the signifier taken in their Saussurean sense, and with syntactics denoting "a set of properties that control its cooccurrence with other signs" (Mel'čuk 2006: 20). As MTT considers meanings and texts to be formal objects that can be described by means of formal languages and specified by formal devices, it puts forward a system of special techniques, symbols, abbreviations, writing conventions (for detailed treatment see Mel'čuk & Milićević 2020: 21–27), used for a rigorous and formalized description of the proposed ECD entry.

Thus, the microstructure of an ECD entry consists of three core zones, described in conformity with each other. The Semantic Zone contains the definition of the lexeme, its semantic label and connotations. The Phonological / Graphematic Zone specifies phonological features of the lexeme (its pronunciation, syllabification, and non-standard prosodic properties (if any)). The Cooccurrence Zone, divided into several sub-zones, presents combinatorial properties of the lexeme: 1) morphological (covering its inflectional paradigm), 2) syntactic (describing active and passive syntactic valence), 3) lexical (based on the notion of a lexical function, associated with semantic derivations of the lexeme paradigmatically and with its collocational potential syntagmatically), and 4) stylistic (including usage labels). The list of lexical functions includes those describing standard lexical paradigmatic relations: synonymy (`Syn`), antonymy (`Anti`), and conversion (`Conv`); derivational relations (`S, A`); syntagmatic lexical relations (adjectival functions (`Magn, Ver, Bon`), support verbs (`Oper`), phasal verbs (`Incep, Fin, Cont`), causative verbs (`Caus, Liqu`), some semantic derivations and several examples of complex lexical functions.

The source of the material for filling up the semantic and phonological zones as well as for stating the lexeme's morphological cooccurrence, stylistic cooccurrence and paradigmatic lexical functions is the dictionaries, while the lexeme's syntactic cooccurrence and syntagmatic lexical functions are studied using the corpus data of conflict political discourse. The use of the corpus data implies 1) checking the frequency of words in the corpus, 2) listing the collocations of the word, conventionally used together in conflict political discourse of the analyzed time span: N + the keyword, the keyword + N, Adj + the keyword, V+ the keyword (the output is a list of wordforms with significance scores for their co-

occurrence with the keyword, ordered from high to low significance) (Fig. 5), 3) comparing the collocates of two words to see how they differ in meaning and usage (Fig. 6). The procedure helps the lexicographer "to sharpen the focus of definitions, highlighting salient facts and omitting remote possibilities, and to formulate explicit rules for choosing among near synonyms" (Kruyt 1995).

HE LP	ⓘ	★	ALL FORMS (SAMPLE) : 100 200 500 WORDS	FR EQ	TOTAL 519 | UNIQUE 181 +
1	ⓘ	★	SANCTION PACKAGE	32	
2	ⓘ	★	SANCTION LIST	29	
3	ⓘ	★	SANCTION WEAPON	21	
4	ⓘ	★	SANCTION PUTIN	17	
5	ⓘ	★	SANCTION EVASION	16	
6	ⓘ	★	SANCTION RULES	14	
7	ⓘ	★	SANCTION TARGETS	14	
8	ⓘ	★	SANCTION ORDERS	11	
9	ⓘ	★	SANCTION LISTS	10	
10	ⓘ	★	SANCTION OLIGARCHS	10	
11	ⓘ	★	SANCTION MEASURES	9	
12	ⓘ	★	SANCTION POLICIES	9	
13	ⓘ	★	SANCTION INDIVIDUALS	8	

Figure 5. Fragment of the collocations list "SANCTION +Noun"

WORD 1 (W1): SANCTION (0.34)

	WORD	W1	W2	W1/W2	SCORE
1	RUSSIA	1164	639	1.8	5.4

WORD 2 (W2): PUNISHMENT (2.97)

	WORD	W2	W1	W2/W1	SCORE
1	RUSSIA	639	1164	0.5	0.2

WORD 1 (W1): SANCTION (0.12)

	WORD	W1	W2	W1/W2	SCORE
1	RUSSIA	1164	370	3.1	27.0

WORD 2 (W2): PENALTY (8.58)

	WORD	W2	W1	W2/W1	SCORE
1	RUSSIA	370	1164	0.3	0.0

WORD 1: COERCION

WORD	W1	W2
RUSSIA	51	50629

WORD 2: SANCTIONS

WORD	W2	W1
RUSSIA	50629	51

Figure 6. Fragment of comparing the keyword and its synonyms

An example of the simplified and user-oriented entry "SANCTION" as it is meant to appear in the ECD of English Political Conflict Lexis is presented below. The word has been chosen for the illustration as it is registered in all the dictionaries of conflict studies used for lexicographic analysis (Table 2). Though the Illustration Zone (exemplifying typical uses of the lexeme) is at times considered redundant, it is quite appropriate in the ECD of English Political Conflict Lexis as it makes it easier for the prospective user to understand a lexicographic description and substantiate the claims about conflict-provoking connotations of the lexeme; thus, the use of the entry lexeme in the corresponding zone (syntactic cooccurrence and syntagmatic lexical functions) is illustrated by actual sentences from the corpus.

SANCTION

The semantic zone

Definition: sanctions of X against Y concerning Z for W – coercive action/s concerning Z taken by country/ies X against the country Y which is considered to have violated international law in W to end the violation

Semantic label: coercive action

Connotation: punishment, ban, boycott, enforcement (negative, conflict-provoking)

The phonological zone

US /ˈsæŋkʃən/ UK /ˈsæŋkʃən/

The cooccurrence zone

(a) *Morphological cooccurrence*: Noun [C, usually plural], Politics;

(b) *Stylistic cooccurrence (usage label):* a conflict-provoking term used in political discourse, international law, and diplomatic discourse;

(c) *Syntactic cooccurrence:*

▪ *Active valence syntactic cooccurrence (Government Pattern)[3]*

X ⇔ I [who imposes sanctions]	Y ⇔ II [against whom the sanctions are imposed]	Z ⇔ III [concerning what the sanctions are imposed – the object of sanctions]	W ⇔ IV [the reason for sanctions]
1. Subjectival→Ncom(x) 2. Subjectival→NP(x) 3. Subjectival→NPabbr(x) 4. Indirect objectival→prep(by)/ NP/NPabbr(x)	1.Indirect-objectival →LprepN(y)	1.Modificative→AoL	1.Indirect-objectival/ modificative→ LprepN
5. Modificative→Nabbr(x)L 6. Modificative→Ngen(x)L 7. Modificative→NPabbr gen(x)L 8. Modificative→Ao(x)L 9. Determinative→Prnposs(x) L	2.Modificative→ Ao(y)L 3. Modificative→ N(y)L	2.Modificative→AoVedL	2.Indirect-objectival/ modificative→ LprepVingN

Examples:

X⇔I

CI.1 Subjectival→Ncom(x): ***America / Europe / the West:*** *Since **America** imposed **sanctions** on Russia over its invasion of Ukraine in late February, it has granted foreign bondholders an exemption to allow them to receive money from Moscow* (Yahoo Finance, 05.05).

CI.2 Subjectival→NP(x): ***the United States / European Union***: ***The United States*** *imposed **sanctions** on Russia: it is accused of gross human rights violations* (CNBC, 15.03).

---

[3] Though basically applying notations proposed by Mel'čuk, in the case of Government Pattern we use standard notations for parts of speech, their grammatical categories and syntactic functions accepted in Modern English Grammar (N – noun, com – common case, gen – genitive case, NP – noun phrase, NPabbr – abbreviated noun phrase, Prn – pronoun, Prnposs – possessive pronoun, A – adjective, o – zero degree of comparison, comp – comparative degree, superl – superlative degree, V – verb, Ved – participle II, Ving – participle I, gerund, VP – verb phrase, Vtrans – transitive verb, prep – preposition, etc. L stands for the entry lexeme).

CI.3 Subjectival→NPabbr(x): **the US / EU:** When **the US** imposed **sanctions** on Russia in April 2018—the US has a long history of imposing sanctions on Russia for a variety of reasons – the volatility in prices was similar (Yahoo Finance, 05.05).

CI.4 Indirect-objectival→prep(**by**)N/NP/NPabbr(x): **by the West / United States / EU:** With Moscow's war of aggression against Ukraine and the **sanctions** imposed **by the West**, these days seem to be over (Seeking Alpha, 16.06).

CI.5 Modificative→Nabbr(x)L: **US sanctions / EU sanctions:** **EU sanctions** against Moscow have snarled the fertilizer trade further (Politico, 11.08).

CI.6 Modificative→Ngen(x)L: **America's / West's / Biden's sanctions**: The new set of **Biden's sanctions** on Russia came in reaction to the country's invasion of Ukraine (Business Insider, 27.02).

CI.7 Modificative→NPabbrgen(x)L: **EU's sanction** list against Russia: Russia, blaming payment issues prompted by the latest round of **EU's sanctions**, cut off the flow of oil in a pipeline that runs through Ukraine (USA Today, 09.08).

CI.8 Modificative→Ao(x)L: **American / European / Canadian / Japanese / international / western** sanctions: Hungary's Prime Minister reiterated his opposition to **European sanctions** against Russia (The Telegraph on MSN.com, 16.09).

CI.9 Determinative→Prnposs(x)L: **the US – their** sanctions: **The West** lifts **its sanctions** against Russia and Russian companies, heightening a stand-off between Russia and the continent (Forbes, 06.09).

Y⇔II

CII.1 Indirect-objectival/modificative→LprepNy: *sanctions **against Russia** / sanctions **on Iran:** The Western **sanctions against Russia** were a reasonable non-violent response* (wral.com, 09.09).

CII.2 Modificative→Ao(y)L: **anti-Russian / Russian / Iranian sanctions:** With **Russian sanctions** in place on Russian oil, choices are limited (Market Watch, 15.08).

CII.3 Modificative→N(y)L: **Russia** sanctions: US unveils new **Russia sanctions**, implements ban on new imports of Russian gold (CNN, 28.06).

Z⇔III

CIII.1 Modificative→AoL: **economic / financial / diplomatic / political / disciplinary / nuclear sanctions:** Russia claims punitive **economic sanctions** imposed by the West are responsible for the indefinite halt to gas supplies via Europe (CNBC, 06.09).

CIII.2 Modificative→AoVedL: **nuclear-related** sanctions: Even if the U.S. lifts **nuclear-related sanctions** under a new deal, numerous other American sanctions on Iran would remain (Politico, 24.08).

W ⇔ IV

CIV.1 Indirect-objectival / modificative→LprepN: *sanctions **for violation** of the international law: Washington and its allies aim to cut off energy imports from Russia in line with sanctions **over its war on Ukraine*** (The Hill, 27.06).

CIV.2 Indirect-objectival / modificative→LprepVingN: *sanctions **for 'beginning' invasion:** European leaders have accused the Kremlin of using*

*its energy to punish and blackmail the bloc over sanctions **for invading Ukraine*** (Forbes, 22.08).

Possible complex syntactic cooccurrence

*CI.1-5+CII.1-3+CIII.1-2+CIV.1-2: As part of **American sanctions against Russia for its unprovoked invasion of Ukraine**, U.S. officials have stepped up efforts to seize the ruling elite's toys and assets* (The New York Times, 10.04).

Impossible

CII.1+CII.2-3: Russia/n sanctions against Russia

▪ ***Passive valence syntactic cooccurrence***

1. Direct-objectival→VtransL: ***impose / announce / evade / face / lift / avoid / enforce / tighten / violate / increase / apply / support / breach / bolster / bypass / ease / join / slap / implement/ introduce / strengthen / expand*** *sanctions: Russia will **face** additional **sanctions** and become even more of a global pariah than it is now* (The Washington Post, 03.08).

2. Indirect-objectival→VprepL: *speak **about** sanctions: The way we tend **to think about** sanctions is an alternative to war* (The New York Times, 01.04).

3. Modificative→LN: *sanction/s **package / list / weapon / rules / targets / regime / policy / strategy / tools / campaign / club**: Its owner is not on the E.U. **sanction list** even* (New York Magazine, 17.03).

4. Modificative/ Direct-objectival→LNverbal: *sanction **enforcement / violations / restrictions / breaches / exemptions**: Sanction restrictions – how the grain sector is adapting to the new reality* (CNN, 12.05).

5. Modificative→AoL: ***new / significant / severe / fresh / tough / strong*** *sanctions: A common ground between the United Kingdom and the U.S. has been its **heavy sanctions** against Russia* (Washington Examiner, 05.09).

6. Modificative→AcomL: ***tougher / harsher / stronger / stricter*** *sanctions: The NATO lawmakers approved a resolution calling for **stronger sanctions** against Russia* (The Denver Post, 30.05).

7. Modificative→AsupL: ***latest / toughest*** *sanctions: The White House is expected to allow **toughest sanctions** on controversial Russian gas pipeline* (CNN, 10.03).

8. Modificative→VingL: ***crippling / sweeping / existing / increasing / punishing / blocking*** *sanctions: It prompted **sweeping** economic **sanctions** against Russia and military support for Ukraine from Washington and its Western allies* (Reuters, 14.09).

9. Modificative→VedL: ***reimposed / imposed / increased / proposed / renewed / coordinated / limited*** *sanctions: He once again called for **increased sanctions** against Russia, including its entire banking sector and oil industry* (Washington Examiner, 18.04).

10. Modificative→Num-N-L: **12-point** sanction / **250$-a-day** sanction: *The EU extended its **six-month sanctions** against Russia* (Time, 09.09).

(d) *Lexical cooccurrence (lexical functions)*

Syn⊂	discipline
Syn⊃	punishment, coercion
Syn∩	penalty

Anti	support, permission, approval
Conv	sanction (in the meaning of approval)[4]
Gener	measure, policy
Figur	blockade, boycott
S_1	sanctioner
S_2	target [of sanctions]
A_1	sanctionable
A_2	sanctionative
A_3	sanctionless
A_4	sanctioned
V_0	sanction
S_{instr}	embargo, ban, restriction, prohibition
	*The next round of sanctions against Russia includes an oil **embargo** (The Guardian on MSN.com).*
S_{loc}	foreign policy, diplomacy, economy
S_{res}	negative effects, material costs
Sing	round of
	*Member nations are expected to approve another **round** of sanctions targeting Russia's defense industry and technology sector (Defense One, 11.07).*
Mult	list of, range of, set of
	*The Biden administration has banned energy imports as part of a **set of** sanctions against Mr. Putin (The New York Times, 08.04).*
$Loc_{in}Centr$	amidst
	*The Russian rouble has devalued in the wake of the invasion and **amidst** sanctions (Fox News, 07.03).*
A_i	under [sanctions]
	*Russian banks are **under** sanctions (The Nation, 06.08).*
Able	sanctionable, sanctioned
	*Department also issued guidance on its website warning that gold-related transactions involving Russia may be **sanctionable** by U.S. authorities (Reuters, 24.03).*
Magn	toughest, harshest, biggest, maximal
	*Russia now finds itself targeted by the **toughest** sanctions ever agreed against Moscow by the EU, US and UK (International Business Times, 26.02).*
[AntiBon + Magn]	lighter, lesser, mild
	*The Treasury department hit other institutions with **lesser** sanctions, including Alfa bank (New York Times, 01.03).*
$Magn^{temp}$	prolonged, continuous
	*European Commission said that the "reinforced, **prolonged** EU sanctions against the Kremlin" send "a strong signal to Moscow (The Washington Post, 16.03).*
$Magn^{quant}$	substantial, relevant
	*The U.S. and Europe have pledged to hold Moscow accountable, with more damaging and **substantial** sanctions against Russia now being readied by western powers (CNBC, 24.02).*

[4] Sanction has two opposite lexical meanings: it can refer to penalizing or disciplining someone or something, or to authorizing or approving something; it is sometimes known as a "Janus word".

Ver	punitive, disciplinary
	*The United States and European allies are looking to step up **punitive** sanctions on Russia for invading Ukraine* (Euronews.com, 06.03).
Bon	appropriate, proper, effective, powerful, *Washington should hit Moscow with the most **effective** sanctions possible against Russia's financial system* (The National Interest, 28.02).
Pred	sanction
	*Singapore will restrict military and high-tech exports to Russia and **sanction** four Russian banks* (Forbs, 03.05).
$Oper_1$	impose, place, put, issue, introduce
	*The U.S. and the U.K. plan to **impose** sanctions against Russia in reaction to the latest developments (*Business Insider, 25.02).
$IncepOper_1$	initiate, levy
	*The United States and European Union **have levied** sanctions on Russia's biggest banks and its elite* (TwinCities.com, 01.03).
$IncepInvolvOper_1$	join
	*Serbia rejects Western calls to **join** sanctions on Russia* (ABC, 25.02).
$ContOper_1$	keep, maintain
	*Evidence that Russian troops murdered hundreds of Ukrainian civilians is leading some U.S. lawmakers to insist that America and its allies **keep** sanctions on Moscow so long as Vladimir Putin remains in power* (Politico, 06.04).
$FinOper_1$	end
	*French far-right leader Marine Le Pen called for Europe **to end** sanctions against Russia to avoid a blackout* (YahooFinance, 02.08).
$LiquOper_1$	remove, waive, exempt
	*A growing number of Italians want **to remove** sanctions against Russia* (Politico, 05.09).
$Oper_2$	face, get
	*Get up to speed: Russia **faces** sanctions for 'beginning' invasion* (CNN, 24.02).
$ContOper_2$	be under
	*If Ukraine will fall, Putin will not stop... especially when Russia **is under** sanctions...* (Washington Examiner, 26.03).
$FinOper_2$	evade, escape, avoid
	*Metals Giant **Avoids** Sanctions. So far, the U.S. and its allies haven't sanctioned Russia's MMC* (MarketWatch, 07.03).
$LiquOper_2$	circumvent, breach, bypass, undermine
	*The U.S. and its allies were shutting down Russia's ability to use gold to **circumvent** sanctions* (The New York Times, 24.03).
$F_1 = IncepPred$ $Plus^{refl}$	extend, expand, lift, enforce, tighten, toughen, increase, escalate, strengthen, intensify
	*The EU must leave itself with an ability to significantly **tighten** and **escalate** sanctions should Russia hit back* (Politico, 31.03).

Though not fully living up to MTT's requirements and its well-developed lexicographic metalanguage, both the project of the proposed dictionary and the entry microstructure utilize some of its basic lexicographic concepts, having theoretical orientation and formalized character as its distinctive properties.

5. Conclusion

The results of the study are significant in at least three respects. Firstly, they contribute new findings and additional evidence, suggesting that computer technologies, including computational and corpus linguistics, are relevant to lexicography. The current findings add to a growing body of literature on computerized compilation of dictionaries, based on large electronic text corpora, proving that corpus-based technologies support lexicographical practice and enhance the quality and consistency of the resulting dictionary. Moreover, the analysis and interpretation of the corpus data by the researcher can be improved by automated linguistic analysis of language samples, which offers different views on the data by various types of sorting and rearranging options according to the criteria set by the lexicographer (frequency lists, reverse lists, collocations with the keyword, distribution over the sources, etc., supported by statistical tools). Secondly, the use of electronic text corpora offers an effective way for compiling a specialized discoursed-based dictionary that particularly focuses on a specific subject area, embedding the lexicographic data in the social, political, and geopolitical contexts. The specific subject area in the proposed dictionary merges political and conflict types of discourse as politics often becomes a platform for conflict deployment and conflict escalation. Thirdly, though based on the data of "language in context", being primarily an object of discourse *analysis,* the dictionary is designed on the principles of MTT from the viewpoint of text *synthesis*, successfully applied to its compilation (including both collecting the data and presenting the information about lexemes). However, the two are not mutually exclusive: the synthetic orientation of the dictionary does not prevent its use for further analysis. Thus, the ECD of English Political Conflict Lexis would cater to specific needs of its targeted user group and would be most helpful for scholars in linguistics, discourse analysis, media and communication, political science, and conflict studies.

The research which is the initial draft stage of compiling the ECD of Political Conflict Lexis has highlighted a number of questions that need to be addressed. They include quality assessment of programming tools, lexicographic sources and methods used, as well as the evaluation of the ECD itself and its microstructure, which calls for an increase in the list of entries and their more user-oriented description. We hope that the answers provided by our further studies will help us compile a full-fledged dictionary of modern political conflict lexis.

REFERENCES

Apresjan, Yurij D. 1974. *Lexical Semantics. Synonymic Means of Language*. Moscow: Nauka. (In Russ.)

Benson, Phil. 2002. *Ethnocentrism and the English Dictionary*. London & New York: Routledge.

Bergenholtz, Henning & Rufus H. Gouws. 2012. What is lexicography? *Lexicos* 22. 31–42.

Chen, Wenge G. 2015. Bilingual lexicography as recontextualization: A case study of illustrative examples in a New English-Chinese Dictionary. *Australian Journal of Linguistics* 35 (4). 311–333.

Chen, Wenge G. 2019. Towards a discourse approach to critical lexicography. *International Journal of Lexicography* 32 (3). 362–388.

DeCessaris, Janet. 2013. Incorporating context into dictionary presentation. In Olga M. Karpova (ed.), *Life beyond dictionaries*, 16–17. Ivanono: Ivanovo state university.

Demeshkina, Tatyana A. & Maria A. Tolstova. 2017. Gender dialectology and dictionaries as its source. *Russian Journal of Lexicography* 12. 83–105. https://doi.org/10.17223/22274200/12/5 (In Russ.)

Kruyt, Johanna G. 1995. Technologies in Computerized Lexicography. *Lexikos* 5. 117–137.

Fesenko, Olga P. 2015. Lexicography of language, speech, and discourse. *Omsk Scientific Bulletin* 1 (135). 52–54. (In Russ.)

Firth, John R. 1957. *Papers in Linguistics.* Oxford: Oxford University Press.

Gerasimova, Marina V. 2020. Culturological support of dictionary entries in the digital multilingual dictionary of metaphors. *Russian Journal of Lexicography* 17. 90–110. https://doi.org/10.17223/22274200/17/5 (In Russ.)

Golev, Nikolay D. 2019. A modern dialect vocabulary discourse dictionary (based on Runet materials): An innovative lexicographic project. *Russian Journal of Lexicography* 16. 113–137. https://doi.org/10.17223/22274200/16/7 (In Russ.)

Gromenko, Elizaveta S. 2020. The problem of term selection for a dictionary of neologisms (based on the material of contemporary periodical popular science publications)*. Russian Journal of Lexicography* 17. 5–25. https://doi.org/10.17223/22274200/17/1 (In Russ.)

Issers, Oksana S. 2015. *Discursive Practices of the Present Time*. Moscow: LENAND. (In Russ.)

Kachru, Braj B. 1995. Introduction. In Braj B. Kachru & Henry Kahane (eds.), *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta 64*, lxiii-lxvi. Tübingen: Max Niemeyer Verlag.

Koshkarova, Natalya N. & Olga A. Solopova. 2021. Cognitive and discursive approach to lexicography: How dictionary reflects place-naming system. In Elena V. Dzyuba & Svetlana A. Yeryomina (eds.), *Cognitive Strategies of Philological Education in Russia and Abroad* III, 69–73. Ekaterinburg: Ural state pedagogical university. (In Russ.)

Kozyrev, Vladimir A. & Valentina D. Chernyak. 2015. *Russian Lexicography: Present and Past Centuries.* Saint-Petersburg: A.I. Herzen RSPU Publishing House. (In Russ.)

Krasnobaeva-Chernaya, Zhanna V. 2020. The terminological data bank "Classification parameters of phraseological units" as an electronic terminographic product: Design experience. *Russian Journal of Lexicography* 18. 117–132. https://doi.org/10.17223/22274200/18/6 (In Russ.)

Manik, Svetlana A. 2019. *English Political Lexicography: Formation, Development, Present State.* Doctor of Philology thesis. Ivanovo: Ivanovo state University. (In Russ.)

Mel'čuk, Igor A. 1995. *The Russian language in the Meaning-Text perspective*. Wien: Wiener Slavistischer Almanach. Moscow: Languages of the Russian Culture School. (In Russ.)

Mel'čuk, Igor. 2006. Explanatory combinatorial dictionary. In Giandomenico Sica (ed.), *Open problems in linguistic and lexicography*, 225–355. Monza (Italy): Polimetrica.

Mel'čuk, Igor A. & Anna Wierzbicka. 2018. Semantic decomposition, and the Meaning-Text approach. *Russian Journal of Linguistics* 22 (3). 521–538. https://doi.org/10.22363/2312-9182-2018-22-3-521-538

Mel'čuk, Igor A. & Jasmina Milićević. 2020. *An Advanced Introduction to Semantics: A Meaning-Text Approach*. Cambridge: Cambridge University Press.

Morkovkin, Valerij V. 1977. *Ideographic Description of Lexis: A Case Study (Analysis of Words with Temporal Meaning in Russian)*. Moscow: Moscow University Publishing House. (In Russ.)

Nielsen, Sandro H. 2013. Foundations and key components of specialized lexicography. In Olga M. Karpova (ed.), *Life beyond dictionaries*, 24. Ivanono: Ivanovo state university.

Plotnikova, Anna M. 2014. New trends in Russian lexicographic discourse. *Novosibirsk State Pedagogical University Bulletin* 3 (19). 22–29. (In Russ.)

Radbil, Timur B. & Vadim V. Saigin. The concept "grekh" ("sin") in the context of the lexicographic description of key concepts of Russian culture. 2019. *Russian Journal of Lexicography* 15. 36–59. https://doi.org/10.17223/22274200/15/3 (In Russ.)

Sandomirskaya, Irina. 2001. *The Book about Motherland: Experience of Discursive Practices Analysis.* Wien: Wiener Slawistischer Almanach Sonderband. (In Russ.)

Sheremetyeva, Svetlana O. 2017. *Linguistic Models and Tools for Processing Patent Claims.* Chelyabinsk: SUSU Publishing Centre.

Skovorodnikov, Alexander P. & Galina A. Kopnina. 2021. On the conception of the modern political labels dictionary. *Russian Journal of Lexicography* 20. 105–118. https://doi.org/10.17223/22274200/20/6 (In Russ.)

Solovyev, Valery, Marina M. Solnyshkina & Danielle M. McNamara. 2022. Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics* 26 (2). 275–316. https://doi.org/10.22363/2687-0088-30161

Van Dijk, Teun A. 1998. *Ideology: A Multidisciplinary Approach.* London: Sage.

Vlavatskaya, Marina V. 2013. *Theoretical Foundations of Combinatorial Linguistics: Lexicological and Lexicographic Aspects*. Doctor of Philology thesis. Novosibirsk: Novosibirsk state technical university. (In Russ.)

Williams, Geoffrey. 2013. Blast from the past: Meaning change through collocational resonance. In Olga M. Karpova (ed.), *Life beyond dictionaries*, 25–26. Ivanono: Ivanovo state university.

Wodak, Ruth & Michael Meyer. 2015. Critical Discourse Studies: History, Agenda, Theory and Methodology. In Ruth Wodak & Michael Meyer (eds.), *Methods of Critical Discourse Studies*, 18–50. London: Sage.

Zemicheva, Svetlana S. 2017. A conception of the electronic thesaurus of a dialect language personality. *Russian Journal of Lexicography* 12. 24–38. https://doi.org/10.17223/22274200/12/2 (In Russ.)

Zholkovsky, Alexander K. & Igor A. Mel'čuk. 1967. On semantic synthesis. *Problems of Cybernetics* 19. 177–238. (In Russ.)

Dictionaries and other sources

ABBYY Lingvo 12. Retrieved from https://www.lingvolive.com/en-us. (accessed 1 July 2022).

BBICDE 1990 – Benson, Morton, Evelyn Benson & Robert Illson. *The BBI Combinatory Dictionary of English*. Special edition (USSR). Moscow: Russian language.

Cambridge Dictionary. Retrieved from dictionary.cambridge.org. (accessed 1 July 2022).

CDDLP 2006–2012 – *Complete Dictionary of a Dialect Language Personality*. Vol. 1–4. Tomsk: Tomsk University Publishing House.

Con 2010 – *Conceptosphere of the Russian Language: Key Concepts and their Representations (based on lexis, phraseology and paroemiology): Project of a Dictionary*. Ekaterinburg: Ural University Publishing House. (In Russ.)

Davletchina, Svetlana B. 2005. *Dictionary of Conflictology*. Ulan-Ude: VSSTU Publishing House. (In Russ.)

DCRL 2002 – *Dictionary of Combinability of the Russian language*. Denisov, Pyotr N. & Valerij V. Morkovkin (eds.). Moscow: Astrel: AST. (In Russ.)

Dictionary.com. Retrieved from https://www.dictionary.com. (accessed 23 July 2022).

Dmitriev, Anatoly V. 2013. *Conflict studies: Dictionary of Terms*. St. Petersburg: SPbGUP. (In Russ.)

DMLR 1948–1965 – *Dictionary of Modern Literary Russian*. Vol. 1–17. Moscow: USSR Academy of Sciences; Nauka. (In Russ.)

DRELI 2007 – Oubin, Ivan I. *Dictionary of Russian and English Lexical Intensifiers*. Moscow: R. Valent.

Karaulov, Yurij N et al. 2002. *Russian Associative Dictionary*. Vol.1–2. Moscow: AST: Astrel. (In Russ.)

Lacroix, U. 1956. *Les mots et les idees. Dictionnaire des termes cadrant avec les idees*. Paris: Fernand Nathan.

MacMillan. Retrieved from https://www.macmillandictionary.com. (accessed 23 July 2022).

Mel'čuk, Igor A. & Alexander K. Zholkovsky. 1984; 2016. *Explanatory Combinatorial Dictionary of Contemporary Russian*. Vienna: Wiener Slawisticsher Almanach 14; Moscow: Global Com: Languages of the Slavonic culture. (In Russ.)

Mel'čuk, Igor, Nadia Arbatchewsky-Jumarie, Lidija Iordanskaja & Suzanne Mantha. 1992. *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques III, Montréal: Les Presses de l'Université de Montréal.

Miller, Christopher A. & Mary E. King. 2003. *A Glossary of Terms and Concepts in Peace and Conflict Studies*. San Jose, Costa Rica: Universidad para la Paz.

Merriman, Hardy & Nicola Barrach-Yousefi. 2021. *Glossary of Civil Resistance: A Resource for Study and Translation of Key Terms*. ICNS Press: International Center on Nonviolent Conflict.

ML 2005, 2009, 2014 – *Man and his Language. Materials of the Complex Linguacultural Learner's Dictionary*. Issues 1–4. Nizhny Novgorod: Nizhegorodsky University Publishing House.

Multitran Dictionary. Retrieved from https://www.multitran.com/en/ru. (accessed 23 June 2022).

NEDRS 2000 – Apresjan, Yurij D. et al. *New Explanatory Dictionary of Russian Synonyms*. Moscow: Languages of the Russian Culture. (In Russ.)

NOW – News on the Web. Retrieved from https://www.english-corpora.org/now/. (accessed 13 February 2022).

Ozhegov, Sergej. I. 1949. *Dictionary of the Russian Language*. Moscow: State Publishing House of Foreign and National Dictionaries. (In Russ.)

Online Etymology Dictionary. Retrieved from https://www.etymonline.com. (accessed 17 May 2022).

Power Thesaurus. Retrieved from https://www.powerthesaurus.org. (accessed 25 June 2022).

Raymond, Walter J. (1930–1992). *Dictionary of politics: Selected American and foreign political and legal terms*. Lawrenceville, Va.: Brunswick Publishing Company.

Shakhmatov 1898 – *Dictionary of the Russian Language*. Shakhmatov, Aleksey A. (ed.). Saint-Petersburg: The Imperial Academy of Sciences Printing House. (In Russ.)

Shipilov, Anatoly I. & Anatoly Y. Antsupov. 2020. *Dictionary of Conflict Studies*. Moscow: Prospect. (In Russ.)

The Law Dictionary. Retrieved from https://dictionary.thelaw.com. (accessed 1 July 2022).

Ushakov 1935–1940 – *An Explanatory Dictionary of the Russian Language*. Ushakov, Dmitry. N. (ed.). Vol. 1–4. Moscow: State Institute "Soviet Encyclopedia". (In Russ.)

VD 1998–2002 – *Vershininsky Dictionary*. Vol. 1–7. Tomsk: Tomsk University Publishing House. (In Russ.)

Webster's. Retrieved from https://www.merriam-webster.com. (accessed 23 July 2022).

Bionotes:
Olga SOLOPOVA is Dr Habil. in Philology, Professor at the Department of Linguistics and Translation at the Institute of Linguistic and International Communications of South Ural State University (National Research University). Her research interests include lexicography, typological linguistics, grammar, metaphor studies, discourse analysis, political linguistics, and diachronic linguistics.
e-mail: o-solopova@bk.ru
https://orcid.org/0000-0003-4170-7267

Tamara KHOMUTOVA is Dr Habil. in Philology, Professor at the Department of Linguistics and Translation Studies at the Institute of Linguistics and International Communications of South Ural State University (National Research University). Her research interests include discourse analysis, integral linguistics, lexicography, political linguistics, cultural linguistics, and corpus linguistics.
e-mail: khomutovatn@susu.ru
https://orcid.org/0000-0001-5244-3960

Сведения об авторах:
Ольга Александровна СОЛОПОВА – доктор филологических наук, профессор кафедры лингвистики и перевода Института лингвистики и международных коммуникаций Южно-Уральского государственного университета (национального исследовательского университета). Ее научные интересы включают лексикографию, типологическую лингвистику, грамматику, изучение метафор, анализ дискурса, политическую лингвистику и диахроническую лингвистику.
e-mail: o-solopova@bk.ru
https://orcid.org/0000-0003-4170-7267

Тамара Николаевна ХОМУТОВА – доктор филологических наук, профессор кафедры лингвистики и перевода Института лингвистики и международных коммуникаций Южно-Уральского государственного университета (национального исследовательского университета). Ее научные интересы включают анализ дискурса, интегральную лингвистику, лексикографию, политическую лингвистику, лингвокультурологию и корпусную лингвистику.
e-mail: khomutovatn@susu.ru
https://orcid.org/0000-0001-5244-3960

Research article / Научная статья

Russian interrogative pronouns as a lexicographic type

Valentina APRESJAN[1,2] ⊠ and Boris IOMDIN[2]

[1] HSE University, Moscow Russia
[2] Vinogradov Russian Language Institute; Moscow, Russia
⊠vapresyan@hse.ru

Abstract

Our study tackles Russian interrogative-relative pronouns (*wh*-words) as a lexicographic type which requires a unified treatment. Our objective is to give a systematic description and explanation of the numerous collocational and constructional properties of the Russian *wh*-words using lexicographic and corpus methods. The dataset and statistics were extracted from the Russian National Corpus, at least 100 examples for each of the pronouns were analysed. Methodologically the study is based on the principles of the Moscow School of Semantics (namely, integral description of language and systematic lexicography) which are to a large extent rooted in the "Meaning⇔Text" theory. They include analysis of linguistic items on all levels of language; a focus on their semantic and combinatorial properties; acknowledged validity of dictionary as an instrument of linguistic research. The paper considers semantic, syntactic and co-occurrence properties shared by many Russian interrogative pronouns and analyzes the reasons for their almost entire lack in the pronouns *zachem* 'what for' and *pochemu* 'why'. As demonstrated in the study, most of the constructional and co-occurrence properties typical of Russian interrogative pronouns (for example, co-occurrence with particles *imenno* 'exactly' and *khot'* 'at least', constructions with *mnogo* 'many', *malo* 'few', etc.) are motivated by the semantics of multiplicity and choice, which are incompatible with 'what for' and 'why'. In addition, as the findings show, different interrogative pronouns have different frequencies of occurrence in the described constructions, which is explained not by their general corpus frequencies or by the animacy hierarchy, but by the compatibility of their semantics with the meanings of multiplicity and choice. The obtained results suggest that combinatorial properties of *wh*-words are motivated by their semantics which, in turn, reflects the meta-linguistic characteristics of the situations to which they refer.

Keywords: *Russian interrogative pronouns, wh-words, co-occurrence, construction, lexicographic type*

Русские вопросительные местоимения как лексикографический тип

В.Ю. АПРЕСЯН[1,2] ✉, Б.Л. ИОМДИН[2]

¹ Национальный исследовательский университет «Высшая школа экономики»
² Институт русского языка им. В.В. Виноградова РАН
✉vapresyan@hse.ru

Аннотация
Работа посвящена русским вопросительно-относительным местоимениям (или *к*-словам) как лексикографическому типу, требующему единого словарного описания. Цель статьи – системно описать и объяснить многочисленные сочетаемостные и конструкционные свойства, характерные для русских вопросительно-относительных местоимений, с использованием лексикографических и корпусных методов. Корпусный материал и статистические данные извлечены из Национального корпуса русского языка; анализируемый объем материала от ста и более вхождений на каждое из местоимений. Методологически мы опираемся на принципы Московской семантической школы, которые во многом восходят к теории «Смысл ⇔ Текст». Важнейшие принципы теории «Смысл ⇔ Текст», унаследованные МСШ, – это установка на многоуровневое языковое описание, внимание к семантическим и комбинаторным свойствам языковых единиц, ценность словаря как инструмента лингвистического исследования (в трактовке МСШ – интегральное описание языка и системная лексикография). В работе описываются общие семантические, синтаксические и сочетаемостные свойства, имеющиеся у многих русских вопросительных местоимений, и анализируются причины отсутствия большей части этих свойств у местоимений *зачем* и *почему*. Как продемонстрировано в работе, большая часть конструкционных и сочетаемостных свойств, характерных для русских вопросительных местоимений (например, сочетаемость с частицами *именно* и *хоть*, образование синтаксических фразем вида *редко кто, мало + к*-слово*, много + к*-слово и т. п.), мотивирована семантикой множественности и выбора, которая несовместима со значением *зачем* и *почему*. Кроме того, как показывает наше исследование, разные вопросительные местоимения обладают разной встречаемостью в описанных конструкциях, что объясняется не их общей корпусной частотностью или же иерархией одушевленности, а совместимостью их семантики со значениями множественности и выбора. Полученные результаты свидетельствуют о наличии глубокой мотивирующей связи между семантическими свойствами вопросительных местоимений и их комбинаторными способностями; первые, в свою очередь, связаны с внеязыковыми характеристиками ситуаций, к которым относятся разные местоимения.
Ключевые слова: *русские вопросительные местоимения, сочетаемость, конструкции, лексикографический тип*

Люша и Боря – Игорю

Мельчук – некоторая константа моей жизни, как мои родители. Он был всегда. И хотя большую часть жизни я знаю замечательного ученого, великого лингвиста Игоря Александровича Мельчука, о котором уже много лет рассказываю студентам, об Игоре я думаю в первую очередь как о замечательном человеке, друге моих родителей и всей нашей семьи – великом и при этом простом и доступном, лишенном какого бы то ни было снобизма, добром, веселом, заботливом, надежном, который всегда рядом, всегда греет своим огнем и теплом, пишет смешные письма в стихах и может подписаться «твой старик Похабыч». Почему-то особенно вспоминаются два эпизода. Один – когда я жила у Мельчуков в Канаде в 1990 году, учась на курсах английского, и Игорь сам готовил мне завтрак и беспокоился о том, чтобы я как следует поела. А второй – когда Игорь с Лидой приезжали в Москву и Игорь прочел лекцию для студентов-лингвистов у нас в Вышке, а прочтя, запрыгнул на стол. Кажется, студенты были приятно поражены тем, что классик лингвистики так хорошо прыгает. На других лекциях они с таким не сталкивались. Игорь, я Вас очень люблю. И желаю, чтобы Вы еще много-много лет могли запрыгивать на столы, цвести и пахнуть (пользуясь Вашими же выражениями). Наяривайте и нахоцуписяривайте! Ваша Люша

Дорогой Игорь Александрович! Ваш талант и драйв могут пленить *кого угодно. Мало кто* умеет с одинаковым рвением сочинять великие лингвистические труды и учить детей складывать бумажные самолетики. *Редко когда* приходится слышать, как человек, чье имя с благоговением повторяют студенты разных поколений, в лесной прогулке по окрестностям Монреаля вдохновенно сочиняет страшилки: «Маленький мальчик грибочков поел». *Хоть кто* оценит Ваш эпистолярный стиль (вот прекрасное пожелание в одном из писем мне: «Вперёд, к!») Этот маленький текст я написал не **абы почему* и не **почему попало*, а чтобы поздравить Вас со знаменательным юбилеем и пожелать Вам ещё долго радоваться жизни и радовать нас! Ваш Боря

1. Введение

В последние десятилетия растет интерес к комбинаторным (лексико-сочетаемостным, конструкционным) свойствам языковых единиц и к многоуровневому лексикографическому описанию как действенному инструменту анализа языка, который в отечественной лингвистике впервые был отчетливо продекларирован в теории «Смысл ⇔ Текст» (Мельчук 1974, Мельчук и др. 1984, Мельчук 1995). Подход Московской семантической школы, во многом

продолжающей и развивающей идеи и принципы теории «Смысл ⇔ Текст», дополнительно предполагает поиск семантической мотивации языковых свойств, и в том числе объединение языковых единиц в так называемые лексикографические типы, т.е. классы лексем, которые в силу своих общих лингвистических свойств требуют единообразного лексикографического описания (Апресян 2009).

Наша работа посвящена русским вопросительным местоимениям (РВМ) как лексикографическому типу. В ней рассматриваются общие и различающиеся свойства так называемых *к*-слов (*кто, когда, куда, где* и т. п.) на всех языковых уровнях: семантика (в том числе полисемия), конструкции, сочетаемость. Известно, что РВМ обладают исключительно богатыми сочетаемостными и конструкционными свойствами, в том числе порождают большое количество синтаксических фразем (Иомдин 2007, Iomdin 2007). Однако разные РВМ в разной степени способны к употреблению в тех или иных конструкциях и контекстах. Насколько нам известно, не существует работ, в которых бы системно и на корпусном материале рассматривались комбинаторные и прочие языковые свойства РВМ в контексте их семантических различий и давалось бы семантическое объяснение тому, что их различает в лексико-сочетаемостном, конструкционном, фразеологическом плане. Цель данного исследования – системно описать и объяснить многочисленные сочетаемостные и конструкционные свойства, характерные для русских вопросительно-относительных местоимений, с использованием лексикографических и корпусных методов.

2. Проблемы описания вопросительно-относительных местоимений как лексикографического типа

Многие из рассматриваемых в работе языковых явлений в том или ином виде анализировались и упоминались в литературе, посвященной семантическим, сочетаемостным и синтаксическим свойствам русских вопросительно-относительных местоимений, в том числе юбиляром. Не ставя себе целью перечислить все работы на данную тему, упомянем лишь некоторые[1]: (Кобозева 1988) о типах вопросов и вторичных функциях вопросительных предложений, (Апресян 1995) об эксскламативных употреблениях вопросительных слов *как* и *какой*, (Булыгина, Шмелев 1997, Падучева 2010, 2018) о невопросительных употреблениях вопросительных предложений, (Апресян 2005) о конструкциях с вопросительными местоимениями *как* и *какой* и параметрическими прилагательными, (Мостовая 2009) об иллокутивных особенностях и семантических функциях конструкций с местоимениями *зачем* и *почему*, (Зевахина 2018) об эксскламативных употреблениях местоимений *как, сколько*

---

[1] Более полный список литературы о вопросительных местоимениях можно найти в работе: Чжан Юэ. Семантика и функции вопросительных предложений со словами «где», «куда» и «откуда»: дисс. … к. ф. н. М., 2019.

и *какой*, (Левонтина, Шмелев 2005, 2018) о местоименных сериях на *угодно, попадя, абы* и пр., (Тестелец, Былинина 2005) об «амальгамах» со значением неопределенности на основе вопросительных местоимений, (В. Апресян 2007) о конструкциях с *мало (ли)* и вопросительными местоимениями, (Кустова 2016, Кустова, Добровольский 2020, В. Апресян, Копотев 2021, 2022) о биместоименных комплексах с вопросительными местоимениями, словарная статья слова *где* М.Я. Гловинской в АС, (Иорданская, Мельчук 2020) об относительном употреблении местоимений *кто* и *который*.

Однако большая часть работ фокусируется на каком-то одном семействе конструкций, на одном типе употреблений, на одном или нескольких местоимениях. При этом обращает на себя внимание тот факт, что РВМ представляют собой достаточно компактный лексикографический тип. Во-первых, у них выделяются общие тенденции в полисемии: так, у большинства РВМ выделяются, помимо вопросительных значений, также значение 'известный говорящему Х' (*Так вот кто тебе помогал!*; *Так вот что ты вчера делал!*; *Куда стоит сходить, так это в Музей импрессионизма*) и значение 'Одни Х-ы так, другие Х-ы иначе' (*У кого пятерка за экзамен, а у кого и двойка*; *Где топят, а где еще не начинали*). Во-вторых, у них есть общие синтаксические свойства – так, многие РВМ могут вводить определительные придаточные (*Вот мой друг, о ком я тебе рассказывал*; *Вот шкаф, где я храню свои вещи*). Наконец, у многих РВМ есть богатейшая конструкционная и лексическая сочетаемость, например, с конструкциями вида *мало кто <что, где>, редко кто <когда, какой>*, с конструкцией редупликации (*Уж кто-кто, а он знает*), с дистрибутивной конструкцией (*кто куда, когда как*), с лексемами *именно, же* (*кто именно, что же* и др.). Однако имеются между ними и различия: например, для местоимений *зачем* и *сколько* нехарактерны определительные употребления (невозможно **Вот цель, зачем я это сделал*, **Вот сумма, сколько я заплатил*), конструкция *мало почему* практически не встречается, в то время как конструкция *мало кто* весьма частотна. Представляется, что для того, чтобы описать РВМ как единый класс слов, с их общими свойствами и индивидуальными различиями, необходим системный анализ массового материала (всех местоимений во всех релевантных употреблениях) с использованием корпусных данных. Однако системного и полного разноуровневого лексикографического описания всех РВМ, выполненного с учетом современных корпусных технологий, в литературе не существует. Кроме того, словари (МАС, БАС, БТС, СУш, СОШ) также не предлагают последовательного описания всех свойств РВМ как лексикографического типа. Таким образом, на данный момент в русистике нет общего представления об устройстве данного лексикографического типа, который является одним из важнейших и центральных не только в русском языке, но и в языке вообще.

Соответственно, мы поставили цель создать единый лексикографический подход к описанию РВМ. Предложенные решения опираются на описания,

составленные В.Ю. Апресян, Ю.Д. Апресяном, М.Я. Гловинской, А.В. Птенцовой, Е.В. Урысон для Активного словаря русского языка². Мы объясняем различия в свойствах разных РВМ на основе различий в их основном – вопросительном – значении. Полученные результаты, как ожидается, могут быть применимы не только к русскому языку, но в какой-то мере и к другим языкам.

3. Методы и материалы

В данной работе мы пользуемся лексикографическими и корпусными методами. Корпусный материал и статистические данные извлечены из Основного корпуса НКРЯ. Методологически мы опираемся на принципы Московской семантической школы (МСШ), в частности на интегральное описание языка и системно-лексикографический подход, предложенные в (Апресян 2010) для Активного словаря русского языка. Многие принципы МСШ восходят к теории «Смысл ⇔ Текст», создателем которой является юбиляр. Важнейшие принципы теории «Смысл ⇔ Текст», унаследованные МСШ, – это установка на многоуровневое языковое описание, внимание к семантическим и комбинаторным свойствам языковых единиц, ценность словаря как инструмента лингвистического исследования (в трактовке МСШ – интегральное описание языка и системная лексикография). Мы опираемся на лексикографический подход, предложенный в (Апресян 2010) для Активного словаря русского языка, а также на корпусное исследование, выполненное на материале Основного корпуса НКРЯ, для того чтобы создать единый шаблон описания РВМ в той мере, в которой их свойства пересекаются. Мы рассматриваем местоимения *кто, что, как, где, куда, откуда, когда, какой, чей, сколько,* а также во многих отношениях противопоставленные им местоимения *зачем* и *почему*[3], но не местоимение *который*, поскольку в современном русском языке оно используется в основном в качестве относительного. Для каждого из этих местоимений исследуется потенциал развития полисемии, а также сочетаемостные и конструкционные свойства, а именно возможность появления в тех или иных лексических и синтаксических контекстах, на основе корпусных данных.

---

² В силу недостатка места мы не приводим здесь образцы описаний, опирающиеся на единый, выработанный нами шаблон: все описания вопросительных местоимений можно найти в соответствующих выпусках АС.

³ Наш анализ *зачем* и *почему* по сравнению с другими РВМ в очень малой степени пересекается с анализом, предложенным в работе (Мостовая 2010), в которой иллокутивные ограничения и возможности конструкций с этими местоимениями (т.е. их функционирование в качестве прямых и риторических вопросов, возражений и пр.) мотивируются пропозициональной природой целевого и причинного значения. Нашей областью интересов являются семантические ограничения на лексическую и конструкционную сочетаемость *зачем* и *почему*, вызывающие невозможность их употребления в большей части конструкций, характерных для русских РВМ, что мы объясняем семантикой заданности, имплицитно содержащейся в их значениях.

4. Результаты и дискуссия

Мы получили два основных результата: во-первых, системное описание свойств РВМ как единого лексикографического типа, во-вторых, семантическое объяснение многих из этих свойств. Мы описали разные употребления, а также разные сочетаемостные и конструкционные свойства РВМ и предложили семантические объяснения тому факту, что для разных местоимений характерна разная степень допустимости и частотности реализаций в тех или иных контекстах и конструкциях. Ниже описываются разные свойства РВМ, приводятся корпусные данные относительно их проявления у разных местоимений, а также излагаются соображения, объясняющие различие в этих проявлениях, в первую очередь противопоставление местоимений *зачем* и *почему* остальным.

4.1. Полисемия

По нашим наблюдениям, у большинства РВМ есть значения «характеризации», в которых они вводят определительные придаточные, ср. например:

(1) *Вот та, кого я люблю*;
(2) *Вот место, где мы познакомились*;
(3) *Вот шкаф, куда можно все сложить.*

Однако у местоимений *зачем*, *почему* и *сколько* есть значение 'известный говорящему', которое, возможно, развилось из определительной конструкции, но сами определительные употребления отсутствуют:

(4) *Вот зачем он позвонил* (но не **Вот цель, зачем он позвонил*);
(5) *Вот сколько он за это заплатил* (но не **Вот количество, сколько он за это заплатил*).

Определительные употребления применительно к цели, количеству и причине выражаются аналитически при помощи местоимения *который*:

(6) *Вот цель, с которой он позвонил*;
(7) *Вот причина, по которой он не звонит.*

Такого рода замены на *который* возможны и часто предпочтительны и у других РВМ, но для перечисленных трех они обязательны.

Помимо «характеризующего» значения у *зачем* и *почему* отсутствует уходящее значение неопределенного нереферентного местоимения ('-нибудь'), которое представлено, в той или иной степени и по крайней мере потенциально, у всех остальных РВМ:

(8) *Может, тебе нужно что?*;
(9) *Может, кто что вспомнит?*

Так, невозможно или странно сказать **Он не пришел почему?* [в значении 'по какой-то причине']; **Он это сделал зачем?* [в значении 'с какой-то

целью']; *?Может, он опоздал почему*; *?Может, он уехал зачем*. Степень (не)приемлемости зависит от конструкции. В вопросе такое употребление совершенно невозможно, в конструкции-предположении с модальным словом – маргинально. По-видимому, это связано и с семантикой местоимений, и с характером вопроса, и со сферой действия операторов, и с некоторыми общими прагматическими соображениями. Вопросы подобного типа, несмотря на наличие вопросительного местоимения, являются общими. Во фразах вида *Тебе ПОМОГАЛ кто?* вопрос относится к глаголу, который находится в фокусе, а местоимение – часть темы, в отличие от специальных вопросов вида *КТО тебе помогал?*, где глагол в теме, а местоимение в фокусе. Таким образом, семантической сферой действия вопроса является наличие человека, который помогал ['Был ли кто-то, кто тебе помогал?']. Если подставить в подобные вопросы местоимения *почему* и *зачем*, то сферой действия вопроса, т.е. областью сомнения, будет наличие причины у ситуации (**Он не ПРИ-ШЕЛ почему?* ['Была ли причина, по которой он не пришел?']) или цели у действия (**Он ПОЗВОНИЛ зачем?* ['Была ли цель, с которой он позвонил?']). Это, как представляется, странно, потому что с точки зрения наивной прагматики у любой ситуации есть причина, а у любого действия – цель.

4.2. Сочетаемостные свойства

4.2.1. Сочетаемость с частицей *именно*

Как показывает корпусное исследование, почти для всех вопросительных местоимений характерно или возможно употребление с фокусной частицей *именно*. Эта частица вводит пресуппозицию существования альтернативы и указывает на множество, из которого происходит выбор (Добровольский, Левонтина 2010, Козлов 2020). Ср. следующий пример, из которого ясно, что *именно* предполагает выделение одного элемента (в данном случае повести «Записки сумасшедшего») из релевантного множества (произведения Гоголя) по определенному признаку (близость к Достоевскому):

(10) *Именно «Записки сумасшедшего» можно назвать вещью Гоголя, наиболее близко стоящей как к форме, так и к методу Достоевского* (И. Золотусский).

Именно часто появляется постпозитивно в прямых и косвенных вопросах с *кто, что, какой* и прочими вопросительными местоимениями. Вообще, подобные конструкции составляют около 10% от всех употреблений *именно* в основном корпусе НКРЯ. Таким образом, можно говорить о том, что семантика *именно*, т.е. идея выбора одного элемента из множества по заданному признаку, «притягивается» к семантике вопросительных местоимений (Stefanowitsch & Gries 2003). Ср. примеры из НКРЯ:

(11) *Николай не может внятно ответить,* **[что]**[4] *именно ему нужно, он даже не знает о чем мечтает* (Интернет-форум).

Здесь речь идет о потенциально неограниченном множестве желаний и нужд человека – настолько большом, что в нем трудно ориентироваться самому субъекту ментального состояния.

(12) *В первое лето по приезде в Вологду, всей семьёй подались мы на пристань с целью поехать на рыбалку, но* **[куда]** *именно, ещё не знаем* (В. Астафьев).

Здесь подразумеваемое множество – это те места, куда можно поехать на рыбалку на вологодской пристани, и оно явно подразумевает более двух возможностей, поскольку при наличии лишь двух альтернатив употребление *именно* не оправданно; ср. странность (13), но не (14):

(13) ?*Кто именно из вас двоих разбил окно?*
(14) *Кто именно из вашего класса разбил окно?*

Это ограничение, по-видимому, имеет прагматическую природу и касается не только *именно*, но и других кванторных слов, например, *только, никто, все, каждый*:

(15) ?*Только я из нас двоих решил задачу*;
(16) ?*Никто из нас двоих не решил задачу.*

Подобные высказывания нарушают максиму информативности Грайса: если в множестве только два элемента, их идентификация не требует когнитивного усилия, на который тем или иным образом указывают кванторные слова.

Приведем частотные данные по сочетаемости вопросительных местоимений с частицей *именно*. Мы указываем общую встречаемость местоимений в корпусе, количество их сочетаний с *именно* и процентную долю употреблений *именно* среди всех вхождений местоимений.

Из табл. 1 видно, что по частотности появления с *именно* вопросительные местоимения образуют некоторую иерархию, которая не соотносится напрямую с их общей встречаемостью в НКРЯ: например, *как, что* и *когда* значительно частотнее, чем *какой*, но относительно реже встречаются с *именно*. При этом часть употреблений с *именно*, особенно у *что* и, в меньшей степени, у *когда* не относится к интересующим нас случаям, поскольку в них представлено союзное значение данных слов, и они не входят в сферу действия *именно*, а, напротив, включают частицу в свою сферу действия; ср. (17):

---

[4] Квадратными скобками мы отмечаем сферу действия частицы *именно*, т.е. в данном случае вопросительное местоимение. Интересно, что вообще *именно* употребляется препозитивно к слову, которое входит в его сферу действия (*Я люблю именно Васю*), однако постпозитивно с вопросительными местоимениями.

(17) *Как показывают приведённые примеры, возможны случаи, когда* [*именно сирконстанты*] *являются теми элементами, которые определяют нормативность синтаксической конструкции* (В. Храковский).

Таблица 1. **Встречаемость вопросительных местоимений с *именно* и в целом**

местоимение	вхождения в НКРЯ, общее число	вхождения в НКРЯ с *именно*	% вхождений с *именно*
какой	342127	2618	0.76%
чей	18184	69	0.37 %
кто	355623	1119	0.3%
сколько	89573	236	0.26%
куда	97225	216	0.22 %
где	323725	615	0.19%
что	3707974	11221	0.15 %
как	1910556	1165	0.06%
откуда	41195	26	0.06 %
когда	572008	432	0.005 %

Поскольку употребляемость вопросительных местоимений с *именно* напрямую не коррелирует с их частотностью, для нее должно быть семантическое объяснение. По-видимому, из высокочастотных местоимений идея выбора наилучшим образом сочетается с семантикой признака (*какой*), поскольку у каждого объекта есть какие-то, потенциально варьирующиеся, признаки. Время и место занимают невысокое положение в данной иерархии – с одной стороны, время и место присущи любой ситуации, с другой стороны – они, по-видимому, очень редко предполагают открытое множество выбора. Вопросы типа *когда именно* и *где именно* (а также их новые разговорные аналоги *когда по времени* и *где территориально*, несколько выходящие за пределы литературной нормы) не отсылают к потенциально бесконечному списку всех возможных времен и мест, а, как правило, просят уточнить точную локализацию или момент времени в некотором ограниченном ситуацией, хотя и не названном эксплицитно, пространстве или временном отрезке. Несколько более часто варьируются направление, количество, субъекты и объекты, ср. относительно более высокую частотность *куда именно* (существенно чаще, чем исходная точка движения), количество, ср. *сколько именно* (а также новое нелитературное *сколько по деньгам*), а также *кто именно* и *что именно*.

Что касается местоимений *зачем* и *почему*, для них употребление с *именно* крайне нехарактерно. Имеющиеся в НКРЯ примеры не относятся к интересующей нас конструкции: частица *именно* в них относится не к вопросительному местоимению (в чью сферу действия она входит), а к другим словам:

(18) *Почему именно* [*через этот банк*] *пойдут деньги пенсионеров?* («Аргументы и факты», 2003.06.04).

Примеров на *зачем именно* в Основном корпусе НКРЯ нашлось 49, при этом только в 21 примере представлена нужная конструкция, и большая часть этих примеров относится к XIX веку; ср.:

(19) *До сих пор для меня тайна, зачем именно она приглашала нас к себе?* (Ф.М. Достоевский).

При этом по общей встречаемости *зачем* (с более чем 75000 вхождениями) превышает *сколько, откуда* и *чей*, а *почему* (с более чем 132000 вхождений) превышает еще и *куда*.

Чем объясняется это отличие *почему* и *зачем* от других вопросительных местоимений? С одной стороны, у всякой ситуации есть либо цель (если она каузирована целенаправленно), либо причина (если она происходит нецеленаправленно), т.е. вопрос о причине или цели ситуации вполне оправдан. Нельзя это и объяснить и неспособностью этих местоимений к фокализации, поскольку они не просто свободно, но даже часто сочетаются с некоторыми другими фокусными частицами, например с *же* и *вот*. Ср. сочетания *почему же* (11473 вхождения), *зачем же* (8625 вхождений), *вот почему* (9080 вхождений) и *вот зачем* (405 вхождений).

Несочетаемость *именно* с *почему* и *зачем* объясняется, по-видимому, тем, что семантика цели и особенно причины плохо совместима с идеей выбора. Потенциальный круг возможных целей совершенного действия или возможных причин состоявшейся ситуации удивительно мало вариативен. При том, что в каком-то смысле причина и цель чаще бывают неизвестны, чем другие аспекты ситуации (они не наблюдаемы, в отличие от субъектов, объектов, места, направления и пр.), по-видимому, в языке заложено представление об их внутренней заданности для каждой конкретной ситуации и действия. Для причины это легко объяснимо: субъект не контролирует причину ситуации и, соответственно, не может выбирать ту или иную причину, по которой ситуация имеет место. Что касается цели, ее заданность и «прикрепленность» к ситуации несколько менее очевидны, однако, явным образом, совершая действие, субъект не выбирает цель из множества доступных и готовых к употреблению, а определяет ее для себя сам. В самом деле, можно сказать *Давай выберем место и время проведения встречи*, но странно сказать ?*Давай выберем цель проведения встречи*, если только речь не идет о том, что сообщить по этому поводу окружающим. Цель или причина могут быть неизвестны, и мы часто о них спрашиваем, но мы не предполагаем, что потенциальный ответ выбирается из множества разных, заранее доступных вариантов. Возможно это связано с тем, что цель и причина характеризуют ситуацию в целом, в то время как субъект, объект, место, время и т. п. представляют собой лишь какой-то ее аспект. Эта имманентная заданность цели и причины и приводит к невозможности или крайней ограниченности сочетания слов, указывающих на причину и цель, с частицей *именно*, в семантику которой заложена идея выбора из множества.

Эта же идея заданности объясняет невозможность или ограниченность употребления *зачем* и *почему* во многих других контекстах, свойственных другим вопросительным местоимениям, которые рассматриваются ниже, и выделяет их в особый класс.

4.2.2. Уточняющие конструкции с из или обстоятельством

Для многих РВМ характерно или возможно употребление в уточняющих конструкциях с *из* или обстоятельством времени или места, функция которых состоит в выборе одного элемента, отвечающего необходимым условиям, из некоторого заранее очерченного множества, или выбор точной локализации или момента времени из заранее названного пространства или временного отрезка. Ср. примеры уточняющих конструкций с разными местоимениями: *кто из них, какой из них, что из этого, когда завтра, откуда из Грузии* и т. п.:

(20) *Не знаю уж, **кто из них** прав, кто виноват* (В. Аксенов);
(21) *Существует множество путей для развития теории, и не ясно, **какой из них** предпочтительнее* («Знание – сила», 2003);
(22) *Барбара спросила: – **Когда завтра** самолет? – В Чанги надо быть к двенадцати* (В. Скворцов);
(23) *– Из Грузии, – без охоты признался мужик. – **Откуда из Грузии?** Поти?* (А. Терехов).

Разные вопросительные местоимения употребляются в разных уточняющих конструкциях. *Кто, что, какой, чей* употребляются в конструкции с *из*, которая указывает на выбор одного из элементов дискретного множества по какому-либо признаку: *Кто из них тебя обидел?; Что из этого тебе не нравится?; Какую из этих книг ты хочешь* и т.п. Эта конструкция характерна и для соответствующих неопределенных местоимений, а также для некоторых кванторных слов и для суперлативов; ср. *кто-то из ее друзей, Сыграй мне что-нибудь из Моцарта, каждый из, некоторые из, немногие из, лучший из.* Местоимение *сколько* также употребляется в конструкции с *из*, в значении 'какая доля от общего количества': *Сколько из этих денег ты потратил?* Эта конструкция характерна и для соответствующих неопределенных местоимений: *Девочке подарили десять конфет, сколько-то из них она сразу съела.*

Для местоимений *кто* и *какой* уточняющая конструкция с *из* характерна в высшей степени, как видно из табл. 2, и практически все примеры на *кто из* и *какой из* являются ее реализацией – доля «шума» ничтожна.

Таблица 2. **Встречаемость местоимений с предлогом *из***

Местоимение	*кто*	*какой*
общая частотность в НКРЯ	355623	342127
вхождения с *из*	10757	3542
% употреблений с *из*	3 %	1 %

Для *где* характерны употребления с обстоятельствами места: *Где вы живете во Франции?, Где они отдыхают на Гавайях?* Конструкции вида *где из* выглядят ненормативными: ??*Где из Франции вы живете?* Таким образом, для *где* множеством выбора является заранее заданное пространство, в котором при помощи уточняющей конструкции фиксируется одна точка: Франция воспринимается именно как недискретное множество, а не как набор дискретных топонимов. То же верно относительно времени: *когда* употребляется в контексте другого обстоятельства, задающего отрезок времени: *когда завтра, когда в следующем году, когда в древности, когда в XIX веке.* Таким образом, для *когда* недискретным множеством выбора является временной период, в котором фиксируется один момент – как бы временное пространство. Аналогичные конструкции характерны и для соответствующих неопределенных местоимений: *где-то в лесу, где-нибудь за городом, когда-то в прошлом, когда-нибудь на следующей неделе.* Такой способ уточнения вообще типичен для места и времени и вне контекста вопросов и местоимений; обстоятельства места и времени часто дублируются: *на площади под часами, в лесу на опушке, в спальне под кроватью; завтра в полдень, в понедельник вечером* и пр.

Куда и *откуда* также употребляются с дублирующими обстоятельствами – направления и источника соответственно: *куда во Францию, откуда с Кавказа.* Однако встречаемость подобных конструкций намного ниже – по-видимому, в силу того, что направление, особенно направление из заранее известного источника, – это изначально нечто намного более точное, чем место и время, и поэтому потребность в уточнении намного ниже. Этим же объясняется и более низкая встречаемость *куда* и *откуда,* нежели *где* и *когда,* с частицей *именно.*

Уточняющие конструкции с местоимением *как* не встречаются с предлогом *из,* однако лексикализованы с другими контекстами, в частности представлены фраземами *как иначе* (1523 вхождения) и *как по-другому* (71 вхождение) со значением 'каким другим образом'. Свободное выражение уточнения с *как* возможно и с наречиями, особенно в компаративе:

(24) *Надо хорошенько обсудить, как это сделать получше* (Н. Трубецкой).

(25) *У «партии власти» другая проблема – как эффективнее потратить гигантский рекламный бюджет* («Газета», 2003.06.20).

Все уточняющие конструкции с *как* также характерны для неопределенных местоимений *как-то* и *как-нибудь: Надо это сделать как-нибудь иначе, Надо стараться как-то понятнее выражаться.*

Уточняющие конструкции возможны, но не типичны для *что, сколько, чей, откуда* (в корпусе представлены единичные примеры). Местоимения *почему* и *зачем* невозможны в уточняющих конструкциях; ср. нестандартность ??*Почему по странной причине ты решил уехать?,* ??*Зачем с благородной целью ты это сделал?* Уточняющие конструкции невозможны для *почему* и *зачем* по той же причине, что и сочетания с *именно* – поскольку семантика *зачем*

и *почему* несовместима с идеей выбора из возможных альтернатив, вводимой уточняющими конструкциями.

Однако встречаемость РВМ в конструкции с *именно* не полностью предсказывает их встречаемость в уточняющих конструкциях. Так, местоимение *что*, которое очень частотно с *именно*, достаточно редко употребляется в конструкции *что из*. Как представляется, это связано с тем, что семантика *что* слишком широка и это мешает данному местоимению применяться по отношению к элементам заранее названного и очерченного множества однородных объектов. Вопрос с *что* часто предполагает в качестве ответа называние класса объекта, а не идентификацию одного конкретного объекта из заранее названного множества; ср. – *Что ты ей подарил? – Книгу* (называется класс объектов).

Даже если в качестве ответа указывается не класс, а конкретный объект, вопрос с *что*, как правило, не включает информации о классе: – *Что ты ей подарил? – «Войну и мир»*. Поэтому конструкция с *из*, в которой практически всегда называется класс объектов, для *что* не слишком характерна. Странно сказать ?*Что из книг ты хочешь,* ?*Что из салатов ты хочешь* и пр. Исключением являются конструкции вида *Что из этого ты хочешь?*, в которых не называется класс объектов, а указывается конкретное множество, или же конструкции с обозначениями таксономических классов (*Что из мебели/одежды тебе нужно?*). В ситуации, когда заранее известно, о какой совокупности объектов идет речь, и нужно указать на один конкретный объект, обладающий заданным признаком, употребляется местоимение *какой*: *Какую из книг ты хочешь?, Какой из салатов ты хочешь?*[5] Это объясняет высокую встречаемость *какой* в уточняющих конструкциях с *из*. Интересно, что *кто* высокочастотно в конструкции с *из*: это, опять-таки, связано с его (более узкой по сравнению с *что*) семантикой. *Кто* в качестве ответа предполагает не класс объектов, а уникальную конкретную идентификацию; ср. *Кого ты любишь? Васю*, при странности – *Кого ты любишь? – Человека <собаку, инженера, одноклассника>*. Соответственно, *кто* «притягивается» к конструкции с *из*, функция которой – уточнение конкретного объекта из заранее названного множества.

4.3. Конструкции с редупликацией

Для вопросительных местоимений характерны конструкции с редупликацией[6]. Есть как минимум три конструкции с редупликацией, которые имеют разный смысл и в разной степени сочетаются с разными местоимени-

---

[5] Русское вопросительное местоимение *какой* имеет две интерпретации в своем первом значении: 'what kind of' (вопрос о признаке объекта) и 'which' (просьба указать на конкретный объект из заранее заданной совокупности): – *Какое платье ты купила? – Шелковое, летнее* ['what kind of']; – *Вот это* ['which'].

[6] Обзор литературы по редупликации см. в (В. Апресян 2018).

ями. Ниже представлена статистика употреблений вопросительных местоимений с редупликацией по данным НКРЯ (без различения семантически различных типов редупликации).

Таблица 3. **Частотность местоимений с редупликацией**

Местоимение	Общая частотность в НКРЯ	Вхождения с повтором	% вхождений с повтором
что	3707974	**2937**	0,07 %
кто	355623	**715**	0,2 %
как	1910556	**549**	0,02 %
куда	97225	388	0,3 %
зачем	66462	**330**	0,5 %
почему	119 077	327	0,27 %
где	323725	298	0,09 %
какой	342127	185	0,05 %
сколько	89573	94	0,1 %
когда	572008	83	0,01 %
откуда	41195	42	0,1 %
чей	18184	11	0,06 %

4.3.1. Редупликация эмфазы

Наиболее универсальна конструкция, которую можно назвать «редупликацией эмфазы». Она возможна для всех местоимений, в том числе для *зачем* и *почему*. Это связано с тем, что она очень мало специфична семантически: эмфатическая редупликация возможна практически для всех частей речи и независимо от семантики слова; ср.

(26) *Всё это, конечно, так, но **тоска, тоска**…* (В. Пьецух);

(27) *– Я люблю вас, Владимир Михайлович, – сказала она и обняла его.
       – **Люблю, люблю**!* (Ю.О. Домбровский).

Конкретное чувство, которое выражает редупликация, зависит от контекста и может сильно различаться в зависимости от семантики повторяемого слова. У вопросительных местоимений повтор обычно передает раздражение, вызванное вопросом, или эмфазу в риторическом вопросе (сожаление, ужас, отчаяние и другие сильные эмоции):

(28) *– Эй, сестрички, как ваши дела? – **Как, как**? – огрызнулась Зина. –
       Никак, вот как* («Мурзилка», 2002);

(29) *– **Куда, куда**? Да на кудыкину гору!* (А. Моторов);

(30) *– Должно быть, не в мою смену. – А **кто, кто** был вместо вас?! –
       в отчаянии крикнула она* (Д. Рубина);

(31) *Ну **зачем, зачем** люди произносят такое количество слов?*
       (А. Волос).

У некоторых редуплицированных местоимений в интерпретации раздражения существуют клишированные грубые продолжения (которые могут также служить ответами на вопросы с одиночными местоимениями): *кто-кто, дед Пихто <конь в пальто>*: *куда-куда, на кудыкину гору*; *почему-почему, по кочану*; *где-где, у тебя на бороде*; *откуда-откуда, от верблюда* и пр. Такого рода фразеологизации объясняют относительную частотность редупликаций у более редких местоимений типа *откуда.*

4.3.2. Редупликация переспроса

Вторая распространенная конструкция с редупликацией вопросительных местоимений используется для переспроса, если человек не расслышал или не понял собеседника, или если то, что он услышал, слишком неожиданно. Способность к употреблению с редупликацией переспроса имеется у всех местоимений, кроме *зачем* и *почему*, для которых эта конструкция нехарактерна. Ср.:

(32) – *Мы едем в Занзибар. –* ***Куда-куда?***;

(33) – *Вам звонила Кассиопея Никандровна. –* ***Кто-кто?***;

(34) – *Нужно заказать новый нефелометр. –* ***Что-что?***;

(35) – *Сорок гривен!* – ***Сколько-сколько?!*** *– Сорок гривен!* (Р. Карцев);

(36) – *Ты откуда? Паренек пролепетал что-то невнятное. Он дрожал всем телом и часто моргал. –* ***Откуда, откуда?*** *– С Едрина. – Едрино? Где это такое?* (Ю. Вяземский).

По-видимому, редупликация такого рода предъявляет как фонетические, так и референциальные требования. С одной стороны, поскольку при редупликации переспроса повторенное местоимение произносится как единое фонетическое целое, она более характерна для коротких, односложных слов (*кто, что, как*). С другой стороны, поскольку иллокутивная цель переспроса – каузировать повтор высказывания, повторяемое высказывание не должно быть слишком длинным. Во всех примерах просьба о повторе касается одного слова или короткой фразы. Местоимения *зачем* и *почему*, с одной стороны, не односложны, с другой стороны – сообщения о цели или причине не могут передаваться одним словом; как правило, они требуют целого придаточного предложения. Таким образом, редупликация переспроса нехарактерна для *зачем* и *почему*: сами эти слова слишком длинны для редупликации с превращением в единое фонетическое слово, и информация, которую требовалось бы повторить, также слишком длинна.

4.3.3. Уступительно-противительная редупликация

Третий тип конструкции с редупликацией, характерный для русских вопросительных местоимений – это уступительно-противительная

редупликация (Feldstein 2016). Она возникает не в вопросительном значении, а в разговорном значении неопределенного местоимения, обсуждавшемся выше; ср.

(37) *Кто-кто, а уж он не подведет* 'Какие-то другие люди могут подвести, но он нет';

(38) *Где-где, а на Канарах всегда хорошая погода* 'В каких-то других местах может быть плохая погода, но не на Канарах'.

При данной редупликации двойное местоимение является единым фонетическим словом, обосабливается и употребляется перед союзом *а*.

Это значение отсутствует у *зачем, почему* и практически не встречается у *сколько*. Соответственно, для них неестественна или невозможна эта редупликация:

(39) ^{??}*Сколько-сколько, но уж пятьсот рублей он тебе точно даст*;

(40) **Зачем-зачем, но уж чтобы ей помочь, он точно пальцем не двинет*;

(41) **Почему-почему, но уж из-за ветра крыша точно не обвалится.*

У остальных вопросительных местоимений количество обособленных конструкций с повтором перед союзом *а* распределяется следующим образом.

Таблица 4. **Редуплицированные вхождения перед союзом *а***

Местоимение	Общая частотность в нкря	Редуплицированные вхождения перед союзом *а*
что	3707974	366
кто	355623	296
где	323725	18
куда	97225	3
когда	572008	2
какой	342127	1
как	1910556	0
чей	18184	0
откуда	41195	0

Как видно из количественных данных, уступительно-противительная редупликация относительно частотна только у местоимений-существительных, однако единичные примеры встречаются и у других местоимений, у которых есть значение неопределенности. Ср.:

(42) ***Чего-чего****, а работы у нас хватает* (В. Аксенов);

(43) ***Кого-кого****, а их-то должны обслуживать по высшему разряду!* (И. Грекова).

Как и в других случаях, встречаемость разных вопросительных местоимений в данной конструкции объясняется их семантикой. Смысл конструкции с редупликацией таков:

(44) *(Уж) X-X, а Y Z* 'Есть самые разные элементы X некоторого множества; многие или все из них не обладают свойством Z; элемент Y этого множества точно обладает свойством Z'.

Поскольку *почему* и *зачем* в принципе несовместимы с идеей альтернативы, для них невозможно употребление в конструкции, указывающей на существование других элементов, принадлежащих тому же множеству – в данном случае множеству релевантных причин или целей. Что касается остальных местоимений, то их распределение регулируется прагматическими факторами. Множество существ или множество объектов, (не) обладающих определенными свойствами – это легитимные и часто встречающиеся множества. Множество мест с определенными свойствами – также достаточно часто встречающийся тип множества. Однако множество признаков с определенными свойствами (*уж какой-какой*), множество способов с определенными свойствами (*уж как-как*) и особенно множество количеств с определенными свойствами (*уж сколько-сколько*) представить себе намного труднее. Поэтому большая часть местоимений, при потенциальной возможности употребления в уступительной конструкции, в реальности практически в ней не встречается. При этом некоторые местоимения формируют фразеологизованные коллокации с этой конструкцией; ср. например *Кто бы говорил*, *Чья бы корова мычала (а твоя бы молчала)* со значением ['Если другим и пристало об этом говорить, то тебе нет'].

4.4. Конструкции с риторическими вопросами

Для вопросительных местоимений характерно употребление в разных типах риторических конструкций. Мы не рассматриваем здесь экскламативы со словами *какой* и *как* (*Как она прекрасна!, Какой из него музыкант* и пр.), описанные в работах (Апресян 1995, Зевахина 2018) и др., а только конструкции, значимые для всех или большей части вопросительных местоимений.

1. Риторические вопросы и восклицания со значением отрицания (*Кто может сравниться с Матильдой моей*);

2. Риторические вопросы и восклицания с отрицанием, со значением всеобщности или разнообразия (*А кто не пьет*?);

3. Риторические вопросы с частицей *-то* со значением 'неизвестно, какой человек/объект/место/направление и пр. будет участником ситуации А1, и будут ли они хорошими в этом качестве' (*Где-то мы будем жить?*);

4. Риторические вопросы или восклицания вида *А кто к нам пришел!*, в ситуациях языковой игры, когда говорящий имеет в виду известного ему конкретного человека (объект, направление и пр.), о котором он собирается сообщить адресату, считая, что его это порадует, или же самого адресата.

5. Риторические вопросы вида *(Да) кто А1 такой <такая>, чтобы А2?*, со значением 'А1 не такой человек, который имеет право делать А2'.

Конструкция вида *(Да) что (ж) это такое* выражает менее специфическую семантику возмущения по поводу имеющей место ситуации:

(45) *– Да что же это такое! – возмутилась Ольга* (В. Белоусова).

Для остальных местоимений, не являющихся существительными, сочетание вообще невозможно.

Исчисление встречаемости каждого из местоимений в каждой из конструкций затруднено в силу того, что они либо не имеют фиксированного синтаксического оформления, либо имеют большое количество синтаксических омонимов.

Первый тип вопросов и восклицаний (с отрицательной интерпретацией) возможен для всех местоимений, включая *зачем* и *почему*, поскольку семантика отсутствия не противоречит их семантике:

(46) *Почему бы машина сломалась?* 'Нет причины, по которой машина могла бы сломаться';
(47) *Зачем бы я стал тебе мешать?* 'Нет цели, которую бы могло преследовать такое поведение'.

Эта конструкция обычно предполагает использование сослагательного наклонения, модального слова, будущего времени или иного показателя неверидикативности: *Чем тут можно помочь <поможешь>?* ['помочь нельзя ничем']; *Как я ему объясню, что его увольняют?* ['объяснить никак невозможно']; *Где нам искать спасения?* ['негде']; *Кто бы устоял перед этим искушением?* ['никто']. Общий смысл этой конструкции такой:

(48) *X бы Y* 'Нет X-а (человека, объекта, ситуации, места, источника, количества, причины, и пр.), который бы имел свойство Y'.

Эта конструкция в высшей степени характерна для *кто* и образует ряд фразеологизованных коллокаций с этим местоимением: *Кто бы мог подумать* 'Никто не думал'*, Кто бы мог вообразить* 'Никто не мог вообразить'*, Кто бы знал* 'Никто не знал'*, Кто бы сомневался* 'Никто не сомневается'*, Кто бы спорил* 'Никто не спорит'.

У *почему* при этом есть своя собственная конструкция с риторическим вопросом, *почему бы не*, которая выражает предложение нечто сделать:

(49) *А почему бы вам не организовать продажу ваших вещей у нас, за Уралом?* («Народное творчество», 2004).

Риторическая конструкция со значением всеобщности в разной степени возможна для большей части вопросительных местоимений. Эта конструкция включает в себя эксплицитное или имплицитное отрицание:

(50) *Чего не сделаешь ради своего ребенка?* 'Сделаешь всё';
(51) *Кому понравится такое хамство?* 'Всякому человеку не понравится такое хамство'.

Она также часто включает частицу *только* в усилительном значении, особенно если глагол стоит в прошедшем времени:

(52) *Кто там только не был!* 'Все, кого можно себе представить, были';

(53) *Чего только она не наготовила!* 'Все, что можно себе представить в данной ситуации, она наготовила'.

Смысл конструкции без частицы *только* можно сформулировать следующим образом:

(54) *X не Y* (*А кто не пьет?*; *Кому охота в воскресенье в шесть утра идти на работу?*) 'Всякий X (человек, объект, ситуация, место, источник, причина и пр.) имеет свойство Y'.

В сочетании с частицей *только* смысл конструкции несколько сдвигается:

(55) *X только не Y* (*Чего только он ради нее не делал* <*Чего только не сделаешь ради друга*; *Куда только я не обращалась!*> 'Все X-ы (люди, объекты, ситуации, места, источники, причины, и пр.), среди X-ов, релевантных для данной ситуации, имеют свойство Y'.

Таким образом, без *только* (преимущественно в будущем времени и настоящем хабитуальном) конструкция интерпретируется генерически – как относящаяся ко всем элементам открытого множества потенциально возможных X-ов. С *только* (преимущественно в прошедшем времени) она воспринимается конкретно-референтно – как описывающая все элементы замкнутого множества тех X-ов, которые можно себе представить в имевшей место ситуации. Это употребление носит более риторический и менее буквальный характер: на самом деле имеется в виду очень большое многообразие X-ов, а не реальная всеобщность.

Эта конструкция невозможна для *зачем* и *почему* по уже многократно упоминавшейся причине: *зачем* и *почему* не совместимы с семантикой множественности, которая вводится, в данном случае, как следствие семантики всеобщности или многообразия. Ср. невозможность **Зачем только не поможешь другу?* [невозможна интерпретация 'У помощи другу есть всякая цель'], при возможности фраз вида *Чего только не сделаешь, чтобы помочь другу!*, **Почему только не вернешься домой* [невозможна интерпретация 'У возвращения домой есть всякая причина'], при естественности *Чего только не отдашь, чтобы вернуться домой!*

Кроме того, эта конструкция прагматически неестественна для *сколько* (в НКРЯ нет подобных примеров), особенно в сочетании с частицей *только*. Без частицы *только* возможно сконструировать относительно естественные примеры со *сколько* в данной конструкции:

(56)   *Сколько не заплатишь, чтобы вернуть здоровье!* ['Заплатишь любую сумму'], поскольку речь идет о неограниченном множестве потенциальных сумм, всякую из которых субъект готов заплатить.

Однако в сочетании с *только* конструкция становится странной:

(57)   ??*Сколько он только не заплатил за свои операции на позвоночнике!*

Это связано с тем, что *только* с прошедшим временем указывает на некоторое множество релевантных денежных сумм, все из которых субъект заплатил, что прагматически крайне маловероятно. То же верно для вопросительного местоимения *когда*. Примеры без *только* в принципе возможны, хотя и не встречаются в НКРЯ:

(58)   *Когда он не ленился и не отлынивал?* 'Он всегда ленился и отлынивал'.

Примеры с *только* прагматически неадекватны, поскольку предполагают разнообразие временных характеристик некоторой конкретной ситуации:

(59)   ?*Когда он только не говорил ей комплименты!*

Некоторые местоимения образуют устойчивые фразеологизованные сочетания с данной конструкцией: *С кем не бывает* ['Не волнуйся, всякий человек может сделать ошибку, подобную той, что сделал ты'], *Чего (на свете) не бывает* ['Случаются самые разные, часто удивительные, вещи']. Точное исчисление встречаемости данной конструкции невозможно в силу ее высокой омонимичности с не-риторическим вопросом: ср. *Кто не придет к Саше на день рождения?* ['Всякий человек придет к Саше на день рождения' или 'Скажи мне, какие люди не придут к Саше на день рождения']. Не-омонимичными являются вхождения с частицей *только*: *Кого я только у него не встречал* ['встречал самых разных людей'], однако они представляют собой лишь часть реально встречающихся риторических вопросов со значением всеобщности. Ниже приводятся частотные данные встречаемости вопросительных местоимений в конструкции с *только не* (местоимение + *только* + *не*) по НКРЯ.

Таблица 5. **Частотность местоимений с *только не***

Местоимение	что	кто	где	как	куда	какой	откуда	чей
Общая частотность в НКРЯ	3707974	355623	323725	1910556	97225	342127	41195	18184
вхождения с *только не*	1170	357	106	100	47	41	12	1

Таблица частотностей риторической конструкции с *только* в интерпретации разнообразия и многочисленности напрямую не коррелирует с частотностью самих вопросительных местоимений (за исключением самой верхней и самой нижней частей таблицы), как это верно и относительно всех других

рассматриваемых конструкций. Чаще всего речь идет о разнообразии объектов и субъектов ситуаций, реже всего – о разнообразии источников и посессоров, а разнообразие мест (*где*), способов (*как*), направлений (*куда*) и признаков (*какой*) занимает промежуточную позицию.

Третья риторическая конструкция, с частицей *-то* и вопросом, возможна для большинства вопросительных местоимений. Реальные корпусные примеры употребления найти непросто, поскольку они встречаются достаточно редко (конструкция стилистически окрашена как уходящая и высокая или поэтическая) и при этом они омонимичны с высокочастотными неопределенными местоимениями на *-то*. Однако возможно сконструировать достаточно естественные примеры употребления. Конструкция имеет следующий смысл:

(60) *X-то Y* (*Куда-то мы с тобой попадем?*; *Как-то мне ему это объяснить?*) 'Неизвестно, какой X (человек/объект/место/ признак/способ/количество/направление) является или будет являться участником некоторой ситуации и будет ли он хорош в этом качестве'.

Для местоимений *зачем* и *почему* эта конструкция нехарактерна в силу ее проспективной ориентации и/или неопределенности. *Зачем* и *почему* ретроспективно ориентированы: мы обычно спрашиваем о причине или цели уже имеющей место ситуации; если же вопрос задается относительно будущей ситуации, говорящий должен быть уверен в том, что она произойдет, что противоречит идее неопределенности, вносимой конструкцией с *-то*; ср. странность ?*Почему-то она ему откажет?*; ?*Зачем-то он ее бросит?* Эти фразы возможны без *-то* (*Почему она ему откажет?, Зачем он ее бросит?*), но в таком случае они предполагаются в качестве реакции на утверждение собеседника, что соответствующие ситуации будут иметь место, и вопрос относится к основаниям высказанного мнения. Для *откуда* и *чей* естественные примеры также трудно сформулировать, поскольку семантика источника и семантика посессора в меньшей степени совместимы с проспективностью и неопределенностью, чем семантика объекта, субъекта, места и пр.; ср. неполную естественность следующих фраз: *Откуда-то мы получим его следующее письмо?*; *Чьи-то руки будут ее обнимать завтра?*

Для большей части вопросительных местоимений также характерны риторические вопросы или восклицания. Они используются в ситуациях языковой игры, когда говорящий имеет в виду известного ему конкретного человека (объект, направление и пр.), о котором он собирается сообщить адресату, считая, что его это обрадует. Такие фразы произносятся с особой просодией – подъемом на последнем слове и растяжением его слогов. Такого рода употребления характерны для разговорной речи, причем для прагматически весьма ограниченного круга контекстов: а именно, ситуаций сюрприза адресату, к которому говорящий обращается достаточно фамильярно.

Для местоимений *какой* и *как* поиск примеров дополнительно осложняется наличием омонимии с широко распространенной восклицательной конструкцией *какой X!*; *Как X!*: *Какой он (потрясающий) музыкант!*; *Какую*

(прекрасную) он подарил мне картину!; *Как (хорошо) мы отдохнули!* Эта конструкция, с одной стороны, описана в многочисленных работах, с другой стороны, характерна только для *какой* и *как*, поэтому здесь мы ее не рассматриваем.

Конструкция имеет следующий смысл:

(61)   *(А) X Y!* (*А что я тебе подарю!*; *А кто к нам придет!*; *А куда мы с тобой поедем!*; *А где сейчас твоя мама!*; *А какую я тебе игрушку подарю!*; *А сколько я выиграл в рулетку!*) 'Говорящий знает, какой X (человек/объект/направление/место/ признак/количество) является или будет являться участником некоторой ситуации, и считает X очень хорошим; адресат этого не знает; говорящий сообщает адресату об этом, считая, что адресату это будет приятно'.

Сочетаемость этой конструкции с разными вопросительными местоимениями ограничена прагматикой сюрприза. Самые естественные ситуации приятного сюрприза – какой-то подарок или чей-то приход, поэтому местоимения *что* и *кто* употребляются в ней наиболее свободно. Для большинства других местоимений конструкция допустима, хотя и с некоторой натяжкой. С другой стороны, некоторые аспекты ситуации, такие как принадлежность (*А чьего мужа я вчера видела в парке!*) или способ (*А как мы разведем костер под дождем!*), с трудом поддаются нужной интерпретации. Для причины и цели такое вообще не представляется возможным; ср. прагматическую невероятность следующих сюрпризов: **А почему он мне позвонил!*; **А зачем он вчера ко мне зашел!*

4.5. Конструкции с *угодно, хочешь, попало, ни попадя, придется, абы* **и пр.**

Мы не будем подробно углубляться в описание этих конструкций, так как они достаточно детально описаны в работах (Левонтина, Шмелев 2005, 2018). Для всех вопросительных местоимений, кроме *зачем* и *почему*, возможно употребление в конструкциях с *угодно, хочешь (хотите), попало* и пр., с интерпретацией свободного выбора (Анна Зализняк, Падучева 2020). Для *зачем* и *почему* эти конструкции практически невозможны (ср. неприемлемость сочетаний **почему угодно, *абы почему, *почему попало* и странность сочетаний *?зачем угодно, ??абы зачем, ?зачем попало*), т.к. выбор, как было многократно показано, исключается семантикой цели и причины.

В разных сочетаниях добавляются дополнительные прагматические оттенки: широты выбора (*угодно*), широких возможностей выбора желаемого (*хочешь*), небрежно сделанного выбора (*попало, ни попадя*), отсутствия достаточного выбора (*придется*), небрежного и плохого выбора (*абы*). Кроме того, хотя все эти конструкции со значением свободного выбора не веридикативны, т.е. не употребляются в фактивных контекстах (что неудивительно для семантики свободного выбора), они по-разному распределяются по

неверидикативным контекстам. Для каких-то конструкций характерна проспективность, для каких-то потенциальность, для каких-то узуальность.

Для конструкции с *угодно* нехарактерно ретроспективное употребление – она употребляется в контекстах будущего и узуальности: ср. невозможность **Он пришел с кем угодно*, при возможности *Приходи с кем угодно, Можно приходить с кем угодно*. Подстановка в этот пример всех других конструкций была бы прагматически неуместной по разным причинам. Фраза вида *Человек начнет винить в своих неудачах кого хочешь, но не себя* неестественна, так как полностью исключает интерпретацию желательности. Ср. следующий пример, в котором речь идет об участии в положительно оцениваемой ситуации, и употребление конструкции свободного выбора с *хочешь* оправдано:

(62) *Она девка красивая, умная, кто хочешь в нее влюбится, в принцессу мою* (А. Маринина).

Безусловно, положительная оценка предиката не является обязательным условием употребления конструкции с *хочешь*; ср. (63). Однако некоторая идея волитивности и заинтересованности в этом примере присутствует:

(63) *А дедушка был не такой человек, чтобы позволить себя бить и уродовать. Он сам мог изуродовать кого хотите* (А. Рыбаков).

При этом возможны употребления, где никакой заинтересованности усмотреть невозможно; ср. следующий пример:

(64) *От такой картины у кого хочешь голова закружится* (А. Геласимов).

По-видимому, у конструкции с *хочешь* есть большое количество ослабленных употреблений, однако ситуации, которые безусловно отрицательно оцениваются говорящим, исключают ее употребление: так, странно сказать

(65) *?Он на нее кому хочешь готов жаловаться.*

Кроме того, для конструкции с *хочешь*, как и для конструкции с *угодно*, характерно проспективное употребление и неестественно ретроспективное:

(66) **От этой картины у кого хочешь голова закружилась.*

Конструкции с *попало* и *ни попадя* указывают, помимо идеи свободного выбора, на то, что эти объекты выбираются без должного тщания и в силу этого имеют низкое качество:

(67) *– Но она же спит направо и налево с кем попало! – закричала Эльвира, выведенная из себя* (А. Слаповский).

В силу этих семантических особенностей – указания на уже многократно сделанный и при этом неправильный, с точки зрения говорящего, выбор – для конструкции с *попало* и *ни попадя* характерно употребление в узуальных кон-

текстах, ретроспективных или актуальных: *Она спала с кем ни попадя*; *Он общается с кем попало*. В силу своей семантики она не употребляется в императивных и проспективных контекстах: нельзя сказать **Общайся с кем попало* (при нормальности *Общайся с кем хочешь, Зови кого угодно*), **На такой девушке кто попало женится* (при нормальности *На такой девушке кто хочешь женится*). Для нее также невозможно употребление в ситуации, когда не может идти речь о реальном выборе субъекта ситуации: ср. естественность *Она встречается с кем попало / с кем ни попадя* (субъект плохо выбирает, с кем встречаться), при странности ?*В нее влюбляется кто попало / кто ни попадя* (от субъекта ситуации ничего не зависит).

Конструкция с *придется* подчеркивает отсутствие возможности выбора и в силу этого невысокое качество объектов: субъект ситуации вынужден удовлетворяться тем, что есть. Эта конструкция употребляется в узуальных контекстах, где речь идет о выборе доступных неудовлетворительных объектов: *Ели что придется* ('то, что было доступно и было не очень хорошим'), *Ночевали где придется* ('там, где была возможность, в не очень хороших местах'). Конструкция с *придется*, как и конструкция с *попало*, не употребляется в императивных и проспективных контекстах: **Приводи кого придется* (при нормальности *Приводи кого угодно*), **Такую статью какой придется журнал примет!* (при нормальности *Такую статью какой угодно журнал примет*).

Конструкция с *абы* имеет, по сравнению с остальными конструкциями свободного выбора, следующие особенности. Она, как и конструкции с *попало* и *ни попадя*, указывает на некачественные объекты, однако, в отличие от них, проспективно и прохибитивно ориентирована. Если *попало* и *ни попадя* указывают на плохо сделанный выбор, конструкция с *абы* предполагает желание избежать такого выбора. Соответственно, она отрицательно поляризована и используется преимущественно в отрицательных контекстах (см. Левонтина, Шмелев 2018), при этом часто в контексте глаголов желания:

(68) *Строил для себя: черепица, стропила обработаны, лестница дуб, я и сам хотел, чтобы не абы кому* (А. Терехов).

Эта конструкция также применима к узуальным отрицательным контекстам и контекстам с внутрисловным отрицанием:

(69) *Ее снимала Энни Лейбовиц, а Энни Лейбовиц абы кого не снимает* («Русский репортер», 2012).

Для *зачем* и *почему* конструкции со значением свободного выбора невозможны, потому что эти местоимения несовместимы с идеей альтернативы. Не исключено, что это типологическая универсалия – ср. отсутствие местоимений свободного выбора цели и причины в английском (хотя любое утверждение на эту тему нуждается в проверке на широкой выборке из генетически разных языков): ср. невозможность *anywhy, *whyever, при наличии *whoever, whichever, whatever, wherever, anybody, anything, anywhere.* В НКРЯ отсутствуют примеры на подобные конструкции.

Таким образом, в данном семействе конструкций сочетания с *угодно* и *хочешь* подчеркивают широту не отрицательно оцениваемого выбора, *попало* и *ни попадя* – плохое качество объектов из-за отсутствия тщательности при выборе, *абы* – желание избежать плохого выбора, а *придется* – его случайность и возможное плохое качество объектов из-за отсутствия достаточных возможностей для выбора.

В табл. 6 представлена встречаемость каждого из местоимений в каждой из конструкций. Некоторые конструкции омонимичны конструкциям с другим значением и при этом обладают высокой частотностью. Например, у многозначной коллокации *как хочешь* есть уступительная интерпретация (*Ты как хочешь, а я поеду*), интерпретация «несвободного выбора» (*Как хочешь, так и делай, но чтобы к завтрашнему дню все было готово*), композициональная интерпретация (*Как ты хочешь ехать, на машине или на самолете?*) и интересующая нас интерпретация свободного выбора (*В этой стране можно жить как хочешь, делать что хочешь, ездить куда хочешь*).

Таблица 6. **Сравнительная частотность конструкций свободного выбора**

Местоимение	Общая частотность в НКРЯ	Вхождения с *угодно*	Вхождения с *хочешь / хотите*	Вхождения с *придется*	Вхождения с *попало*	Вхождения с *ни попадя*	Вхождения с *абы*
что	3707974	4652	≈ 2360 (≈60 % от 3934	≈ 207 (≈ 20 % от 1028)	978	98	24
как	1910556	1080	≈ 866 (≈25 % от 3467	204	673	2	**84**
когда	572008	238	≈ 79 (≈30 % от 262	35	15	3	-
кто	355623	1936	≈ 542 (≈70 % от 775	33	427	69	37
Какой	342127	1649	≈ 93 (≈20 % от 465	11	46	1	20
где	323725	561	≈ 64 (≈20 % от 320	220	398	31	5
сколько	89573	**2185**	904	3	4	-	-
куда	97225	705	≈ 45 (≈6 % от 748	55	331	27	4
откуда	41195	31	17	5	12	1	-
чей	18184	47	5	-	1	-	-

В целом сочетаемость местоимения с конструкциями свободного выбора отражает отчасти их частотность в корпусе, отчасти типичную структуру событий: в первую очередь важны актанты, вводимые местоимениями-существительными *что* и *кто*, т.е. агентивные и пациентивные роли, их признаки (*какой*), а также основные сирконстанты – место (*где*), направление к цели, в силу goal bias (*куда*), способ (*как*), и, в меньшей степени, время (*когда*). Количество (*сколько*), принадлежность (*чей*) и направление от источника (*откуда*) либо менее важны для большинства ситуаций, либо менее способны к вариативности, вводимой семантикой свободного выбора.

Кроме того, среди конструкций есть частотные (с *угодно,* с *хочешь / хотите*), а есть очень редкие (с *абы,* с *ни попадя*), что связано с регистровыми и семантическими особенностями: крайней разговорностью (*попало, ни попадя*), устарелостью (*абы*), сложной семантикой (*абы, придется, попало, ни попадя*). При этом некоторые реализации конструкций явно лексикализованы и идиоматизированы, и такие сочетания обладают высокой частотностью или большей частотностью, чем можно было бы ожидать, учитывая частотность самого местоимения. Например, в серии *абы* лексикализовано сочетание *абы как*. При том, что само местоимение *как* в два раза менее частотно, чем *что*, сочетание *абы как* в 3,5 раза более частотно, чем *абы что. Абы как* – это, по сути, лексическая функция AntiBon со значением 'плохо, кое-как'.

В серии *угодно* лексикализовано сочетание *сколько угодно*. При том, что в НКРЯ *сколько* в 21 раз менее частотно, чем *как*, сочетание *сколько угодно* в 2 раза более частотно, чем *как угодно*. По сути, *сколько угодно* – это лексическая функция Magn со значением 'очень много'.

4.6. Конструкции с *мало, редко, много*

Семантика и свойства данных конструкций анализируются в работах (Апресян, Иомдин 1990, Иомдин 2006, В. Апресян 2007), поэтому здесь мы приводим лишь количественные данные по каждому из местоимений; ср. примеры употребления *Мало кто умеет так работать с детьми; Много кто туда заходит; Редко кто добирается до нашей деревни.* Легко заметить, что частотны в данных конструкциях только местоимения, которые наилучшим образом сочетаются с семантикой дискретности, привносимой кванторами *мало, много* и *редко* – т.е. с *что* и *кто*. Очевидным образом, сочетания с *зачем* и *почему* невозможны, в силу их несовместимости с идеей альтернативы и, соответственно, множественности. Запрос в НКРЯ был сформулирован таким образом, чтобы по возможности избежать «шума».

Таблица 7. **Частотность конструкций с *много, мало* и *редко***

Местоимение	Общая частотность в НКРЯ	Вхождения с *много*	Вхождения с *мало*	Вхождения с *редко*
Что	3707974	**1208**	**2410**	28
Как	1910556	-	-	-
Когда	572008	-	9	**269**
Кто	355623	60	**3324**	719
Какой	342127	-	41	**72**
Где	323725	43	49	**110**
Сколько	89573	-	-	-
Куда	97225	7	11	4
Откуда	41195	1	-	-
Чей	18184	1	3	1

4.7. Конструкции с *как (бы) ни*

Уступительные конструкции со значением предела вида *Кто (бы) ни обращался к нему, он всем отказывал*; *Когда (бы) и ни приезжал, ее никогда не было дома* подробно рассматриваются в (В. Апресян 2015), поэтому здесь мы приводим лишь количественные данные. Из табл. 8 видно, что наиболее частотна конструкция вида *как (бы) ни*. Она лексикализована в значении 'несмотря на предельно высокую степень X, имеет место Y' для сочетания *как ни* и в значении 'несмотря на возможную предельно высокую степень X, будет иметь место Y'; ср. *Как он ни старался, ничего не вышло*, *Как это ни печально, нам нужно уезжать*; *Как бы ты меня ни убеждал остаться, я уеду*. Поскольку главное значение конструкции – это значение достигнутого или возможного количественного предела, можно предположить, что значение степени, варьирующееся вплоть до предела, выражаемое в реализации *как*, наиболее востребовано языком. Сочетание *как ни* лексикализовано и даже фразеологизовано в различных устойчивых сочетаниях со значением уступительности: *как ни крути, как ни верти, как ни поверни, как ни посмотри, как ни старайся, как ни странно, как ни грустно* и т.д.

Таблица 8. **Частотность конструкций с *(бы) ни***

Местоимение	Общая частотность в НКРЯ	Вхождения с *бы ни*	Вхождения с *ни*
что	3707974	2103	12289
как	1910556	2576	14742
когда	572008	56	424
кто	355623	583	1175
какой	342127	600	974
где	323725	343	550
сколько	89573	459	1633
куда	97225	306	1807
откуда	41195	38	549
чей	18184	13	1

Также достаточно частотно варьирование объекта, вплоть до того, что он становится любым (*что (бы) ни*), и количества (*сколько*). Для местоимений *почему* и *зачем* реализация с *ни* невозможна, т.к. она предполагают **осуществленную** множественность причин и целей:

(70) **Почему он ни приходил, я его не принял*;
(71) **Зачем он ни приходил, я его не принял.*

Сочетания *почему бы ни* и *зачем бы ни* потенциально возможны, так как выражают лишь **предполагаемую** множественность причин и целей и встретились в корпусе. Этим же объясняется возможность (хотя и ограниченная) местоимений *почему* и *зачем* соединяться с маркером неопределенности *-нибудь*, который допускает потенциальную множественность[7]: *Если почему-*

---

[7] Мы благодарим за это соображение и за примеры нашего анонимного рецензента.

нибудь я его не найду, то просто-напросто позову полицейского сыскного агента и прикажу отыскать (А.И. Куприн); *А если почему-нибудь не сможешь приехать в Ялту, то пусть она их вышлет…* (Л. Вертинская); *Вероятно, надо зачем-нибудь, чтобы я погиб, вот и подсластили финал* (Д. Быков).

4.8. Конструкции с *хоть*

Все вопросительные местоимения, кроме *зачем* и *почему*, употребляются в конструкции с частицей *хоть*:

(72) *Она сидела до зари / И говорила: – Подари / Хоть шаль, хоть что, хоть полушалок* (О. Мандельштам).

В этой конструкции местоимения выступают в разговорном и/или устаревшем значении, отчасти синонимичном неопределенным местоимениям с *-нибудь* и *-то*: *Подари хоть что ≈ Подари хоть что-нибудь*. Это значение весьма редко, так что общее количество таких сочетаний невелико даже у самых частотных местоимений. Значение частицы *хоть* в данном сочетании таково (мы приводим отчасти модифицированное толкование из В. Апресян 2015: 239):

(73) 'Считая, что желательная ситуация А2 невозможна, говорящий или субъект готов удовлетвориться любым А1 из того же класса ситуаций, что и А2'.

Иными словами, смысл данной конструкции таков: говорящий соглашается на любой, даже самый нестандартный, маленький, плохой или иным образом неподходящий объект (признак, способ, место и т.д.) в качестве участника ситуации. Таким образом, конструкция с *хоть* – это разновидность конструкции свободного выбора, в сочетании с семантикой уступительности: что угодно (свободный выбор), даже самое малое или неподходящее (уступительность). Соответственно, в конструкции с *хоть* каждое из местоимений семантически трансформируется, обычно следующим образом: *хоть что* – это 'что-то небольшое или недостаточно хорошее', *хоть кто* – это 'кто-то, не удовлетворяющий стандартам заинтересованного участника ситуации', *хоть как* 'не очень хорошо', *хоть какой* 'не очень хороший', *хоть сколько* 'немного', *хоть где* 'в не очень хорошем или подходящем месте' и т.д.

Часто конструкция с *хоть* интерпретируется именно в значении малого и плохого полюса, как в примерах выше, однако возможна и противоположная интерпретация; ср.:

(74) *– С тобой говорить, Евпраксия, невозможно, до того твой характер испортился. – Вы хоть чей характер испортите!* (А.Ф. Писемский);

(75) *Жених хоть кому, а всё-таки учители ходят, часа не теряет* (Д.И. Фонвизин) ['жених любой, даже самой хорошей невесте'].

Для неопределенных местоимений вторая интерпретация невозможна: ср. естественность

(76)   *Нам бы хоть куда-нибудь на каникулы выбраться, хоть в деревню*

при странности

(77)   *??Они на каникулы могут хоть куда-нибудь поехать – хоть на Мальдивы, хоть на Сейшелы.*

Ниже приводятся данные сочетаемости этой конструкции с разными местоимениями. Для местоимений *почему и зачем* употребление с *хоть* невозможно – сочетаний **хоть почему,* *хоть зачем в НКРЯ не встретилось. Это связано опять-таки с тем, что семантика свободного выбора предполагает выбор и множественность, которые несовместимы с семантикой причины и цели.

Примеры были вручную очищены от «шума» – например, от фраз с запятой после *хоть*: *Хорошо хоть, что…*; *Ты знаешь хоть, куда…* и т.д. Кроме того, была исключена из рассмотрения предикативная идиома *хоть куда* со значением 'очень хороший'.

Таблица 9. **Частотность местоимений с *хоть***

Местоимение	Общая частотность в НКРЯ	Вхождения с *хоть*
что	3707974	150
как	1910556	90
когда	572008	18
кто	355623	**303**
какой	342127	**237**
где	323725	16
сколько	89573	57
куда	97225	≈ **103** (≈24 % от 429
откуда	41195	2
чей	18184	2

Данные распределения местоимений по конструкциям не отражают их относительной частотности, а мотивируются семантикой конструкции. Она вводит представление о некоторой шкале, по которой выстроены элементы множества, причем говорящий соглашается даже на элементы, находящиеся на полюсах – обычно на плохом и/или малом. Наиболее частотны сочетания с местоимениями, которые обозначают градуируемые с точки зрения качества или масштаба сущности – например, признаки (*какой*), но в гораздо меньшей степени время (*когда*).

4.9. Дистрибутивная конструкция

Дистрибутивная биместоименная конструкция вида X Y (*кто как, когда как, кто о чем, кому что*) и пр., выражающая смысл 'Разным X-ам свойственны разные Y', рассматривается в работах (Шведова 1998, Кустова 2016,

Кустова, Добровольский 2020), а также подробно анализируется на базе корпусного исследования в (В. Апресян, Копотев 2021, 2022), поэтому здесь она не разбирается. Отметим лишь, что у нее есть несколько наиболее частотных реализаций (*кто как, кому как, когда как, кто о чем, кто куда*), есть невозможные реализации (**зачем почему,* *почему зачем) и есть большое количество потенциальных или редких реализаций (*кому чей, что зачем* и пр.). Частотность реализаций этой конструкции регулируется причинами, которые подробно рассматриваются в работах (В. Апресян, Копотев 2021, 2022). Заметим кратко, что к дистрибутивным конструкциям тяготеют, в первую очередь, местоимение *кто* в позиции первого элемента и местоимения места, направления, способа (*куда, где, как*) в позиции второго элемента, что объясняется обсуждавшимися семантическими, прагматическими и коммуникативными причинами.

5. Заключение

В заключение можно отметить: полученные результаты свидетельствуют о том, что большая часть русских вопросительных местоимений действительно представляет собой лексикографический тип, объединенный сходной семантической деривацией (полисемия), а также сочетаемостными и конструкционными свойствами, который должен получить единое лингвистическое описание, в частности, в словаре. Местоимения *зачем* и *почему* противопоставлены остальным вопросительным местоимениям в большей части своих языковых свойств. В частности, они не развивают значений, которые свойственны другим РВМ, и не употребляются в конструкциях и лексических контекстах, характерных для остальных РВМ. Как было показано, это связано с их семантическими и прагматическими особенностями. Большая часть сочетаемостных и конструкционных свойств РВМ связана с идеей выбора, которая не сочетается с семантикой *зачем* и *почему*, в силу того, что в языке заложено представление о невозможности выбора причины, вызывающей ситуацию, и цели, мотивирующей действие. Полученные результаты могут быть использованы при описании вопросительных местоимений в других языках.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

Апресян В.Ю. Фраземы с наречиями малого количества: *мало ли // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007»* (Бекасово, 30 мая – 3 июня 2007 г.) / под ред. Л.Л. Иомдина,

Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея. М.: Издательский центр РГГУ, 2007. С. 16–22. [Apresyan, Valentina Yu. 2007. Idioms with adverbs of small quantity: *Malo li.* In Leonid L. Iomdin, Nataliya I. Laufer, Aleksandr S. Narin'yani & Vladimir P. Selegei (eds.), *Computation linguistics and intellectual technologies*. Moscow: RGGU Publ. (In Russ.)].

Апресян В.Ю. Уступительность: механизмы образования и взаимодействия сложных значений в языке. М.: Языки славянской культуры, 2015. [Apresyan, Valentina Yu. 2015. Ustupitel'nost': mekhanizmy obrazovaniya i vzaimodeistviya slozhnykh znachenii v yazyke (Concession: Mechanisms for the formation and interaction of complex meanings in language). Moscow: Yazyki slavyanskoy kul'tury. (In Russ.)].

Апресян В.Ю. Русские конструкции с синтаксической редупликацией цветообозначений: корпусное исследование // *Russian Journal of Linguistics*. 2018. Т. 22. № 3. С. 653–674. [Apresyan, Valentina Yu. 2018. Russian constructions with syntactic reduplication of colour terms: A corpus study. *Russian Journal of Linguistics* 22 (3). 653–674. (In Russ.)].

Апресян В.Ю., Копотев М.В. Абсолютивные дистрибутивные конструкции с вопросительными местоимениями в русском языке. *Вопросы языкознания*. В печати. 2022. [Apresyan, Valentina Yu. & Mikhail V. Kopotev. In print. Absolyutivnye distributivnye konstruktsii s voprositel'nymi mestoimeniyami v russkom yazyke (Absolute distributive constructions with interrogative pronouns in Russian). *Voprosy yazykoznaniya*. (In Russ.)].

Апресян В.Ю., Копотев М.В. *Кто о чем, а вшивый о бане.* В печати. 2022. [Apresyan, Valentina Yu. & Mikhail V. Kopotev. In print. *Kto o chem, a vshivyi o bane (To harp on)*. (In Russ.)].

Апресян Ю.Д. Правила взаимодействия значений и словарь // *Русский язык в научном освещении*. 2005. Т. 1. № 9. С. 7–45. [Apresyan, Yurij D. 2005. The rules of interaction of meanings and the dictionary. *Russkii yazyk v nauchnom osveshchenii* 1 (9). 7–45. (In Russ.)].

Апресян Ю.Д. Исследования по семантике и лексикографии в 2-х томах. Т. 1. Парадигматика. М.: Языки славянских культур, 2009. [Apresjan, Yuriy D. 2009. Issledovaniya po semantike i leksikografii. T. 1. Paradigmatika (Studies in semantics and lexicography. Vol 1. Paradigmatics). Moscow: Yazyki slavyanskikh kul'tur Publ. (In Russ.)].

Апресян Ю.Д. Введение // *Проспект активного словаря русского языка*. Авторы: В.Ю. Апресян, Ю.Д. Апресян, Е.Е. Бабаева, О.Ю. Богуславская, И.В. Галактионова, М.Я. Гловинская, Б.Л. Иомдин, Т.В. Крылова, И.Б. Левонтина, А.В. Птенцова, А.В. Санников, Е.В. Урысон. Отв. ред. Ю.Д. Апресян. Москва, 2010. С. 17–54. [Apresjan, Yuriy D. 2010. Introduction. In Valentina Yu. Apresyan, Yurij D. Apresyan, Elizaveta E. Babaeva, Olga Yu. Boguslavskaya, Irina V. Galaktionova, Marina Ya. Glovinskaya, Boris L. Iomdin, Tatiana V. Krylova, Irina B. Levontina, Anna V. Ptensova, Andrei V. Sannikov, Elena V. Uryson. 2010. *Prospect of the Active Dictionary of the Russian Language*. Moscow: Yazyki slavyanskikh kultur Publ. 17–54. (In Russ.)].

Апресян Ю.Д., Иомдин Л.Л. Конструкции типа НЕГДЕ СПАТЬ в русском языке: синтаксис и семантика // *Семиотика и информатика*. М., 1990. Вып. 29. С. 3–89. [Apresyan, Yuriy D. & Leonid L. Iomdin. 1990. Konstruktsii tipa NEGDE SPAT' v russkom yazyke: sintaksis i semantika (Constructions of the *negde spat'* type in Russian: Syntax and semantics). *Semiotika i informatika* 29. 3–89. (In Russ.)].

Булыгина Т.В., Шмелев А.Д. Языковая концептуализация мира (на материале русской грамматики). М.: Школа «Языки русской культуры», 1997. [Bulygina, Tatyana V. & Aleksey D. Shmelev. 1997. Yazykovaya kontseptualizatsiya mira (na materiale russkoi

grammatiki) (Linguistic conceptualization of the world (based on Russian grammar)). Moscow: Shkola "Yazyki russkoi kultury". (In Russ.)].

Добровольский Д.О., Кустова Г.И. Русские дистрибутивные конструкции типа *кто куда*: когнитивные стереотипы и межъязыковые соответствия // *Слова, конструкции и тексты в истории русской письменности*. Сб. статей к 70-летию академика А.М. Молдована. М., СПб.: Нестор-История, 2021. С. 357–378. [Dobrovol'skii, Dmitrii O. & Galina I. Kustova. 2021. Russkie distributivnye konstruktsii tipa *kto kuda*: kognitivnye stereotipy i mezh"yazykovye sootvetstviya (Russian distributive constructions like *kto kuda*: Cognitive stereotypes and interlingual correspondences). *Words, Constructions and Texts in the History of Russian Writing*. Moscow, St. Petersburg: Nestor-Istoriya. (In Russ.)].

Добровольский Д.О., Левонтина И.Б. О синонимии фокусирующих частиц (на материале немецкого и русского языков) // *Компьютерная лингвистика и интеллектуальные технологии*: По материалам ежегодной Международной конференции «Диалог». 2012. № 11. С. 138–150. [Dobrovolskii, Dmitriy O. & Irina B. Levontina. 2012. Synonymous focus particles in German and Russian. *Computational Linguistics and Intellectual Technologies* 11. 138–150. (In Russ.)].

Зализняк Анна А., Падучева Е.В. Русское адъективное местоимение ВСЯКИЙ: семантика и идиоматика // *Вопросы языкознания*. 2020. № 1. С. 7–26. [Zalizniak, Anna A. & Elena V. Paducheva. 2020. Russian adjective pronoun *vsyakii*: Semantics and idiomatics. *Voprosy yazykoznaniya* 1. 7–26. (In Russ.)].

Зевахина Н.А. Сложноподчиненные предложения с восклицательной интерпретацией в русском языке // *Русский язык в научном освещении*. 2018. № 1. Т. 36. С. 200–227. [Zevakhina, Natalya A. 2018. Slozhnopodchinennye predlozheniya s vosklitsatel'noi interpretatsiei v russkom yazyke (Complex sentences with exclamatory interpretation in Russian). *Russian Language and Linguistic Theory* 1 (36). 200–227. (In Russ.)].

Иомдин Л.Л. Новые наблюдения над синтаксисом русских фразем // Bozhena Chodzhko, Elzhbieta Feliksiak, Marek Olesiewicz (eds.). *Obecnosc*. Białystok: Uniwersytet w Białymstoku, 2006. С. 247–281. [Iomdin, Leonid L. 2006. Novye nablyudeniya nad sintaksisom russkikh frazem (New observations on the syntax of Russian phrasemes). In Bożena Chodźko, Elżbieta Feliksiak & Marek Olesiewicz (eds.), *Obecnosc*, 247–281. Białystok: Uniwersytet w Białymstoku. (In Russ.)].

Иомдин Л.Л. Русские конструкции малого синтаксиса, образованные вопросительными местоимениями // *Мир русского слова и русское слово в мире*. Материалы XI Конгресса Международной ассоциации преподавателей русского языка и литературы. Sofia: Heron Press, 2007. С. 117–126. ISBN 978-954-580-213-3. [Iomdin, Leonid L. 2007. Russkie konstrukcii malogo sintaksisa, obrazovannye vopsrositelnymi mestoimeniyami (Russian minor syntax constructions formed by interrogative pronouns). 117–126. (In Russ.)].

Иорданская Л.Н., Мельчук И.А. Местоимения *кто* и *который* в определительном придаточном // *Труды Института русского языка им. В.В. Виноградова*. 2020. № 2. С. 107–136. [Iordanskaja, Lidiya N. & Igor A. Melchuk. 2020. Pronouns *kto* 'who' and *kotoryi* 'which' in a relative clause. *Trudy Instituta russkogo yazyka im V. V. Vinogradova* 2. 107–136. (In Russ.)].

Кобозева И.М. О первичных и вторичных функциях вопросительных предложений // *Текст в речевой деятельности: Перевод и лингвистический анализ*. М.: Институт языкознания АН СССР, 1988. С. 39–45. [Kobozeva, Irina M. 1998. O pervichnykh i vtorichnykh funktsiyakh voprositel'nykh predlozhenii (On the Primary and Secondary functions of interrogative sentences). *Text in Speech Activity: Translation and Linguistic Analysis*. Moscow: Linguistics Instutute. 39–45. (In Russ.)].

Козлов А.А. *Именно*. [Электронный ресурс] // https://www.academia.edu/45577248/Именно (дата обращения: 30.03.2022). [Kozlov, Aleksey A. 2020. Imenno (*Namely*). https://www.academia.edu/45577248/Именно (accessed 30 March 2022) (In Russ.)].

Кустова Г.И. Дистрибутивные биместоименные конструкции типа *кто куда* // *Компьютерная лингвистика и интеллектуальные технологии*: По материалам ежегодной Международной конференции «Диалог». № 11. М.: РГГУ, 2016. С. 355–369. [Kustova, Galina I. 2016. Distributive bipronominal constructions. *Computational Linguistics and Intellectual Technologies* 11. 355–369 (In Russ.)].

Левонтина И.Б., Шмелев А.Д. Малоизученные единицы со значением незаданности критериев выбора в русском языке // *Логический анализ языка. Квантификативный аспект языка* / под ред. Н.Д. Арутюновой. М.: Индрик, 2005. С. 638–651. [Levontina, Irina B. & Aleksey D. Shmelev. 2005. Understudied items with the meaning of unspecified choice criteria in Russian. In Nina D. Arutyunova (ed.), *Logicheskii analiz yazyka. Kvantifikativnyi aspekt yazyka*, 638–651. Moscow: Indrik. (In Russ.)].

Левонтина И.Б., Шмелев А.Д. *Абы*: корпусное исследование в аспекте синхронии и диахронии // *Компьютерная лингвистика и интеллектуальные технологии*: По материалам ежегодной Международной конференции «Диалог». 2018. № 17. Т. 24. С. 355–369. [Levontina, Irina B. & Aleksey D. Shmelev. 2018. The Russian *aby*: Corpus-driven research (synchrony and diachrony). *Computational Linguistics and Intellectual Technologies* 17 (24). 436–447. (In Russ.)].

Мельчук И.А. Опыт теории лингвистических моделей «Смысл ⇔ Текст». М., 1974 (2-е изд., 1999). [Melchuk, Igor A. 1974. Opyt teorii lingvisticheskih modelei Smysl ⇔ Tekst (An experience of a theory of linguistic models: Meaning ⇔ Text). Moscow: Nauka. (In Russ.)].

Мельчук И.А. Русский язык в модели «Смысл ⇔ Текст». Москва-Вена, 1995. [Melchuk, Igor A. 1995. Russkii yazyk v modeli Smysl ⇔ Teskt (Russian Language in Meaning ⇔ Text Theory). Moscow–Vienna. (In Russ.)].

Мельчук И.А., Жолковский А.К., Апресян Ю.Д., Альперин В.А, Апресян Т.А., Бабицкий К.И., Бирюлин Л.А., Богданова Е.Ю., Богуславский И.М., Вольф Е.М., Гончар Л.А., Ельницкий Л.Л., Иорданская Л.Н., Казавчинская Т.Я., Коссек Н.В., Крысин, Л.Л., Никитина С.Е., Перцов Н.В., Разлогова Е.Э., Саввина Е.Н., Санников В.З., Телия В.Н., Шаляпина З.М. Толково-комбинаторный словарь современного русского языка: Опыты семантико-синтаксического описания русской лексики. Wien: Wiener Slavistischer Almanach, 1984. [Melchuk, Igor A., Alexander K. Zholkovsky, Yurij D. Apresyan, Vadim A. Alperin, Tamara A. Apresjan, Konstantin I. Babickij, Leonid A. Birjulin, Elena Ju. Bogdanova, Igor M. Boguslavskij, Elena M. Volf, Ludmila A. Gonchar, Elnitsky, L.L., Lidia N. Iordanskaja, Tamara Ja. Kazavčinskaja, Natalia V. Kossek, Leonid P. Krysin, Serafima E. Nikitina, Nikolai V. Percov, Elena E. Razlogova, Vladimir Z. Sannikov, Elena N. Savvina, Zoya M. Shalyapina, Veronika N. Telia. 1984. *Explanatory Combinatory Dictionary of Russian*. Moscow–Vienna. (In Russ.)].

Мостовая Л.А. Отрицание в предложениях с вопросительными словами *зачем* и *почему* // *Русский язык в школе*. 2009. № 7. С. 65–69. [Mostovaya, Lidiya A. 2009. Otricanie v predlozheniyah s voprositel'nymi slovami *zachem* i *pochemu* (Negation in sentences with question words *what for* and *why*). *Russian Language at School* 7. 65–69. (In Russ.)].

Падучева Е.В. Высказывание и его соотнесенность с действительностью: Референциальные аспекты семантики местоимений. М.: Изд-во ЛКИ, 2010. [Paducheva, Elena V. 2010. Vyskazyvanie i ego sootnesennost' s deistvitel'nost'yu: Referentsial'nye aspekty semantiki mestoimenii (The utterance and its correlation with reality: Referential aspects of the semantics of pronouns). Moscow: LKI Publ. (In Russ.)].

Падучева Е.В. Русские местоимения свободного выбора // *Russian Linguistics.* 2018. № 3. Т. 42. С. 291–319. [Paducheva, Elena V. 2018. Russian free-choice pronouns. *Russian Linguistics* 42 (3). 291–319. (In Russ.)].

Тестелец Я.Г., Былинина Е.Г. О некоторых конструкциях со значением неопределенных местоимений в русском языке: амальгамы и квазирелятивы // *Хэндаут доклада*, представленного на семинаре ИППИ РАН «Теоретическая семантика. 2005. Т. 15. [Testelets, Yakov G., & Elizaveta G. Bylinina. 2005. O nekotorykh konstruktsiyakh so znacheniem neopredelennykh mestoimenii v russkom yazyke: amal'gamy i kvazirelyativy (On some constructions with the meaning of indefinite pronouns in Russian: Amalgams and quasi-relatives). *Handout of a talk* at the IITP RAS Theoretical semantics seminar. Vol. 15. (In Russ.)].

Шведова Н.Ю. Местоимение и смысл: класс русских местоимений и открываемые ими смысловые пространства. М.: Азбуковник, 1998. [Shvedova, Natalya Yu. 1998. Mestoimenie i smysl: klass russkikh mestoimenii i otkryvaemye imi smyslovye prostranstva (Pronoun and meaning: The class of Russian pronouns and the semantic spaces they open). Moscow: Azbukovnik. (In Russ.)].

Feldstein, Ronald. 2016. Russian Concessives/Adversatives Using Reduplicated Pronouns of the type "Уж чем-чем, а.." (Uzh chem-chem, a..). *Trudy Instituta russkogo yazyka im. V.V. Vinogradova RAN*. Materialy mezhdunarodnoi nauchnoi konferentsii "Grammaticheskie protsessy i sistemy". Moscow. 327–341.

Iomdin, Leonid. 2007. Russian idioms formed with interrogative pronouns and their syntactic properties. *Meaning – Text Theory*. Proceedings of the 3rd International Conference on Meaning – Text Theory. Klagenfurt, Austria. May 21–24, 2007. Wiener Slawistischer Almanach. Sonderband 69. München – Wien. ISSN 0258-6835. ISBN 978-3-87690-xxx-x. 179–189.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2). 209–243.

Словари

АС – Активный словарь русского языка / Отв. ред. акад. Ю. Д. Апресян. Тт. 1–2: А–Г. М.: «Языки славянских культур», 2014. [Yuriy D. Apresyan (ed.), Aktivnyy slovar' russkogo yazyka. Vols. 1–2. Moscow: Yazyki slavyanskoy kul'tury. (In Russ.)].

БАС – Большой академический словарь русского языка / гл. ред. К. С. Горбачевич, А. С. Герд и др. М.; СПб.: Наука, 2004–. Т. 1–. [Bol'shoi akademicheskii slovar' russkogo yazyka (The Great Academic Dictionary of the Russian Language]) (Vols. 1–). K. S. Gorbachevich, A. S. Gerd et al. (Eds.). Moscow, St Petersburg, Nauka Publ., 2004–.]

БТС – Большой толковый словарь русского языка / Сост., гл. ред. С.А. Кузнецов. СПб.: Норинт, 1998 [Kuznetsov S. A. (ed.). (1998), Large Explanatory Dictionary of Russian. Norint, St. Petersburg].

МАС – Словарь русского языка : в 4 т. / под ред. А.П. Евгеньевой. 2-е изд. испр. и доп. М. : Русский язык, 1981–1984. [Evgenyeva A. P. (ed.)., Russian Language Dictionary. 1981–1984. Moscow: Russian Language Publ.].

НКРЯ – Национальный корпус русского языка (старая версия). https://ruscorpora.ru/old/.

СОШ – Ожегов С.И. и Н.Ю. Шведова. Толковый словарь русского языка. М.: Азъ, 1992 [Ozhegov S.I., Shvedova N.Yu., Tolkovyj slovar' russkogo jazyka (Explanatory Dictionary of Russian). 1992. Moscow: Az].

СУш – Толковый словарь русского языка / Под ред. Д.Н. Ушакова. М.: Гос. ин-т Сов.эн-цикл.; ОГИЗ; Гос. изд-во иностр. и нац. словарей, 1934–1940 [Ushakov D.N. (ed.).

(1934-1940), Tolkovyj slovar' russkogo jazyka (Explanatory Dictionary of Russian). Moscow: OGIZ].

Bionotes:
Valentina APRESYAN holds a PhD in Linguistics from the University of Southern California. She is currently an Associate Professor at the School of Linguistics, Higher School of Economics (HSE University). She is also a leading researcher at the Sector for Theoretical Semantics, Vinogradov Russian Language Institute at the Russian Academy of Sciences. Her research interests fall within semantics and pragmatics. She has also done extensive work in lexicography. Her works have been published, among others, in *Intercultural Pragmatics*, *Russian Linguistics*, *Voprosy yazykoznaniya*.
e-mail: vapresyan@hse.ru
https://orcid.org/0000-0003-1348-2224

Boris IOMDIN currently is a Leading Researcher at the Sector for Theoretical Semantics at Vinogradov Russian Language Institute, Moscow, and Researcher at the Russian Language Sector in Yandex. His research interests include: lexicography, semantics, variability, semantic complexity, neologisms, and self-sufficient linguistic problems.
e-mail: iomdin@ruslang.ru
https://orcid.org/0000-0002-1767-5480

Сведения об авторах:
Валентина Юрьевна АПРЕСЯН имеет докторскую степень по лингвистике Университета Южной Калифорнии. В настоящее время является профессором Школы лингвистики НИУ ВШЭ и ведущим научным сотрудником сектора теоретической семантики Института русского языка им. Виноградова РАН. Ее исследовательские интересы относятся к семантике и прагматике. Она также ведет обширную работу в области лексикографии. Имеет многочисленные публикации, в том числе в журналах "Intercultural Pragmatics", "Russian Linguistics", «Вопросы языкознания».
e-mail: vapresyan@hse.ru
https://orcid.org/0000-0003-1348-2224

Борис Леонидович ИОМДИН – в настоящее время ведущий научный сотрудник Сектора теоретической семантики Института русского языка имени В.В. Виноградова в Москве и научный сотрудник сектора русского языка Яндекса. Область научных интересов включает лексикографию, семантику, вариативность, семантическую сложность, неологизмы, самодостаточные лингвистические задачи.
e-mail: iomdin@ruslang.ru
https://orcid.org/0000-0002-1767-5480

Research article / Научная статья

Aspectual pairs:
Prefix vs. suffix way of formation

Valery SOLOVYEV ⊠, **Vladimir BOCHKAREV**
and Venera BAYRASHEVA

Kazan Federal University, Kazan, Russia
⊠maki.solovyev@mail.ru

Abstract
In linguistic theory, there is no common point of view on the question of whether verbs in aspectual pairs are in inflectional or derivational relations. At the same time, the prefix and suffix methods of forming aspectual pairs are contrasted in this respect. The publications (e.g. Janda & Lyashevskaya 2011) pointed out the need to develop new quantitative approaches to this aspect of the text corpus. We propose two new approaches that compare the quantitative characteristics of aspectual pairs of both types. One approach is based on the Google Books Ngram corpus and analyzes the dynamics of the frequency of the use of words in pairs. The aspectual pairs from the databases created by Janda and Lyashevskaya are considered. For a numerical assessment of the degree of proximity of the frequency graphs, the Pearson correlation coefficients were used. The second approach introduces a numerical characteristic of the semantic proximity of verbs in pairs using modern computer methods. Semantic proximity of verbs is calculated as a standard cosine measure between vectors representing the compatibility of the considered verbs in the corpus. Several computer models and text corpora are considered. Both proposed approaches did not reveal significant numerical differences in semantic proximity between verbs in aspectual pairs with prefix and suffix pairing. This is in good agreement with the results of an early study by Janda and Lyashevskaya (2011). Together with the results of this work, our research shows that the suffixal and affixal ways of forming aspectual pairs have an equal status in terms of their classification as inflectional or derivational.
Keywords: *aspectual pairs, inflection, word formation, quantitative analysis, corpora, Russian*

Аспектуальные пары: префиксальный *vs.* суффиксальный способ образования

В.Д. СОЛОВЬЕВ ✉, В.В. БОЧКАРЕВ, В.Р. БАЙРАШЕВА

Казанский федеральный университет, Казань, Россия
✉maki.solovyev@mail.ru

Аннотация

В литературе нет общей точки зрения по вопросу, находятся ли глаголы в аспектуальных парах в словоизменительных или словообразовательных отношениях, при этом префиксальный и суффиксальный способы образования аспектуальных пар часто противопоставляются в этом отношении. В публикациях (например, Janda & Lyashevskaya 2011) указывалось на необходимость развития в этом вопросе новых квантитативных подходов на корпусах текстов. В данной статье мы предлагаем два новых подхода, сопоставляющих квантитативные характеристики аспектуальных пар русских глаголов обоих типов. Один подход основан на корпусе Google Books Ngram и анализирует динамику частот употребления слов в парах. Рассматриваются аспектуальные пары из баз данных, созданных Л. Яндой и О. Ляшевской. Для численной оценки степени близости графиков частот использован коэффициенты корреляции Пирсона. Второй подход вводит числовую характеристику семантической близости глаголов в парах с использованием современных компьютерных методов. Семантическая близость глаголов рассчитывается как стандартная косинусная мера между векторами, репрезентирующими сочетаемость рассматриваемых глаголов в корпусах. Рассмотрено несколько компьютерных моделей и корпусов текстов. Ни один из предложенных подходов не выявил существенных численных различий в степени семантической близости между глаголами в аспектуальных парах при префиксальном и суффиксальном способах образования пар. Это находится в хорошем согласии с результатами раннего исследования Л. Янды и О. Ляшевской (Janda & Lyashevskaya 2011). Совместно с результатами этой работы наши исследования показывают, что суффиксальный и аффиксальный способы образования аспектуальных пар имеют равный статус с точки зрения их классификации как словоизменительных или словообразовательных.

Ключевые слова: *аспектуальные пары, словоизменение, словообразование, квантитативный анализ, корпусы, русский язык*

1. Введение

Русский вид (в зарубежной терминологии «аспект») является одной из наиболее сложных грамматических категорий, вызывающей много споров среди лингвистов и трудной для усвоения иностранцами. Значительный вклад

в формализацию и понимание природы этого понятия внесен И.А. Мельчуком в фундаментальной монографии "Курс общей морфологии" в 4-х томах. В настоящей работе мы опираемся на изложенные в ней представления.

Несмотря на длительную историю изучения русского вида остается неясным, как следует его трактовать с точки зрения грамматического статуса – находятся ли глаголы в аспектуальных парах в словоизменительных или словообразовательных отношениях. Основные точки зрения следующие.

1. Вид – словоизменительная категория (это классический подход В.В. Виноградова (Виноградов 1972, Мельчук 1997, Тихонов 1998).

2. Вид – словоклассифицирующая категория (Русская грамматика 1980, Петрухина 2000, Janda et al. 2013, Плунгян 2000).

3. Вид – словообразовательная категория (Всеволодова 1997, Filip 1999).

4. Префиксальное образование глаголов совершенного вида (перфективов) от базовых глаголов несовершенного вида (имперфективов) – это словообразование, а суффиксальное образование вторичных имперфективов от перфективов – словоизменение (пожалуй, наиболее часто представляемая точка зрения (Маслов 1978/2004, 1984, Бондарко 1971, Исаченко 1954/2003, Горбова 2017)).

Разумеется, эти работы различаются нюансами трактовки перечисленных понятий.

Таким образом, эти два способа видовой трансформации тесно связаны с грамматическим статусом аспектуальных пар. Отметим, что видовые пары образуются разным способом. Перфективы образуются из базовых имперфективов [1] с помощью приставок[2], а вторичные имперфективы образуются из перфективов добавлением суффиксов). Целью статьи является проверка квантитативными методами последней из перечисленных 4-х точек зрения. В последнее десятилетие проблема, является ли русский вид словоизменительной или словообразовательной категорией, активно исследовалась в работах Е.В. Горбовой и В.С. Храковского (Горбова 2014, 2015а, 2015б, 2017, Храковский 2015, 2018), однако к консенсусу они так и не пришли. Проблематика русского вида подробно освещена в монографии (Зализняк и др. 2015). По рассматриваемому вопросу точка зрения, представленная в этой монографии, следующая: «вопрос вообще не имеет однозначного решения… что отражает реально двойственную природу категории вида в русском языке» (с. 20).

К сожалению, общепризнанного формального (операционного) критерия разграничения словообразования и словоизменения тоже нет. В различных работах (Dressler 1989, Plank 1991, Перцов 1996, Мельчук 1997) обсуждались разные признаки этих категорий. Авторы работы (Singh 1980) вообще настаивают на отмене этого разграничения. В (Мельчук 1997) предлагается ввести промежуточную категорию квазиграммем. Все же есть общее интуитивное понимание различия между ними, хотя бы в прототипических случаях: при

---

[1] В последнее время используется также термин 'симплексы' (Gorbova 2022).

[2] Другое название – префигированные (Gorbova & Chuikova 2022).

словоизменении само слово не меняется, его лексическая семантика сохраняется, меняется лишь его форма, а при словообразовании меняется слово, т.е. его значение. Эта позиция представлена и в формализованном определении в (Мельчук 1997). Таким образом, мы приходим к вопросу о том, меняется ли семантика слова (кроме чисто видовой) при префиксальной и суффиксальной трансформации.

Однако, как известно, семантика не формализуема или крайне сложно формализуема. Предложенная А. Вежбицкой теория семантических примитивов (Вежбицкая 1999) к данной задаче плохо применима, прежде всего, в силу крайней сложности формул представления семантики, а также в силу сохраняющегося субъективизма в построении таких формул. Противоречивая оценка многих пар глаголов как аспектуальных или нет в работах (Зализняк, Микаэлян 2012, Кузнецова, Янда 2013) показывает, что интуиция даже ведущих лингвистов-аспектологов недостаточна для установления семантического тождества. Это приводит к сделанному (Janda & Lyashevskaya 2011: 201) выводу, что "At the level of intuitive analysis, it has not been possible to resolve this debate" и, следовательно, для решения этих проблем нужно применять статистические методы к корпусным данным.

Итак, исследовательский вопрос, решаемый в данной статье: в какой степени меняется семантика слов при суффиксальном и префиксальном способах образования аспектуальных пар? Проверяемая гипотеза: семантическая близость между глаголами в аспектуальных парах перфектив – вторичный имперфектив будет больше, чем между глаголами в парах базовый имперфектив – перфектив. В статье мы ограничиваемся рассмотрением основных способов образования аспектуальных пар: приставочного для образования перфектива и суффиксального для образования вторичного имперфектива.

В данной статье мы предлагаем два подхода к получению объективной численной оценки степени семантической близости слов, составляющих аспектуальные пары. Оба подхода применены к аспектуальным парам, получаемым префиксацией и суффиксацией. В итоге мы не нашли заметных количественных различий между этими двумя способами, которые позволили бы их четко дифференцировать как словообразовательные или словоизменительные. Это хорошо согласуется с работой (Janda & Lyashevskaya 2011), в которой также не найдено таких различий при применении еще одного количественного метода. В этой работе сравнивались грамматические профили – распределения форм глагола в Национальном корпусе русского языка – и был получен следующий результат: "we do not find reportable differences between the grammatical profiles of aspectual partners formed with prefixes as opposed to suffixes" (Janda & Lyashevskaya 2011: 213).

В этой статье мы используем предложенную в (Janda et el. 2013) терминологию: перфективы в аспектуальных парах с базовыми имперфективами будем называть естественными, а иные перфективы – специализированными. Например, *прочитать* – естественный перфектив, составляющий аспектуальную пару с *читать*, а *дочитать* – специализированный перфектив.

2. Теоретические предпосылки

В этом разделе мы кратко представим необходимый теоретический материал. В соответствии с заявленными целями уточним принципиально важные для данного исследования понятия словообразования и словоизменения. Мы отталкиваемся от классического труда И.А. Мельчука «Курс общей морфологии» в 4-х томах. Хотя эти понятия и хорошо известны, все же И.А. Мельчук в своей монографии подробно их рассматривает. Это связано, в частности, с тем, что между словоизменением и словообразованием нет четкой границы, эти понятия, скорее, градуированные.

Словоизменение и словообразование уместно рассматривать с двух точек зрения: семантико-лексикографической и формальной. Под семантико-лексикографической понимается вопрос: являются ли два слова, находящихся в словоизменительных/словообразовательных отношениях, вариантами одной лексемы или это две разных лексемы. Это определяет в том числе и представление слов в словарях – должны ли они описываться в одной словарной статье или в двух разных.

В (Мельчук 1997) автор исходит из следующих постулатов: «Дериватемы различают разные лексемы» (с. 274), где под дериватемой понимается выражаемое морфологическими средствами словообразовательное значение. В этой монографии используется также понятие граммемы как значения, принадлежащего к словоизменительной категории. При этом постулируется (с. 258): «граммемы различают **формы** “одного и того же слова”» (выделение И.А. Мельчука). Таким образом, по Мельчуку, при словоизменении мы остаемся в пределах «одного и того же слова», т.е. лексическая семантика не меняется. А при словообразовании получаются две разные лексемы с разной семантикой (учитывая, что абсолютная синонимия практически невозможна).

И.А. Мельчук предлагает также формальное определение словоизменительной категории. Чтобы не усложнять изложение в данной статье, мы отсылаем читателя к первоисточнику (Мельчук 1997: 249). Чуть уточненный вариант этого определения приведен в свежей статье И.А. Мельчука с соавторами (Иорданская и др. 2020). Содержательно это определение сводится к двум требованиям – обязательности и регулярности словоизменения. Словообразование не обязано обладать этими свойствами (Мельчук 1997: 271). Русский вид (под названием ‘Аспект V’) описывается в этой монографии в главе II Семантические словоизменительные значения (Мельчук 1998).

Свойства обязательности и регулярности словоизменения рассматривались и в других работах. Так, в (Сичинава 2011) при обсуждении грамматического статуса вида приводятся следующие доводы (далее – цитата):

- выражение при приставочной перфективации совместно со словообразовательными значениями (в пользу словообразовательной трактовки);
- отсутствие общих морфологических показателей вида (против словоизменительной трактовки);

- однозначная охарактеризованность основного массива словоформ по виду (в пользу словоизменительной трактовки);
- высокая регулярность образования вторичных имперфективов (в пользу словоизменительной трактовки).

К сожалению, применение только формальных параметров не позволяет однозначно классифицировать вид. Так, регулярность является градуальным понятием, и не вполне очевидно, трактовать ли применение набора приставок при префиксальном образовании перфектива как регулярное или нет. Ряд более тонких формальных характеристик рассмотрены в работе (Горбова 2017).

Вернемся к семантико-лексикографическому подходу. Как уже отмечалось, критическим является вопрос, совпадает ли лексическая семантика у членов аспектуальной пары. Для глаголов пар 'перфектив – вторичный имперфектив' большинство исследователей считают, что их семантика идентична. Известный пример Ю.Д. Апресяна (Апресян 1995: 112) *пить – выпить – выпивать* демонстрирует наличие дополнительной лексической семантики у вторичного имперфектива. Более того, это отнюдь не единичный случай. В (Сичинава 2011) отмечено, что вторичный имперфектив «семантически сдвинут в сторону итератива». Несовпадение лексической семантики в этих парах трактуется в (Горбова 2017) как «следствие лексикализации (деграмматикализации) грамматической формы».

Соотношение семантики базового имперфектива и приставочного перфектива является предметом давних дискуссий. Совпадение их лексической семантики эквивалентно отсутствию собственной семантики у приставок в составе перфектива. Отправной является точка зрения (Виноградов 1972, Тихонов 1998), что в аспектуальных парах приставки этого типа являются «чистовидовым», то есть не несут собственной семантики, а выражают только вид. Однако есть и альтернативные точки зрения. Гипотеза Вея–Схоневелда (Schooneveld 1978) гласит, что эти приставки имеют собственную семантику, но она накладывается на семантику корня. В последнее десятилетие идея наличия собственной семантики у приставок в аспектуальных парах имперфектив – приставочный перфектив активно отстаивалась в работах Л. Янды (Янда 2012, Janda et al. 2013, Кузнецова, Янда 2013). Л. Янда трактует вид как классифицирующую категорию.

В.А. Плунгян также считает русский вид словоклассифицирующей категорией, хотя и с оговорками (Плунгян 2000: 125). В этой работе приводятся интересные наблюдения типологического и диахронического характера. По данным В.А. Плунгяна, во многих языках мира аспект носит словоизменительный характер с тенденцией превращения в словоклассифицирующую категорию (Плунгян 2000: 294). Суффиксальные аспектуальные пары В.А. Плунгян трактует как словоизменительные, однако отмечает, что такие пары не являются доминирующими. Отмечается, что в последние 200–300 лет число таких пар возрастает (Плунгян 2000: 126). Подход В.А. Плунгяна к словоизменению/словообразованию в целом близок к подходу И.А. Мельчука, но менее формализован.

Отметим интересную работу (Иорданская и др. 2020), в которой пары глаголов движения типа *лететь/летать* трактуются как находящиеся в словоизменительном отношении, несмотря на лексикографическую традицию их описания в разных словарных статьях. В работе обосновывается точка зрения, что различия в семантике глаголов в таких парах можно описать как грамматические, определяемые граммемами *однонаправленное/ненаправленное* движение. Это пример того, как, казалось бы, разные слова могут быть представлены как формы одного слова. Таким образом, разграничение словоизменение и словообразования является совершенно нетривиальной проблемой.

3. Данные и методы

3.1. Подход 1: динамика частотности употребления слова

Первый из двух предлагаемых нами подходов основан на использовании корпуса Google Books Ngram (в дальнейшем сокращенно GBN). Корпус содержит для русского языка более 80 миллиардов слов (в 100 раз больше, чем Национальный корпус русского языка) и охватывает более 2 веков. Сервис визуализации Ngram Viewer (https://books.google.com/ngrams/) позволяет строить графики частот встречаемости слов и словосочетаний по годам.

Коллекция GBN содержит только тексты книг, и в этом смысле она не является всеохватывающей. С другой стороны, методика создания этого корпуса была такова (Michel et al. 2011): полное сканирование книг (более 15 % от всех изданных в мире книг) более чем из 40 крупнейших библиотек по всему миру, включая библиотеки университетов Оксфорда, Гарварда и др. (ENA, September 20, 2022)[3]. Это обеспечило репрезентативное представление всех аспектов человеческой культуры, в той мере, в которой они представлены в изданных книгах, хранящихся в ведущих библиотеках. Вопросы, связанные со сбалансированностью и репрезентативностью GBN, обсуждаются в (Solovyev 2019, Richey & Taylor 2020).

Наша основная идея состоит в следующем. Если два слова находятся в словоизменительном отношении и, соответственно, имеют тождественную лексическую семантику, то частота употребления этих двух слов будет меняться синхронно, в зависимости от степени востребованности слов с этой семантикой в тот или иной период времени. Если же два слова находятся в словообразовательных отношениях, т.е. представляют собой два разных слова, то частоты их употребления могут меняться не синхронно, в зависимости от востребованности каждого из этих слов. Это следует, в частности, из дистрибутивной гипотезы (Sahlgren 2008). Таким образом, для того, чтобы оценить префиксальный и суффиксальный способы образования аспектуальных пар с этой точки зрения, требуется сопоставить динамику частот словоупотребления соответствующих глаголов. В GBN употребление слова репрезентируется временным рядом частот в последовательные годы. Мы

---

[3] http://www.google.com/intl/en/googlebooks/library/partners.html

используем интервал времени с 1920 г., чтобы избежать влияния старой орфографии (до реформы 1917 г.), не всегда точно отражаемой в GBN.

Степень синхронности изменения частот употребления определяется близостью форм кривых – временных рядов и может быть подсчитана. Общей проблематике близости временных рядов и методам его оценки посвящена работа (Batyrshin et al. 2014). Для численной оценки степени близости будем использовать коэффициенты корреляции Пирсона, как часто используемый для оценки близости временных рядов. Данный подход был впервые анонсирован на конференции 15th Slavic Cognitive Linguistics Conference (Соловьев 2017). Обоснование подхода изложено в статье (Соловьев 2022).

Будут использованы два набора данных. Аспектуальные пары с префиксальным способом видообразования взяты из базы данных Exploring Emptiness (Janda et al. 2013), (ENA, September 20, 2022)[4]. Пары с суффиксальным типом видообразования – из базы данных видовых пар (О. Ляшевская 2016 (ENA, September 20, 2022)[5]).

При работе с временными рядами частот из GBN следует иметь в виду следующую особенность. Для низкочастотных слов возрастает влияние на форму графиков случайных факторов. Поэтому целесообразно анализировать по отдельности слова с разной частотностью. Будем сравнивать только слова, отличающиеся по частоте не более, чем в 10 раз. Т.е. делим все слова на группы с частотами f: $f > 10^{-4}$, $10^{-4} > f > 10^{-5}$ и т.д. Частота подсчитывается как средняя на интервале 1920–2008. Рассматриваются лишь аспектуальные пары, оба глагола которых принадлежат к одной частотной группе. В табл. 1 приведены усредненные коэффициенты корреляции по Пирсону временных рядов частот глаголов в аспектуальных парах.

Таблица 1. **Средний коэффициент корреляции Пирсона частот глаголов в аспектуальных парах/**
Table 1. **Average Pearson correlation coefficient of verb frequencies in aspectual pairs**

Частота, log10	Префиксальное видообразование	Суффиксальное видообразование
-4	0,65	0,61
-5	0,66	0,68
-6	0,53	0,58
-7	0,31	0,39
-8	0,19	0,27
-9	-	0,29
Всего пар	673	7800

Как видно из таблицы, для наиболее высокочастотных глаголов с $f > 10^{-4}$ корреляция оказывается несколько выше для префиксального способа видообразования. Для следующих двух частотных групп – для суффиксального способа видообразования, причем тоже не намного выше. Для низкочастотных глаголов в следующих группах коэффициент корреляции имеет небольшие значения в силу фактора, о котором было сказано выше. Таким образом, при таком подсчете существенной разницы не обнаруживается.

---

[4] http://emptyprefixes.uit.no/

[5] http://ru-eval.ru/go/resources.html

Одним из возможных влияющих на результат факторов может быть мно-гозначность глаголов. Это снижает значения коэффициентов корреляции, так как график частот словоупотреблений строится один для всех значений слов, усредняя частоты для всех его значений. Отберем некоторое число эталонных пар – достаточно высокочастотных глаголов, не являющихся омонимами или полисемичными словами с заведомо различными значениями, и рассчитаем коэффициенты корреляции для них. В табл. 2 приведены рассматриваемые пары слов. Список подготовлен Л. Яндой.

Расчет проводится так же, как и для полного набора аспектуальных пар. Результаты приведены в табл. 3.

Таблица 2. **Пары базовый имперфектив – естественный перфектив
и специализированный перфектив – вторичный имперфектив /**
Table 2. **Pairs of basic imperfective – natural perfective
and specialized perfective – secondary imperfective**

Базовый имперфектив	Естественный перфектив	Специализированный перфектив	Вторичный имперфектив
делать	сделать	переделать	переделывать
казаться	показаться	оказаться	оказываться
просить	попросить	выпросить	выпрашивать
ставить	поставить	приставить	приставлять
хранить	сохранить	охранить	охранять
прятать	спрятать	припрятать	припрятывать
плевать	наплевать	оплевать	оплевывать
растить	вырастить	отрастить	отращивать
жрать	сожрать	пожрать	пожирать
нюхать	понюхать	вынюхать	вынюхивать
щупать	пощупать	нащупать	нащупывать
играть	сыграть	выиграть	выигрывать
верить	поверить	проверить	проверять
звонить	позвонить	перезвонить	перезванивать
рисовать	нарисовать	срисовать	срисовывать
влечь	повлечь	привлечь	привлекать
жечь	сжечь	зажечь	зажигать
мерзнуть	замерзнуть	промерзнуть	промерзать
копать	выкопать	раскопать	раскапывать
красть	украсть	обокрасть	обкрадывать

Таблица 3. **Средний коэффициент корреляции Пирсона частот глаголов
в аспектуальных парах из таблицы 2 /**
Table 3. **Average Pearson correlation coefficient of verb frequencies in aspectual pairs from table 2**

Частота, log10	Префиксальное видообразование	Суффиксальное видообразование
-5	0,94	-
-6	0,75	0,72
-7	0,71	0,72
-8	-	0,53

Здесь минус означает отсутствие пар в данной частотной группе. Как и ожидалось, для эталонных пар коэффициент корреляции заметно выше, чем

в табл. 1. Аспектуальные пары обоих типов трансформации присутствуют только в двух частотных группах: $10^{-5} > f > 10^{-6}$ и $10^{-6} > f > 10^{-7}$. В обоих группах коэффициенты корреляции почти равны.

Таким образом, при таком подходе различия в степени семантической близости при суффиксальном и префиксальном способах образования вида для эталонных пар также не обнаружены.

3.2. Подход 2: семантическое сходство в векторной семантике

Хорошо известной является дистрибутивная гипотеза, согласно которой лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения. Основываясь на ней в компьютерной лингвистике разработаны методы векторной семантики, позволяющие формализовать семантику слова как вектор частот слов, с которыми оно вместе встречается в текстах корпуса. Семантическое сходство между словами легко вычислить, одним из стандартных математических методов, например, как косинусную меру сходства между векторами слов. Предложены различные методы построения векторов: skipgram, fasttext и др. (Mikolov et al. 2013, Bojanowski et al. 2017, Pennington et al. 2014). Данный подход и его последующие модификации широко используются в различных задачах обработки текстов, таких как классификация текстов, вопросно-ответные системы, суммаризация текстов (Sivakumar et al. 2020, Yang et al. 2019). Для русского языка существуют системы, в которых уже построены вектора для сотен тысяч слов (Grave et al. 2018, Kutuzov & Andreev 2015). Мы используем эти методы для расчета семантического сходства слов в аспектуальных парах.

В таблице 4 приведена величина сходства (в интервале от 0 до 1; 1 – максимальное сходство) для нашего набора эталонных пар. Для расчетов применялся инструментарий Rusvectores (ENA, September 20, 2022)[6], использующий модель skipgram, обученную на корпусе Тайга (Shavrina & Shapovalova 2017). В нее включены также значения сходства между базовым имперфективом и специализированным перфективом. Сходство приведено в столбцах, расположенных между соответствующими словами.

В приложении приведены две таблицы, в которых использованы другие математические модели и другие корпуса. Численные данные в них немного отличаются от данных таблицы 4, но демонстрируют те же эффекты. В таблице S1 сходство рассчитывалось по предобученной модели fasttext без дополнительного дообучения, а в таблице S2 по skipgram, обученной на корпусе НКРЯ и Википедия. Различие в корпусах состоит в том, что Тайга состоит почти полностью (95%) из художественных текстов. Прочерк в таблицах ниже означает, что в данной вычислительной модели отсутствуют необходимые слова.

---

[6] https://rusvectores.org/ru/misc/

Таблица 4. **Величина сходства между словами в векторной семантике,**
модель skipgram, корпус Тайга /
Table 4. **The value of the similarity between words in vector semantics,**
skipgram model, Taiga corpus

№	Естественный перфектив	Сходст-во	Базовый импер-фектив	Сходст-во	Специализи-рованный перфектив	Сходст-во	Вторичный имперфектив
1	сделать	0,73	делать	0,20	переделать	0,73	переделывать
2	показаться	0,64	казаться	0,21	оказаться	0,69	оказываться
3	попросить	0,82	просить	0,45	выпросить	0,60	выпрашивать
4	поставить	0,81	ставить	0,30	приставить	0,78	приставлять
5	сохранить	0,52	хранить	0,38	охранить	0,50	охранять
6	спрятать	0,81	прятать	0,68	припрятать	0,82	припрятывать
7	наплевать	0,84	плевать	0,41	оплевать	0,50	оплевывать
8	вырастить	0,73	растить	0,27	отрастить	0,69	отращивать
9	сожрать	0,62	жрать	0,60	пожрать	0,48	пожирать
10	понюхать	0,76	нюхать	0,34	вынюхать	0,59	вынюхивать
11	пощупать	0,70	щупать	0,46	нащупать	0,48	нащупывать
12	сыграть	0,65	играть	0,27	выиграть	0,81	выигрывать
13	поверить	0,76	верить	0,19	проверить	0,81	проверять
14	позвонить	0,77	звонить	0,71	перезвонить	0,75	перезванивать
15	нарисовать	0,80	рисовать	0,42	срисовать	0,55	срисовывать
16	повлечь	0,63	влечь	0,39	привлечь	0,76	привлекать
17	сжечь	0,58	жечь	0,38	зажечь	0,82	зажигать
18	замерзнуть	0,63	мерзнуть	0,52	промерзнуть	0,74	промерзать
19	выкопать	0,70	копать	0,50	раскопать	0,71	раскапывать
20	украсть	0,50	красть	0,46	обокрасть	0,67	обкрадывать
Среднее		0,70		0,41		0,67	

4. Обсуждение результатов

Анализируя данные таблиц, следует отметить следующее.

1. Сравнение сходства в парах «базовый имперфектив – естественный перфектив» с парами «базовый имперфектив – специализированный перфектив».

Как и следовало ожидать, сходство между словами в парах «базовый имперфектив – естественный перфектив» (столбец 3) значительно выше по сравнению с парами «базовый имперфектив – специализированный перфектив» (столбец 5). Различие в семантике слов в парах «базовый имперфектив – естественный перфектив» обусловлено только различием в виде, а в парах «базовый имперфектив – специализированный перфектив» кроме вида семантика глаголов различается еще и семантикой приставки. Это подтверждает правильность предложенного метода вычисления семантической схожести в данной задаче. Отметим, что не только в среднем, но и в каждой строке значение в третьем столбце выше, чем в пятом (исключение в строке 9 табл. S2 приложения со словом *жрать*, где имеет место равенство значений).

2. Сравнение сходства в парах «базовый имперфектив – специализированный перфектив» с парами «специализированный перфектив – вторичный имперфектив».

Аналогичное соотношение мы видим и при сопоставлении пар «базовый имперфектив – специализированный перфектив» с парами «специализированный перфектив – вторичный имперфектив». Степень сходства в парах "специализированный перфектив – вторичный имперфектив" (столбец 7) почти во всех случаях больше, чем в парах «базовый имперфектив – специализированный перфектив» (столбец 5). Обратим внимание на интересное исключение.

В табл. 4 и S1 в паре *пожрать – пожирать* степень близости слов меньше, чем в паре *жрать – пожрать*. Это указание на то, что *пожрать – пожирать* не является чистой аспектуальной парой, *пожирать* имеет какую-то дополнительную семантику по сравнению с *пожрать*. Проведем лингвистический анализ этих слов по их толкованиям, чтобы выявить возможные расхождения в семантике. Возьмем толкования в словаре Малом академическом словаре (Евгеньева 1981–84). *Пожрать* означает то же, что и *поесть* с дополнительным значением 'с жадностью' (о животных) или грубое протонародное выражение, если говорится о человеке. Слово *пожирать* кроме этого имеет еще и переносные значения: 1) Быстро, с жадностью и интересом прочитывать (книги, журналы и т. д.). Сочетание ^{??} *пожрал книги* маргинально. В Google это словосочетание встречается 27 раз, в то время как *пожирал книги* – 6600 раз. 2) Поглощать, требовать для себя что-л. в больших количествах. 3) Уничтожать, сжигая (об огне, пламени). 4) Захватывать целиком (о чувствах). Таким образом, между словами *пожрать* и *пожирать* действительно есть значительные семантические различия, кроме видовых.

То, что суффиксальное образование вторичного имперфектива не всегда является чисто видовой трансформацией ранее отмечалось, например, в работе (Апресян 1995: 112), в которой указано, что глагол *выпивать* имеет дополнительное значение кратности, а чисто видовой парой с *выпить* является именно *пить*.

3. Сравнение средних значений сходства. Средние значения близости в парах «базовый имперфектив – естественный перфектив» и «специализированный перфектив – вторичный имперфектив» очень близки и существенно превышают среднее значение в парах «базовый имперфектив – специализированный перфектив».

При расчетах по модели skipgram с корпусом Тайга (табл. 4) в 10 случаях значение близости выше для пар «базовый имперфектив – естественный перфектив», в 9 случаях – для пар «специализированный перфектив – вторичный имперфектив» (в одном случае равенство). При расчетах по модели fasttext (табл. S1) в 8 случаях значение близости выше для пар «базовый имперфектив – естественный перфектив», в 10 случаях – для пар «специализированный перфектив – вторичный имперфектив» (в двух случаях равенство).

При расчетах по модели skipgram с корпусами НКРЯ и Википедия (табл. S2) в 10 случаях значение близости выше для пар «базовый имперфектив – естественный перфектив», в 8 случаях – для пар «специализированный перфектив – вторичный имперфектив».

Таким образом, хотя точные значения величины близости для пар слов несколько различаются для разных вычислительных моделей, общие закономерности сохраняются. Оба способа образования аспектуальных пар – префиксальный и суффиксальный – в одинаковой степени меняют семантику слов. В случаях значительного расхождения значений близости пар слов анализ значений слов по словарям позволяет объяснить причины расхождения. Например, в паре *зажечь – зажигать* семантическое сходство значительно выше, чем в паре *жечь – сжечь*. Семантик слов *зажечь, зажигать* практически совпадает, в то же время между словами *жечь, сжечь* имеется заметная разница: *жечь* имеет значение 'вызывать ожег', которого нет у глагола *сжечь*.

Полученные численные данные могут быть использованы для получения и других результатов. Рассчитаем коэффициенты корреляции по Пирсону между столбцами 3, 5, 7; результаты приведены в табл. 5.

Таблица 5. **Корреляция между разными способами образования пар /**
Table 5. **Correlation between different ways of pairing**

	Модель fastText	Модель skipgram+Тайга	Модель skipgram+НКРЯ+Wiki
Префиксальная аспектуальная пара – суффиксальная аспектуальная пара	0,173	0,018	0,332
Префиксальная аспектуальная пара – способ глагольного действия	0,065	0,097	0,040
Суффиксальная аспектуальная пара – способ глагольного действия	−0,441	−0,150	−0,371

Между префиксальным и суффиксальным способами образования аспектуальных пар корреляция близости пар глаголов слабая (по Чеддоку). Это же касается и сравнения: префиксальные аспектуальные пары – способ глагольного действия. В то же время при сопоставлении данных по суффиксальному способу образования аспектуальных пар с парами, классифицируемыми как способ глагольного действия, неожиданно обнаруживается значительно более высокий уровень корреляции, причем с отрицательным знаком. Для модели fastText эта корреляция характеризуется как умеренная, близкая к средней: −0,441. Качественно это означает, что чем меньший вклад вносит приставка в семантику специализированного перфектива по сравнению с базовым имперфективом, тем в большей степени вторичный имперфектив будет семантически отличаться от специализированного перфектива.

Например, в таблице 4 средняя степень близости в парах «базовый имперфектив – специализированный перфектив» равно 0,41, а в парах «специализированный перфектив – вторичный имперфектив» – 0,67. При этом у

рассмотренного выше глагола *жрать* в паре *жрать – пожрать* степень близости 0,60, заметно выше средней, что указывает на высокую степень семантической близости, хотя *пожрать* и специализированный перфектив, а в паре *пожрать – пожирать* только 0,48 – значительно ниже среднего, что указывает на меньшую семантическую близость, обусловленную дополнительным значением слова *пожирать,* как было указано выше. В то же время у пары *делать – переделать* степень близости всего 0,20 (меньше средней), обусловленная ясным различием в семантике, а у пары *переделать – переделывать* степень близости 0,73 (больше средней). Эта пара является чисто аспектуальной (Ушаков 1994). Насколько нам известно, такой эффект ранее не описывался. Он требует дополнительного изучения.

5. Заключение

С появлением больших корпусов текстов появилась надежда на то, что применение статистических, компьютерных методов позволит прояснить вопрос о том, находятся ли глаголы в аспектуальных парах в словоизменительных или иных грамматических отношениях. Для постулирования словоизменительного или словообразовательного (+словоклассифицирующего) статуса аспектуальных пар, получаемых суффиксальным или префиксальным способом, принципиальным является определение того, в какой степени глаголы в паре являются семантически близкими. В значительном числе работ предполагалось, что суффиксальный способ образования аспектуальных пары является словоизменительным, а префиксальный – словообразовательным (+словоклассифицирующим). Соответственно, можно было бы ожидать, что суффиксальный способ не меняет семантики глаголов, он меняет только вид, а префиксы вносят определенный вклад в глагольную семантику, помимо видовой.

В статье проверялась эта гипотеза. Предложены два различных подхода для получения точных количественных оценок семантической близости глаголов в аспектуальных парах. Оба подхода реализованы в двух вариантах. При этом нигде не обнаружено заметной разницы в степени семантической близости глаголов в префиксальных и суффиксальных аспектуальных парах. Таким образом, ответ на сформулированную в начале статьи гипотезу: «семантическая близость между глаголами в аспектуальных парах перфектив – вторичный имперфектив будет больше, чем между глаголами в парах базовый имперфектив – перфектив» отрицательный.

Аналогичный результат при применении совершенно иного подхода – а именно сопоставлении грамматических профилей глаголов – был получен и в работе (Janda & Lyashevskaya 2011). Найти количественные различия в свойствах суффиксальных или префиксальных аспектуальных пар не удается.

Это означает, что основной признак отнесения грамматической категории к словоизменению или словообразованию – сохранение или не сохранение лексической семантики – не позволяет разграничить префиксальный и

суффиксальный способы видообразования с точки зрения их грамматического статуса. Многолетнюю дискуссию о противопоставлении грамматического статуса аспектуальных пар, образованных префиксальным и суффиксальным способами, по нашему мнению, можно считать практически завершенной.

Обратим внимание на то, что второй из предложенных подходов – с вычислением векторной близости между словами на основе больших корпусов текстов – широко используется в компьютерной лингвистике (Solovyev et al. 2022), но лишь в последнее время начал применяться для решения собственно лингвистических задач (Ryzhova & Paperno 2022, Shavrina 2021). Между тем он может оказаться перспективным в исследованиях, связанных с семантической близостью слов.

Вклад авторов

Соловьев В.Д. – основная идея, выполнение расчетов (табл. 4, 5, S2), квантитативный анализ результатов, написание текста статьи. *Бочкарев В.В.* – выполнение расчетов (табл. 1, 3). *Байрашева В.Р.* – лингвистический анализ примеров. *Янда Л.* – подготовка набора слов (табл. 2), *Исламов М.И.* – выполнение расчетов (табл. S1).

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

Апресян Ю.Д. Избранные труды. Т. 2. М.: Школа "Языки русской культуры", 1995. [Apresjan, Jury D. 1995. Izbrannye Trudy (Elected Works). V. 2. M.: Shkola "Yazyki russkoi kul'tury"].

Бондарко А.В. Вид и время русского глагола. М.: Просвещение 1971. [Bondarko, Alexander V. 1971. Vid i vremya russkogo glagola (Type and tense of the Russian verb). M.: Prosveshchenie (In Russ.)].

Виноградов В.В. Русский язык. М.: Высш. шк., 1972. [Vinogradov, Victor V. 1972. Russkii yazyk (Russian language). M.: Vyssh. shk. (In Russ.)].

Вежбицкая А. Семантические универсалии и описание языков. М.: Языки русской культуры, 1999. [Wierzbicka, Anna. 1999. Semanticheskie universalii i opisanie yazykov (Semantic universals and the description of languages). M.: Yazyki russkoi kul'tury. (In Russ.)].

Всеволодова М.В. Семный состав глагольного слова. (К вопросу типологии вида) // *Типология вида: проблемы, поиски, решения.* Тезисы международной научной конференции. Московский государственный университет имени М.В. Ломоносова. Филологический факультет. М.: Изд-во Моск. ун-та, 1997. [Vsevolodova, Maiya V. 1997. Semnyi sostav glagol'nogo slova. (K voprosu tipologii vida). (The semantic composition of the verb word. (On the question of the typology of species)) *Tipologiya vida: problemy, poiski, resheniya.* Tezisy mezhdunarodnoi nauchnoi konferencii. Moscow State University named after M. V. Lomonosov. Philological department. M.: Izd-vo Mosk. un-ta. (In Russ.)].

Горбова Е.В. Заметки о видообразовании русского глагола и словоизменительной vs. словоклассифицирующей трактовке вида // *Acta Linguistica Petropolitana*. Вып. X. Ч. 3. СПб.: Наука, 2014. С. 181–211. [Gorbova, Elena V. 2014. Notes on aspectual formation of the Russian verb and on inflectional vs. word-classifying interpretation of aspect. *Acta Linguistica Petropolitana* 10 (3). 181–211. St. Petersburg: Nauka. (In Russ.)].

Горбова Е.В. Видообразование русского глагола: префиксация и/или суффиксация? // *Вопросы языкознания*. 2015a. № 1. С. 7–37. [Gorbova, Elena V. 2015a. Aspectual formation of the Russian verb: Prefication or suffixation? *Voprosy Yazykoznaniya* 1. 7–37. (In Russ.)].

Горбова Е.В. Словоизменение или словообразование, или является ли русский вид «видом славянского типа»? // *Аспектуальная семантическая зона: типология систем и сценарии диахронического развития*. Сборник статей V Международной конференции Комиссии по аспектологии Международного комитета славистов. Киото, 13–15 ноября 2015 / Китадзё М. (сост.). Киото: Tanaka Print. С. 62–72. 2015б. [Gorbova, Elena V. 2015b. Inflection or derivation, or Is the Russian aspect «Slavicstyle aspect»? *Aspektual'naya semanticheskaya zona: tipologiya sistem i stsenarii diakhronicheskogo razvitiya*. Sbornik statei V Mezhdunarodnoi konferentsii Komissii po aspektologii Mezhdunarodnogo komiteta slavistov. Kioto, 13–15 noyabrya 2015. Kitajo M. (comp.). Kyoto: Tanaka Print. 62–72. (In Russ.)].

Горбова Е.В. Русское видообразование: словоизменение, (слово)классификация или набор квазиграммем? (еще раз о болевых точках русской аспектологии) // *Вопросы языкознания*. 2017. № 1. С. 24–52. [Gorbova, Elena V. 2017. Aspectual formation of Russian verbs: Inflection, derivation, or a set of quasigrammemes? («sore points» of Russian aspectology revisited). *Voprosy Yazykoznaniya* 1. 24–52. (In Russ.)].

Евгеньева А.П. (ред.) Словарь русского языка: в 4 т. / АН СССР, Ин-т рус. яз. М.: Русский язык, 1981–1984. [Evgenyeva, Anastasia P. (ed.). 1981–1984. Slovar' russkogo yazyka: V 4-h t. (The dictionary of the Russian language in 4 volumes) / AN SSSR, In-t rus. yaz. M.: Russkii yazyk. (In Russ.)].

Зализняк А.А., Микаэлян И.Л. О некоторых дискуссионных моментах аспектологической концепции Лоры Янды // *Вопросы языкознания*. 2012. №6. С. 48–65. [Zalizniak, Anna A. & Irina L. Mikaelian. 2012. O nekotoryh diskussionnyh momentah aspektologicheskoii koncepcii Lory Jandy (On some discussion points in Laura Yanda's aspectological concept). *Voprosy yazykoznaniya* 6. 48–65. (In Russ.)].

Зализняк А.А., Шмелев А.Д., Микаэлян И.Л. Русская аспектология. В защиту видовой пары. М.: Языки славянских культур, 2015. [Zalizniak, Anna A., Aleksej D. Shmelev & Irina L. Mikaelian. 2015. Russkaya aspektologiya. V zashhitu vidovoi pary (Russian Aspectology. In defense of the species pair.). M.: Yazyki slavyanskih kul'tur. (In Russ.)].

Иорданская Л.Н., Крылосова С.Г., Мельчук И.А., Михель П.А. Категория направленности у русских парных глаголов перемещения (на примере глаголов ЛЕТЕТЬ / ЛЕТАТЬ) // *Вопросы языкознания*. 2020. № 1. 27–64. [Iordanskaja, Lidija N., Svetlana G. Krylosova, Igor A. Melcuk & Polina A. Mihel'. 2020. Kategoriya napravlennosti u russkih parnyh glagolov peremeshcheniya (na primere glagolov LETET' / LETAT') (Directionality in Russian paired verbs of displacement (by the example of the verbs FLY / FLY)). *Voprosy yazykoznaniya* 1. 27–64. (In Russ.)].

Исаченко А.В. 1954/2003. *Грамматический строй русского языка в сопоставлении со словацким. Морфология.* Т. I–II. М.: Языки славянской культуры, 2003. (Исаченко А.В. Грамматический строй русского языка в сопоставлении со словацким: Морфология. Часть первая. Братислава, 1954.) [Issatschenko, Aleksandr V. 1954/2003. *Grammaticheskii stroi russkogo yazyka v sopostavlenii so slovatskim. Morfologiya* (Grammatical system of Russian versus Slovak. Morphology). Vol. I–II. Moscow: Yayki

Slavyanskoi Kul'tury (Issatschenko, Aleksandr V. 1954. Grammaticheskii stroi russkogo yazyka v sopostavlenii so slovatskim: Morfologiya (Grammatical system of Russian versus Slovak. Morphology). Part 1. Bratislava. (In Russ.)].

Кузнецова Ю.Л., Янда Л.А. Приставки в свете когнитивной лингвистики и типологии (отклик на статью А.А. Зализняк и И.Л. Микаэлян) // *Вопросы языкознания.* 2013. №4. С. 87–96. [Kuznecova, Julija L. & Laura A. Janda. 2013. Pristavki v svete kognitivnoi lingvistiki i tipologii (otklik na stat'yu Anna A. Zalizniak i Irina L. Mikaelian) (Prefixes in the light of cognitive linguistics and typology (response to the article of Anna A. Zaliznyak & Irina L. Mikaelian). *Voprosy yazykoznaniya* 4. 87–96. (In Russ.)].

Маслов Ю.С. 1978/2004. Избранные труды. Аспектология. Общее языкознание. М.: Языки славянских культур, 2004. С. 305–364. (Маслов Ю.С. К основаниям сопоставительной аспектологии // Маслов Ю.С. (отв. ред.) Проблемы современного теоретического и синхронно-сопоставительного языкознания. Вып. 1: Вопросы сопоставительной аспектологии. Л.: Изд-во ЛГУ, 1978. С. 4–44.) [Maslov, Yury S. 1978/2004. Izbrannye trudy. Aspektologiya. Obshchee yazykoznanie (Selected works. Aspectology. General linguistics). Moscow: Yazyki Slavyanskikh Kul'tur. 305–364. (Maslov, Yury S. 1978. Towards the fundamentals of comparative aspectology. In Yury Maslov (ed.), *Problems of Modern Theoretical and Synchronic Comparative Linguistics. Vol. 1: Issues of Comparative Aspectology*, 4–44. Leningrad: Leningrad State Univ. Publ. (In Russ.)].

Маслов Ю.С. Очерки по аспектологии. Л.: ЛГУ, 1984. [Maslov, Yury S. *Ocherki po aspektologii* (Essays on Aspectology). L.: LGU. (In Russ.)].

Мельчук И.А. 1997. Курс общей морфологии. Т. 1. М.: Прогресс, Языки русской культуры, Вена: WSA. [Mel'chuk, Igor A. 1997. Kurs obshhei morfologii (Course in General Morphology). V. 1. M.: Progress, Yazyki russkoi kul'tury, Vena: WSA. (In Russ.)].

Мельчук И.А. 1998. Курс общей морфологии. Т. 2. М.: Языки русской культуры, Вена: WSA. [Mel'chuk, Igor A. 1998. Kurs obshhei morfologii (Course in General Morphology). V. 2. M.: Yazyki russkoi kul'tury, Vena: WSA. (In Russ.)].

Перцов Н.В. 1996. Грамматическое и обязательное в языке // *Вопросы языкознания.* № 4. С. 39–61. [Percov, Nikolay V. 1996. Grammaticheskoe i obyazatel'noe v yazyke (Grammatical and obligatory in language). *Voprosy yazykoznaniya* 4. 39–61. (In Russ.)].

Петрухина Е.В. Аспектуальные категории глагола в русском языке в сопоставлении с чешским, словацким, польским и болгарским языками. М.: МГУ, 2000. [Petruhina, Elena V. 2000. Aspektual'nye kategorii glagola v russkom yazyke v sopostavlenii s cheshskim, slovatskim, pol'skim i bolgarskim yazykami (Aspectual Categories of the Verb in Russian Compared to Czech, Slovak, Polish, and Bulgarian). M.: MGU. (In Russ.)].

Плунгян В.А. Общая морфология. Введение в проблематику. М.: УРСС, 2000. [Plungian, Vladimir A. 2000. Obshchaya morfologiya. Vvedenie v problematiku (General Morphology. Introduction to the Problematics). Mju: URSS. (In Russ.)].

Русская грамматика: научные труды. В 2 томах / Н.Ю. Шведова (Гл. ред.). — М.: Институт русского языка им. В.В. Виноградова РАН, 2005. [Russkaya grammatika: nauchnye trudy. V 2 tomakh (Russian grammar: Scientific works. In 2 volumes) / Natalia Shvedova (ed.). 2005. M.: Institut russkogo yazyka im. V.V. Vinogradova RAN. (In Russ.)].

Сичинава Д.В. Вид. Материалы для проекта корпусного описания русской грамматики (http://rusgram.ru). На правах рукописи. М., 2011. [Sichinava, Dmitrii V. 2011. Vid. Materialy dlya proekta korpusnogo opisaniya russkoi grammatiki (http://rusgram.ru). Na pravah rukopisi (Materials for the project of corpus description of Russian grammar (http://rusgram.ru). On the rights of manuscript). M. (In Russ.)]. (accesed 13 October 2012).

Соловьев В., Бочкарев В., Янда Л. Аспектуальная система в русском языке: семантический вклад приставок и суффиксов // *15th Slavic Cognitive Linguistics Conference*, Book of Abstracts. С. 75–76. 2017. [Solovyev, Valery, Vladimir Bochkarev & Laura Janda. 2017. Aspektual'naya sistema v russkom yazyke: semanticheskii vklad pristavok i suffiksov (The aspect system in Russian: Semantic contribution of prefixes and suffixes). *15th Slavic Cognitive Linguistics Conference*, Book of Abstracts. 75–76. (In Russ.)].

Соловьев В., Бочкарев В. Дело об аспектуальных парах открывается вновь // *Вестник Томского государственного университета.* Филология. Вып. 78. 2022. [Solovyev, Valery & Vladimir Bochkarev. 2022. Case for aspectual pairs reopened. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya* 78. (In Russ.)].

Тихонов А.Н. Русский глагол: проблемы теории и лексикографирования. М.: Academia, 1998. [Tikhonov, Alexander N. 1998. *Russkii glagol: problemy teorii i leksikografirovaniya* (The Russian verb: Problems of theory and lexicographic description). Moscow: Academia. (In Russ.)].

Ушаков Д.Н. (ред.) Tolkovyi slovar' russkogo yazyka: V 4 t. М.: Сов. Энцикл, 1994. [Ushakov, Dmitry N. (ed.). 1994. Tolkovyi slovar' russkogo yazyka: V 4 t. (The Dictionary of the Russian language: In 4 vols.). M.: Sov. encik. (In Russ.)].

Храковский В.С. Категория вида в русском языке: болевые точки // *Аспектуальная семантическая зона: типология систем и сценарии диахронического развития*. Сборник статей V Международной конференции Комиссии по аспектологии Международного комитета славистов. Киото, 13–15 ноября 2015 г. / Китадзё М. (сост.). Киото: Tanaka Print. С. 321–334. 2015. [Xrakovskij, Viktor S. 2015. The category of aspect in Russian: Problematic issues. *Aspektual'naya semanticheskaya zona: tipologiya sistem i stsenarii diakhronicheskogo razvitiya.* Sbornik statei V Mezhdunarodnoi konferentsii Komissii po aspektologii Mezhdunarodnogo komiteta slavistov. Kioto, 13–15 November 2015 g. Kyoto: Tanaka Print. 321–334. (In Russ.)].

Храковский В.С. И опять о болевых точках категории вида // *Вопросы языкознания*. 2018. № 1. С. 105–118. [Xrakovskij, Viktor S. 2018. Aspect's "sore points" revisited once more. *Voprosy yazykoznaniya* 1. 105–118. (In Russ.)].

Янда Л.А. Русские приставки как система глагольных классификаторов // *Вопросы языкознания*. 2012. № 6. С. 3–47. [Janda, Laura A. 2012. Russkie pristavki kak sistema glagol'nyh klassifikatorov (Russian prefixes as a system of verb classifiers). *Voprosy yazykoznaniya* 6. 3–47. (In Russ.)].

Batyrshin, Ildar, Valery Solovyev & Vladimir Ivanov. 2014. Time series shape association measures and local trend association patterns. *Neurocomputing* 175. 924–934

Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5. 135–146.

Dressler, Wolfgang. 1989. Prototypical differences between inflection and derivation. *STUF – Language Typology and Universals* 42 (1). 3–10.

Gorbova, Elena V. & Oksana Iu. Chuikova. 2022. Suffixal imperfectivation of prefixed verbs: Record breakers and outsiders (based on the dictionary, corpus and Runet). *Computational Linguistics and Intellectual Technologies.* Papers from the Annual International Conference "Dialogue". 21.

Grave, Edouard, Piotr Bojanowski, Prakhar Gupta, Armand Joulin & Tomas Mikolov. 2018. *Learning Word Vectors for 157 Languages. Proceedings of the International Conference on Language Resources and Evaluation* (*LREC 2018*). https://www.aclweb.org/anthology/L18-1550 (accessed 21 September 2022)

Janda, Laura & Olga Lyashevskaya. 2011. Aspectual Pairs in the Russian National Corpus. *Scando-Slavica* 57 (2). 201–215.

Janda, Laura, Anna Endresen, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nesset & Svetlana Sokolova. 2013. *Why Russian Aspectual Prefixes aren't Empty.* Bloomington: Slavica.

Filip, Hana. 1999. *Aspect, Eventuality Types, and Nominal Reference.* London.

Kutuzov, Andrey & Igor Andreev. 2015. Texts in, meaning out: Neural language models in semantic similarity task for Russian. *Proceedings of the Dialog 2015 Conference*, Moscow, Russia. 133–144.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Aiden & Adrian Veres. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331 (6014). 176–182.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

Pennington, Jeffrey, Richard Socher & Christopher D. Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing* (*EMNLP*). 1532–1543.

Plank, Frans. 1991. Inflection and derivation. *EUROTYP working papers* 10.

Richey, Sean & Benjamin J. Taylor. 2020. Google Books Ngrams and political science: Two validity tests for a novel data source *PS: Political Science & Politics* 53. 72–77. https://doi.org/10.1017/S1049096519001318.

Ryzhova, Daria & Denis Paperno. 2022. Constructing typological questionnaire with distributional semantic models. In Ekaterina Rakhilina, Tatiana Reznikova & Daria Ryzhova (eds.), *The Typology of Physical Qualities*. Amsterdam: John Benjamins Publishing Company.

Sahlgren, Magnus. 2008. The Distributional Hypothesis. From context to meaning. Distributional models of the lexicon in linguistics and cognitive science. [Special issue]. *Rivista di Linguistica* 20 (1). 33–53.

van Schooneveld, Cornelis. 1978. *Semantic Transmutations: Prolegomena to a Calculus of Meaning. V. 1: The Cardinal Semantic Structure of Prepositions, Cases, and Paratactic Conjunctions in Contemporary Standard Russian*. Bloomington,

Shavrina, Tatiana & Olga Shapovalova. 2017. To the methodology of corpus construction for machine learning: "Taiga" syntax tree corpus and parser. *Proceedings of the International Conference «Corpus Linguistics–2017»*. St. Petersburg. 78–84.

Shavrina, Tatiana. 2021. Methods of computational linguistics in the evaluation of artificial intelligence systems. *Voprosy yazykoznaniya* 6. 117–138.

Singh, Rajendra & Alan Ford. 1980. Flexion, dérivation et Panini. *Amsterdam Studies in the Theory and History of Linguistic Science Amsterdam* 20. 323–332.

Sivakumar, Soubraylu, Lakshmi Sarvani Videla, T. Rajesh Kumar, J. Nagaraj, Shilpa Itnal & D. Haritha. 2020. Review on Word2Vec Word Embedding Neural Net. *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE. 282–290.

Solovyev, Vavery, Bochkarev Vladimir & Akhtyamova Svetlana. 2019. Google Books Ngram: Problems of representativeness and data reliability. *Communications in Computer and Information Science* 1223. Springer, Cham. 147–162. https://doi.org/10.1007/978-3-030-51913-1_10

Solovyev, Valery, Marina Solnyshkina & Danielle McNamara. 2022. Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics* 26 (2). 275–316. https://doi.org/10.22363/2687-0088-30161

Zhilin, Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov & Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237.

Интернет-ресурсы

Database Exploring Emptiness, http://emptyprefixes.uit.no/ (accessed 05 September 2022).
База данных суффиксальных аспектуальных пар [Baza dannyh suffiksal'nyh aspektual'nyh par], http://ru-eval.ru/go/resources.html (accessed 12 February 2020).

ПРИЛОЖЕНИЕ

Таблица S1. **Величина сходства между словами в векторной семантике, модель fasttext /**
Table S1. **The value of the similarity between words in vector semantics, fasttext model**

№	Естествен-ный перфектив	Рассто-яние	Базовый имперфектив	Рассто-яние	Специализи-рованный перфектив	Рассто-яние	Вторичный имперфектив
1	сделать	0,76	делать	0,51	переделать	0,76	переделывать
2	показаться	0,74	казаться	0,63	оказаться	0,63	оказываться
3	попросить	0,82	просить	0,59	выпросить	0,65	выпрашивать
4	поставить	0,81	ставить	0,41	приставить	0,82	приставлять
5	сохранить	0,59	хранить	0,39	охранить	0,64	охранять
6	спрятать	0,80	прятать	0,65	припрятать	0,64	припрятывать
7	наплевать	0,89	плевать	0,49	оплевать	0,71	оплевывать
8	вырастить	0,74	растить	0,45	отрастить	0,78	отращивать
9	сожрать	0,63	жрать	0,61	пожрать	0,47	пожирать
10	понюхать	0,71	нюхать	0,51	вынюхать	0,71	вынюхивать
11	пощупать	0,67	щупать	0,45	нащупать	0,76	нащупывать
12	сыграть	0,76	играть	0,48	выиграть	0,75	выигрывать
13	поверить	0,72	верить	0,40	проверить	0,75	проверять
14	позвонить	0,76	звонить	0,54	перезвонить	0,63	перезванивать
15	нарисовать	0,75	рисовать	0,55	срисовать	0,80	срисовывать
16	повлечь	0,64	влечь	0,51	привлечь	0,79	привлекать
17	сжечь	0,66	жечь	0,59	зажечь	0,77	зажигать
18	замерзнуть	0,75	мерзнуть	0,60	промерзнуть	0,76	промерзать
19	выкопать	0,61	копать	0,58	раскопать	0,71	раскапывать
20	украсть	0,69	красть	0,50	обокрасть	0,53	обкрадывать
Сред-нее		0,73		0,52		0,70	

Таблица S2. **Величина сходства между словами в векторной семантике,**
модель *skipgram*, корпус НКРЯ и Википедия /
Table S2. **The value of the similarity between words in vector semantics,**
skipgram model, RNC corpus and Wikipedia

№	Естествен-ный перфектив	Рассто-яние	Базовый имперфек-тив	Рассто-яние	Специализи-рованный перфектив	Рассто-яние	Вторичный имперфектив
1	сделать	0,77	делать	0,19	переделать	0,73	переделывать
2	показаться	0,69	казаться	0,38	оказаться	0,76	оказываться
3	попросить	0,79	просить	0,53	выпросить	0,62	выпрашивать
4	поставить	0,86	ставить	0,39	приставить	0,69	приставлять
5	сохранить	0,53	хранить	0,39	охранить	0,45	охранять
6	спрятать	0,80	прятать	0,72	припрятать	0,67	припрятывать
7	наплевать	0,65	плевать	0,51	оплевать	0,56	оплевывать

Окончание табл. S2

№	Естествен-ный перфектив	Рассто-яние	Базовый имперфек-тив	Рассто-яние	Специализи-рованный перфектив	Рассто-яние	Вторичный имперфектив
8	вырастить	0,61	растить	0,27	отрастить	0,58	отращивать
9	сожрать	0,51	жрать	0,51	пожрать	0,54	пожирать
10	понюхать	0,69	нюхать	-	вынюхать	-	вынюхивать
11	пощупать	0,65	щупать	0,51	нащупать	0,70	нащупывать
12	сыграть	0,70	играть	0,27	выиграть	0,82	выигрывать
13	поверить	0,78	верить	0,33	проверить	0,82	проверять
14	позвонить	0,75	звонить	0,63	перезвонить	0,60	перезванивать
15	нарисовать	0,80	рисовать	0,53	срисовать	0,52	срисовывать
16	повлечь	0,69	влечь	0,34	привлечь	0,73	привлекать
17	сжечь	0,56	жечь	0,49	зажечь	0,76	зажигать
18	замерзнуть	0,57	мерзнуть	0,51	промерзнуть	0,62	промерзать
19	выкопать	0,68	копать	0,45	раскопать	0,63	раскапывать
20	украсть	0,44	красть	0,39	обокрасть	-	обкрадывать
Среднее		0,69		0,42		0,66	

Bionotes:
Valery SOLOVYEV is Doctor Habil. of Physics and Mathematics, Professor, Chief Researcher of "Text Analytics" Research Lab at the Institute of Philology and Intercultural Communication of Kazan Federal University, Kazan, Russia. He is a member of the Presidium of the Interregional Association for Cognitive Research, author of four monographs and over 60 publications on text complexity and computational linguistics.
e-mail: maki.solovyev@mail.ru
https://orcid.org/0000-0003-4692-2564

Vladimir BOCHKAREV is Research Fellow of the scientific laboratory "Linguistics and artificial intelligence" at Kazan Federal University, Kazan, Russia. His area of interest is quantitative linguistics.
e-mail: vbochkarev@mail.ru
https://orcid.org/0000-0001-8792-1491

Venera BAYRASHEVA holds a PhD in Physics and Mathematics. She is Associate Professor at the Department of Theoretical Cybernetics at Kazan Federal University, Kazan, Russia. Her areas of interest are computational and theoretical linguistics.
e-mail: vbayrasheva@gmail.com
https://orcid.org/0000-0002-1729-034X

Сведения об авторах:
Валерий Дмитриевич СОЛОВЬЕВ – доктор физико-математических наук, профес-сор, главный научный сотрудник НИЛ «Текстовая аналитика» Института филологии и межкультурной коммуникации Казанского федерального университета, Казань,

Россия. Он является членом президиума Межрегиональной ассоциации когнитивных исследований, а также автором четырех монографий и более 60 публикаций по проблемам сложности текста и компьютерной лингвистики.
e-mail: maki.solovyev@mail.ru
https://orcid.org/0000-0003-4692-2564

Владимир Владимирович БОЧКАРЕВ – научный сотрудник научно-исследовательской лаборатории «Лингвистика и искусственный интеллект». Сфера его интересов – квантитативная лингвистика.
e-mail: vbochkarev@mail.ru
https://orcid.org/0000-0001-8792-1491

Венера Рустамовна БАЙРАШЕВА – кандидат физико-математических наук, доцент кафедры теоретической кибернетики Казанского федерального университета. Сфера ее научных интересов – компьютерная и теоретическая лингвистика.
e-mail: vbayrasheva@gmail.com
https://orcid.org/0000-0002-1729-034X

Book review / Рецензия

Review of Mel'čuk, Igor & Milićević, Jasmina. 2020
An Advanced Introduction to Semantics.
A Meaning-Text Approach.
Cambridge: Cambridge University Press

Svetlana IVANOVA🆔✉

Pushkin Leningrad State University, St. Petersburg, Russia
✉s.ivanova@lengu.ru

Рецензия на книгу
Mel'čuk, Igor & Milićević, Jasmina. 2020
An Advanced Introduction to Semantics.
A Meaning-Text Approach.
Cambridge: Cambridge University Press

С.В. ИВАНОВА🆔✉

Ленинградский государственный университет имени А.С. Пушкина,
Санкт-Петербург, Россия
✉s.ivanova@lengu.ru

This review is a certain digression from what a typical review for a journal is supposed to be. The laws of the genre are violated for a number of reasons. No doubt this review has a personal touch, as the name of Igor Mel'čuk cannot help

evoking a storm of memories and associations for those who started or did their research back in the 70-s and early 80-s. And that is the generation I belong to. That's why the book under review is not only a book for me, it is not only about linguistics, even though it is a great contribution to it – it is part of history for me as a member of the research community in Russia whose life in linguistics started in the early and mid-80s. In fact, this review could have been entitled "Forty years later". About forty years ago my Ph.D. course started at the Maurice Thorez Institute of Foreign Languages in Moscow. We, Ph.D. students from different parts of the Soviet Union, wrote our theses. We worked in the libraries, talked to each other and shared our findings not only in the classrooms of the well-known school of foreign languages but also in the kitchen in the dorms in Usachevka – a great place where future researchers were bred. Among this shared knowledge transferred by word of mouth in the kitchen (a symbolic place for those who lived back then) was the name of Igor Mel'čuk whispered with amazement and awe. We all knew his name, we avidly read his articles in the Problems of Structural Linguistics (Мельчук 1968, 1972) series but could not refer to them in our reference lists, it was one of the conventions, an unwritten rule we all knew and had to abide to. We read his books and articles (mostly articles) though, we admired him, we all respected his stance as a researcher. Since then, Igor Mel'čuk has delighted his followers with numerous books, and I cannot help mentioning some of the latest publications (Mel'čuk 2018, 2021). Now about forty years later, yours truly, inspired and humbled, is writing a review of the book by one of the Mohicans of linguistics. Holding this book, I again feel as a Ph.D. student in Moscow who is privileged to do this job and overwhelmed with responsibility.

Natural language semantics has been a great challenge for linguistics since day one. After all, what is there in language that makes it a salient means of communication? It is meaning. That puts semantics in the limelight of linguistic studies and makes it one of the greatest challenging objects for linguists to describe. Various schools in linguistics approach it from different angles, which results in a diversity of answers. Igor Mel'čuk and Jasmina Milićević offer their take on semantics which draws upon and incorporates achievements of numerous schools of linguistic thought but, first and foremost, it fits in the Meaning-Text Theory. The authors start with fundamentals (Part I), concentrate on meaning in language and its description (Part II), and then discuss Meaning-Text model of semantics (Part III).

In Fundamentals (Part I) the authors paint a general picture to show the correlation between language as a bigger entity and semantics as one of its components. Then the layout of the book displays the logic of zooming in on lexical semantics which afterwards is followed with the part about meaning in the framework of the text. Nevertheless, in the preliminary notes (p. xix) the authors give a fair warning that the book cannot and, I would say, should not be read linearly as language itself is not a linear structure, everything in it is interwoven and interconnected. In fact, this is a great advantage of the book since you can go to

whatever section you are interested in and try to fathom those acute questions which need answering.

First of all, the authors differentiate between two meanings of the term 'semantics'. On the one hand, semantics is a component of language and, on the other, it is a branch of linguistics. Semantics in this latter meaning is a relatively new discipline. It is located on the crossroads of linguistics and some other spheres of research like cognitive science, psychology, artificial intelligence and so on that have vested interest in semantics, as linguistic meaning is not only the pivot of language structure but also a mysteriously elusive product of communication that could be approached and described from different perspectives for its enigmatic nature to be disclosed.

After giving some general remarks on the nature of semantics as a branch of linguistics, the authors switch over to the object of their studies which is semantics as a component of language, the prime of linguistic description. Since semantics is integrated into the system of language as a component, it could be described along the same lines as the bigger entity. Language as this bigger entity is understood as "a set of rules encoded in the brains of its speakers that establish a correspondence between meanings of language and their expression, or texts of language" (p. 4). Meanings are expressed by texts and in texts and thus could be extracted from them. Obviously, meanings and texts are linked together by means of rules which "constitute language proper" (p. 11). Consequently, linguistic meaning is a formal description (p. 71) and it operates as "shallow" meaning ('non-pragmatic, non-extralinguistic, non-encyclopedic meaning') opposite to "deep" meaning which is accessible through life and situational experience (encyclopedic together with pragmatic knowledge and referential identification) as well as logical capacities (p. 73–74). There are three aspects of linguistic meaning: propositional, communicative and rhetorical (p. 76). Propositional meaning is "the meaning that targets the state of affairs described by this expression – that is, entities and facts in the world, as well as the relations between them, including the Speaker's interior states, such as his thoughts, attitudes, desires, etc." (p. 76). Communicative meaning is tied to the Speaker's intentions whereas rhetorical meaning has to do with the Speaker's stylistic preferences and intentions (p. 77).

After putting linguistic meaning in relation to language as a system of formal rules, the authors define the former. Linguistic meaning is a complex entity which "is described in terms of discrete semantic units – semantemes and semantic dependency relations between them" (p. 79). A semanteme is interpreted as "a lexical meaning – that is, the signified of a full lexical unit of language" (p. 79) and characterized by structural complexity. The authors introduce a whole set of notions that make up semantic metalanguage and perform the role of instruments of linguistic description of linguistic meaning, including semanteme, arguments, semantic actants, semantic dependency.

Semantemes are represented by two major classes, those of semantic predicates and semantic names. If semantic predicates are an "incomplete," or "binding,"

meanings (when used by the Speaker, they require that some other meanings, called its arguments, be expressed alongside it) (p. 83), "a semantic name is a complete and non-binding meaning; it cannot have arguments" (p. 85). The term semantic dependency is introduced to show how the semantic predicate is related to semantic actants or arguments. These two types of semantemes are the basic instruments of semantic decomposition which is necessary not only for defining words per se but also will be further used to describe the propositional meaning. Thus, homogeneity of semantic representations on different levels of the language system is ensured. The notion of semantic decomposition is crucial for this approach as it reveals the hierarchical structure of linguistic meaning which "is composed of clearly identifiable units" (p. 89). It also makes it possible to represent the internal structure of linguistic meaning on any level of language structure. This is a particularly important tenet of the authors' conception as the authors show further how it all works in lexicography and in the text and how text semantics can be formalized for machine translation, for one.

If Part I addresses fundamental problems of semantics, Part II deals with lexical meaning and the application of lexical meaning description in lexicography. As the authors state, lexicography cuts through all branches of linguistics as it studies words from all angles. Since the authors maintain that semantics may be formalized, they posit what a lexicographic definition should be. This definition includes formal description of meaning. Lexicographic practice embraces all kinds of units: lexical units (lexemes and idioms), collocations, and cliches. Besides, the authors introduce another opposition which could be of use in meaning description: lexeme *vs* phraseme which is "a phrase consisting of at least two lexemes that is paradigmatically constrained" (p. 105). If collocations and cliches as examples of phrasemes are well-known, it is of interest to get familiar with nominemes (p. 111) and pragmatemes (p. 112).

The formalization of the meaning concerns connotation as well. The authors understand connotation as "a semantic characteristic which, in language L, is attributed to the entities denoted by l(exeme) but which does not constitute a part of its meaning and, consequently, is not a component of l(exeme)'s lexicographic definition" (p. 135). Though the authors believe in linguistic intuition, they maintain that (1) connotations should be supported by linguistic evidence, (2) they cannot be part of the lexicographic definition of the lexeme and (3) they can be indicated in the semantic zone of the dictionary entry, under a special heading "Connotation". The authors also dwell upon lexical relations (synonymy, antonymy, polysemy, conversion) and lexical functions. Both chapters are a vivid illustration how the idea of formalization may be applied in semantics.

Part III puts the problem under consideration into the perspective of the 'Meaning-Text' theory. In the long run, the description of linguistic meaning is necessitated by the fact that words do not exist separately. Lexical semantics is revealed in sentences, that's why the 'Meaning-Text' theory is essential for the

application of the formal procedures. Semantic networks as an embodiment of formalism are used to reveal the propositional meaning of lexical units and that of utterances, in particular, sentences (p. 255). The sentential meaning, the authors argue, is established by means of indicating dependency relations. "Deep-syntactic relations are "generalized" syntactic relations, each subsuming several concrete surface-syntactic relations" (p. 294). The three fundamental distinctions in syntax, according to the authors, are: coordination ~ subordination; weak subordination ~ strong subordination; modification ~ actancy (p. 295). According to this theory, linguistic knowledge is represented as a huge inventory of correspondences between thought and speech, and semantics is viewed "as a component, or module, of the linguistic system, whose functioning is simulated by a corresponding linguistic model" (p. xvii). Within this approach, semantics is viewed formally and gets its formal representation in the system of rigorous notions, specified by about eighty mathematical-like definitions. The sentential meaning is presented as semantic representation which is an aggregate embracing semantic structure, the semantic-communicative structure, the rhetorical structure, and the referential structure (p. 257). All of them are represented as networks with nodes or tree diagrams.

All in all, when it comes to analysis, language boils down to form and meaning as a linguistic sign unites the plane of content and the plane of expression, the signified and the signifier. Thus, according to the authors' stance, semantics is inseparable from formal representation. The third constant used to describe semantics is function as that's what the unity of form and meaning is employed for. The authors are consistent in their approach and use these three constants of linguistic description to reveal semantics as the "crucial component of human language" (p. i). Semantics is represented as a system of rigorous rules and notions with an emphasis on formal modelling (p. xvii).

As a reviewer I can add that the view on semantics presented in the book by Igor Mel'čuk and Jasmina Milićević ties together many a thread woven into the fabric of present-day linguistic theory. The authors put together the legacy of well-known linguistic schools to the advantage of the approach they present. For example, the Saussurean postulate that language is a system of systems is reflected in the idea that the Text-Meaning model at large works on all levels of language: semantic, syntactic, morphological, phonological (p. 13–14). The idea that semantics consists of a lexicon and grammar (p. 18) relates to M.A.K. Halliday's idea about lexicogrammar which was suggested by him within functional linguistics and afterwards this approach was considered as one of the major principles in cognitive linguistics. The understanding of language as a structure goes back to Ferdinand de Saussure as well as to American structuralists according to who language can be presented formally. When we read that "the lexical stock is a psychological and neurological reality, namely, particular information stored in the brains of speakers" (p. 99), we cannot but remember Bloomfieldian linguistics. The

claim that "every language presents a unique conceptualization of the world; this phenomenon is often referred to as specific articulation of extralinguistic reality, which is "built into" a language and which it imposes on its speakers" (p. 81) is in line with Humboldt's lingua-philosophical heritage, or Potebnya's 'close' and 'distant' meanings. Obviously, the statement that "languages differ widely in the quantity of information that they can "squeeze" into their semantemes" (p. 81) corresponds to the tenets of modern cultural linguistics. Moreover "semanteme packaging" is different in different languages which can be exemplified with verbs of motion in English in contrast to Russian or Spanish. Semantic decomposition which presupposes that complex meanings may be represented with simpler meanings (using cause verbs and such) is the development of the ideas of generative semantics. The way lexical meaning is described and the range of instruments used for this type of analysis shows close ties with Russian school of semantics. There is a certain correlation of Meaning-Text theory with Anna Wierzbicka's Natural Semantic Metalanguage with a certain digression outlined by the authors: "Whereas for us semantic primitives represent a goal, for Wierzbicka they are a starting point: she posits several dozens of universal primitive meanings called semantic primes (such that they have lexical – or at least morphological – expressions in all the languages of the world) and uses them to describe all lexical and grammatical meanings in all languages" (p. 92). Another theory that is close to the authors' understanding of how semantic components function is the theory of semantic roles elaborated by Charles Fillmore: semantic roles are associated with semantic actants (p. 96). The authors' stance was influenced by the generative grammar tradition (Noam Chomsky's deep and surface structures) and Roman Jacobson's ideas (texts are explained in Roman Jacobson's terms as "something immediately perceptible", whereas linguistic meaning is "something conceivable and translatable").

To sum up, the book by Igor Mel'čuk and Jasmina Milićević is undoubtedly a great read for students of linguistics and linguists at large, especially researchers working in the field of linguistic semantics, machine translation, lexicography, language learning and teaching, to name just a few. This book is another good reason to delve into Mel'čuk's ideas, reflect on them and admire the parsimonious and elegant ways a true researcher can approach one of the most mysterious and challenging objects of linguistic description. Bravo, Igor Alexandrovich! Encore!

REFERENCES

Mel'čuk, Igor A. 1968. A model for Hungarian declension (The noun). *Problemy strukturnoy lingvistiki* (Problems of structural linguistics). Moscow: Nauka. 344–373. (In Russ.).

Mel'čuk, Igor A. 1968. On possessive forms of the Hungarian noun. *Problemy strukturnoy lingvistiki* (Problems of structural linguistics). Moscow: Nauka. 326–341. (In Russ.).

Mel'čuk, Igor A. 1972. On suppletivity. *Problemy strukturnoy lingvistiki* (Problems of structural linguistics). Moscow: Nauka. 396–438. (In Russ.).

Mel'čuk, Igor A. 2018. Anna Wierzbicka, Semantic Decomposition, and the Meaning-Text Approach. *Russian Journal of Linguistics* 22 (3). 521–538. https://doi.org/10.22363/2312-9182-2018-22-3-521-538

Mel'čuk, Igor. 2021. *Ten Studies in Dependency Syntax*, Walter de Gruyter GmbH, Berlin/Boston.

Bionote:
Svetlana IVANOVA is Dr Habil. in Philology, Professor, Head of the Department of Romance and Germanic Philology and Linguodidactics of Pushkin Leningrad State University. Her research interests include text linguistics, English grammar, discourse analysis, and media linguistics.
e-mail: s.ivanova@lengu.ru
https://orcid.org/0000-0002-0127-9934

Сведения об авторе:
Светлана Викторовна ИВАНОВА – доктор филологических наук, профессор, заведующий кафедрой романо-германской филологии и лингводидактики, Ленинградский государственный университет имени А.С. Пушкина. Сфера ее научных интересов: лингвистика текста, английская грамматика, дискурс-анализ, медиа-лингвистика.
e-mail: s.ivanova@lengu.ru
https://orcid.org/0000-0002-0127-9934

Book review / Рецензия

Review of Mel'čuk, Igor. 2021. *Ten Studies in Dependency Syntax.* Berlin/Boston: De Gruyter Monton

Olga SOLOPOVA ⓘ ✉

South Ural State University (National Research University), Chelyabinsk, Russia
✉o-solopova@bk.ru

Рецензия на книгу Mel'čuk, Igor. 2021. *Ten Studies in Dependency Syntax.* Berlin/Boston: De Gruyter Monton

О.А. СОЛОПОВА ⓘ ✉

*Южно-Уральский государственный университет
(национальный исследовательский университет), Челябинск, Россия*
✉o-solopova@bk.ru

The book by Igor Mel'čuk, one of the greatest living scholars of modern philology, presents a number of case studies carried out within the Meaning-Text approach with its insistence on scrutinizing the strictly dependent nature of syntax.

Beginning in the 1960s, the work on the development of a functional model of Natural Language grew into a linguistic theory, which came to be known as the Meaning-Text theory, and into the study of the system of logical rules, which constitutes a functional model of language. Though the Meaning-Text approach is

connected to the traditional study of language structure in its purest state, it differs in its specific emphasis on linguistic synthesis, which is considered more important than linguistic analysis. It could well be said that a readiness to analyze anything is one of the distinctive marks of modern linguistics (a position shared by the current reviewer). However, it is no longer possible to claim to analyze text comprehension without first studying text production. Thus, as Mel'čuk puts it, "text synthesis is obviously primary to analysis" (Mel'čuk 2021: 10). As the author declares in his prologue, "the Meaning-Text approach presupposes three levels of syntactic description for a sentence: the deep-syntactic representation, the surface-syntactic representation, and the deep-morphological representation" (Mel'čuk 2021: 1), with the focus of the book being on the surface-syntactic module of the Meaning-Text model.

There are four parts of the book, collectively broad in linguistic range, all covering major topics.

Part I (Chapter 1) gives a brief overview of the Meaning-Text model, its three postulates, main notions, rules (semantic, deep-syntactic, surface-syntactic, morphological), basic structures of the model of linguistic representations (deep-syntactic, surface-syntactic, surface-morphological), paradigmatic (both free and restricted) and syntagmatic lexical choices. This part explains every concept in a comprehensive manner, which is crucial for understanding the syntactic dependencies and general inventories discussed in the following parts. In fact, Igor Mel'čuk has taken the transition between the surface-syntactic representation of the sentence and its deep-morphological representation, and married them to a useful functional model so that the reader may easily apprehend the author's theory of the world of syntax and syntactic dependencies.

In Part II (Chapters 2–5), the author turns to the notion of surface-syntactic relations found in various languages (Russian, English, German, French, Korean, Hindi, and many others), which are grouped according to their syntactic properties, starting with the strongest subordinative dependencies and going towards the weakest coordinative links: actantial, modifying, attributive, auxiliary, circumstantial, etc. (more than a hundred types). Various surface-syntactic relations have been discovered by the author to prove his theory, and it becomes apparent that these formulations can be extended and deepened (Chapter 2). This part lays the groundwork for the comparative study of languages, and particularly those languages that diverge in many respects from a syntactical point of view and are fundamentally different in character. For instance, Chapter 3 tackles a popular and controversial linguistic problem of basic surface-syntactic relations and the notions of syntactic subject and direct object. Mel'čuk proposes a set of rigorous definitional parameters and discusses several complex cases involving the syntactic subject. The emphasis in this chapter is on the coding, or definitional properties of the syntactic subject; its characterizing, or behavioral aspects, are discussed only insofar as they bear on the former. Chapter 4 continues the focus on the issue of basic surface-syntactic relations and the problem of the so-called "multiple subjects

and objects" in Korean. Fascinated by the syntactic phenomena of Korean, the author attacks the problem within the framework of the Meaning-Text perspective basing his analysis on general typological considerations, dependent syntactic representation and a formal system of linguistic notions and terms. Examining strings of same-case Korean nouns, he formulates "privileges" (the author's term) of the surface syntactic subject and object in the language and convincingly demonstrates that Korean has neither "multiple subjects" nor "multiple direct objects": "what is theoretically not possible is impossible in any of the possible worlds (in any language)" (Mel'čuk 2021: 203–204). Chapter 5 is a perceptive analysis of the syntactic organization of genitive adnominal dependents in Russian: it contains a list of the semantic relations between a noun and its syntactic nominal dependent in the genitive case without preposition, and an inventory of surface syntactic relations within phrases of this type.

Part III (Chapters 6–9) cracks "some hard nuts in syntax" by dependency description, in an account that is pleasingly metaphorical without a loss of its formal and even formalized character. The major purpose of this part is to analyze four hot topics which constitute a challenge for today's linguistics. Chapter 6 brings a typology of relative clauses to the level of modern linguistic theory. Determined to elaborate and perfect the conceptual apparatus of linguistics and propose a rigorous definition of the relative clause, Igor Mel'čuk attempts to refine and standardize the terminology (modifier, clause, relative, head / top node, antecedent), reformulating some concepts that have been used in previous studies, defining the relative clause and its types: restrictive and descriptive, differentiating each from constructions often confused with relative clauses and, finally, sketching a typology of restrictive relative clauses, each featuring a distinctive syntactic structure and allowing different deep-morphological realizations. The chapter marks an important transition in typological research on relative clauses. In Chapter 7, the author addresses a complicated and significant issue of binary conjunctions, sketches their general typology and offers an inventory of Russian binary conjunctions, proposing the surface syntactic description of sentences containing these conjunctions. The next chapter (8) represents a crucial shift in the way in which a passive construction in Mandarin Chinese has been understood. The author provides research evidence that shows that the category of voice does not exist in this language. The point that deserves particular attention in this chapter, in my view, concerns the requirements proposed for scientific definition (both substantive and formal). The principles that are formulated might well apply to the definition of linguistic phenomena and concepts in other fields. Chapter 9 explores pronominal idioms in a case study of a blasphemous noun in the Russian language. Mel'čuk illustrates their surface-syntactic structure, their possible lexicographic description and offers a universal typology of phrasemes, including major classes of lexemic, morphemic, and syntactic instances.

In Part IV (Chapter 10), the author discusses word order in Russian within the framework of the Meaning-Text approach, examining the operation of linearization

and following his own step-by-step strategy. The two steps presented here are input and output representations, defined in terms of dependencies and major classes of linearization rules, their form and their interaction. The importance of Mel'čuk's work here consists in the primacy of this work on word order, which formally presents both the syntactic input structure and the linearization rules.

There is a creative dimension in Igor Mel'čuk's logical and practical approach to presenting the material. The work, which at times has a technical aspect, vastly expands the "syntactic" horizons of the reader. The incorporation of various schemes and tables with rigorous definitions, clear examples and comprehensive explanations makes this an excellent read, as they greatly assist the reader's understanding of dependency syntax. Mel'čuk breaks down syntactic structures so that their meanings are potentially accessible to all.

He shows, with great force and erudition, that the Meaning-Text approach affects other disciplines and fields of study. Meaning-Text models are of high practical utility and crucial for understanding the way philology is applied in various disciplines in the modern world: Natural Language Processing (machine translation, text generation, etc.), psycholinguistic experimentation, teaching and learning languages, manufacturing dictionaries, grammars, and manuals. This book is also important for its typological perspectives: Igor Mel'čuk has refined existing typologies and offered new inventories of syntactic phenomena. Moreover, the examples to illustrate the key points of his theory are taken from more than seventy languages, comprising different language families (Austronesian, Afro-Asiatic, Indo-European, Niger-Congo, Sino-Tibetan, Trans-New-Guinea, etc.), some unrelated isolates (Basque, Burushaski, Seri) and some dead languages (Ancient Chinese, Ancient Greek, Biblical Hebrew, Sanskrit).

Communication involves two parties, the author and the reader, and Igor Mel'čuk always tries to draw the reader into his theory, encouraging them to engage with the text through his specific tone and style of writing. The emphasis is on communicative interaction: the author directly addresses the reader so that the latter will feel more involved with the story of dependency syntax, inviting them to solve linguistic problems and puzzles. The author's comments and remarks, combined with his omniscient point of view, help the reader to better understand the nature of the Meaning-Text approach, exemplified in ten studies in dependency syntax. His metaphors, analogies and jokes used to explain complex phenomena paint a vivid syntactic picture and, in my opinion, deserve as much interpretation as the text itself. On this point, parenthetically, among my favorites are a joke about a Soviet-era military medical cadet and an enema, and a comparison of a woman to a syntactic subject.

From a personal perspective, an ideal title for the review could be "A Journey into the Jungle of Syntax". The field of syntax may at times appear to be a kind of jungle, a thriving system that comprises masses of topics discussed, each having its own basic notions, definitions, principles to study, numerous typologies, examples in foreign languages and challenges to the reader's survival. As a guide in these

difficulties, Igor Mel'čuk assists his reader to find a way through the bushes and trees of syntactic structures. The correct patterns and perspectives are conveyed by means of schemes, tables, symbols, abbreviations and codes. Thanks to the clarity of Mel'čuk's theoretical accounts, then, the complexity of dependency syntax is rendered much less daunting for readers of all levels of linguistic competence.

Bionote:
Olga SOLOPOVA is Dr Habil. in Philology, Professor at the Department of Linguistics and Translation at the Institute of Linguistic and International Communications of South Ural State University (National Research University). Her research interests include typological linguistics, grammar, metaphor studies, discourse analysis, and diachronic linguistics.
e-mail: o-solopova@bk.ru
https://orcid.org/0000-0003-4170-7267

Сведения об авторе:
Ольга Александровна СОЛОПОВА – доктор филологических наук, профессор кафедры лингвистики и перевода института лингвистики и международных коммуникаций ФГАОУ ВО Южно-Уральский государственный университет (национальный исследовательский университет). Сфера ее научных интересов: типологическая лингвистика, грамматика, метафорология, дискурсология, диахроническая лингвистика.
e-mail: o-solopova@bk.ru
https://orcid.org/0000-0003-4170-7267