

ВЕСТНИК НОВОСИБИРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

Научный журнал
Основан в ноябре 1999 года

Серия: Лингвистика и межкультурная коммуникация

2019. Том 17, № 1

СОДЕРЖАНИЕ

Прикладная и компьютерная лингвистика

<i>Григорьев Ю. Д., Мартыненко Г. Я.</i> Комбинаторные варианты рифм в сонетах Рильке: квантитативно-типологический подход	5
<i>Масевич А. Ц., Захаров В. П.</i> Частотное поведение прилагательных цвета в русских поэтических текстах	21
<i>Зибер И. А., Мороз Г. А.</i> Исследование акустической вариативности <i>s</i> методом главных компонент	49
<i>Азеркович И. Л.</i> Использование мер семантической близости для распознавания кореференции в русском языке	65
<i>Муравьев Н. А., Ольшевская М. Ю.</i> Подходы к составлению лексических минимумов в России и за рубежом: проблемы и перспективы	78
<i>Сибирцева В. Г., Фролова Н. Х.</i> Исследование потенциала компьютерных программ для стилистического и переводческого анализа текста и его практическое применение	90

Лингвокультурная когнитивистика и психолингвистика

<i>Бочкарев А. Е.</i> О любопытстве и смежных понятиях в русском языковом сознании	102
<i>Трофимова Ю. М.</i> Ментальные аспекты лингвистической реконструкции (на материале «Гетики» Иордана)	115
<i>Соловьева И. В.</i> Социокультурный статус формулы извинения <i>pardon</i> в английском языке по данным корпусных баз	125
<i>Себрюк А. Н.</i> Лингвистическая дискриминация и лингвистическая апроприация как примеры феномена двойственного сознания в США	134
<i>Кафтанов Р. А.</i> Динамика образа войны в русском языковом сознании (психолингвистический аспект)	149
Информация для авторов	161

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Том 17, № 1
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2019, vol. 17, no. 1

VESTNIK

NOVOSIBIRSK STATE UNIVERSITY

Scientific Journal
Since 1999, November
In Russian

Series: Linguistics and Intercultural Communication

2019. Volume 17, № 1

CONTENTS

Applied and Computer Linguistics

- Grigoriev, Yuri D., Martynenko, Gregory Ya.* Combinatorics of Rhyming Variants in Sonnets by R. M. Rilke: A Quantitative and Typological Approach 5
- Masevich, Andrey Ts., Zakharov, Victor P.* Frequency Behavior of Color Adjectives in Russian Poetic Texts 21
- Sieber, Inna A., Moroz, George A.* Estimating the Acoustic Variation of *s* via Principal Component Analysis 49
- Azerkovich, Ilya L.* Using Semantic Relatedness Measures in Coreference Resolution for Russian 65
- Muravyev, Nikita A., Olshevskaya, Maria Yu.* Approaches to the Composition of Lexical Minima in Russia and Abroad: Problems and Prospects 78
- Sibirtseva, Vera G., Frolova, Natalia H.* Testing Computer Programs' Potentialities in Stylistic and Translational Comparative Text Analysis 90

Cultural Linguistics, Cognitive Studies and Psycholinguistics

- Bochkarev, Andrey E.* Curiosity and Related Concepts in the Russian Linguistic Consciousness 102
- Trofimova, Yulia M.* Linguistic Reconstruction and Its Mental Foundations (Based on Jordanes' *Getica*) 115
- Soloveva, Inna V.* Language Corpora Data as the Source of Sociocultural Information about Pardon Apology 125
- Sebryuk, Anna N.* Linguistic Discrimination and Linguistic Appropriation as the Examples of Double Consciousness in the U.S. 134
- Kaftanov, Roman A.* Dynamics of the Image of War in the Russian Language Consciousness (Psycholinguistic Aspect) 149
- Instructions to Contributors 161

Editor-in-Chief Dr. Sc. (Linguistics) Maria K. Timofeeva (Novosibirsk)
Deputy Editor-in-Chief Dr. Sc. (Linguistics), Prof. Michèle Debrenne (Novosibirsk)
Executive Secretary Cand. Sc. (Linguistics), Assoc. Prof. Anatoli F. Fefelov (Novosibirsk)
Secretary Alevtina N. Leonova (Novosibirsk)

Foreign Text Editor

Cand. Sc. (Linguistics), Assoc. Prof. Olga A. Ryzhkina (Novosibirsk)

Members of the Editorial Board

Cand. Sc. (Linguistics), Assoc. Prof. Natalia V. Kozlova (Novosibirsk)
Dr. Sc. (Philology), Prof. Alexey D. Palkin (Moscow)
Cand. Sc. (Philology), Assoc. Prof. Ivan A. Remorov (Novosibirsk)
Cand. Sc. (Linguistics), Assoc. Prof. Olga A. Ryzhkina (Novosibirsk)
Cand. Sc. (Philology), Assoc. Prof. Anna L. Solomonovskaya
PhD (Cross-Cultural Psychology), Assistant Prof. Snežana Stupar-Rutenfrans (Utrecht, Netherlands)
Cand. Sc. (Linguistics), Senior Researcher Victor P. Zakharov (St. Petersburg)

Editorial Board of the Series

Academician of the Russian Academy of Sciences, Prof. A. E. Anikin (Novosibirsk)
Dr. Sc. (Philology), Prof. V. E. Gorshkova (Irkutsk), Dr. Sc. (Linguistics), Senior researcher L. L. Iomdin (Moscow),
Dr. Sc. (Linguistics), Prof. N. B. Koshkareva (Novosibirsk),
Cand. Sc. (Linguistics), Doctorat National en sciences du langage A. M. Lavrentev (Lyon, France),
Dr. Sc. (Linguistics), Prof. O. N. Aleshina (Taipei, The Republic of China, Taiwan),
Dr. Sc. (Linguistics), Prof. D. B. Nikulicheva (Moscow), Dr. Sc. (Philology), Prof. L. G. Panin (Novosibirsk),
Dr. Sc. (Linguistics), Prof. Z. G. Proshina (Moscow),
Dr. Sc. (Philosophy, Neuroscience), Assoc. Prof. A. N. Savostyanov (Novosibirsk),
Dr. Sc. (Linguistics), Prof. I. Ya. Selyutina (Novosibirsk), Dr. Sc. (Philology), Prof. I. V. Silantyev (Novosibirsk),
Dr. Sc. (Linguistics), Prof. V. A. Stepanenko (Irkutsk), Dr. Sc. (Linguistics), Prof. Ye. F. Tarasov (Moscow),
Dr. Sc. (Linguistics), Prof. N. V. Ufimtseva (Moscow), PhD (Philology) V. Obry (Mulhouse, France),
PhD (Philology), Assoc. Prof. I. Chekhova (Sofia, Bulgaria),
Dr. Sc. (Linguistics), Prof. I. V. Shaposhnikova (Novosibirsk),
Dr. Sc. (Linguistics), Prof. N. N. Shirobokova (Novosibirsk)

*The journal is published quarterly in Russian since 1999
by Novosibirsk State University Press*

The address for correspondence

Institute of Humanities, Novosibirsk State University
1 Pirogov Street, Novosibirsk, 630090, Russian Federation

Tel. +7 (383) 363 42 23

E-mail address: lingua@vestnik.nsu.ru

On-line version: <http://elibrary.ru>

УДК 81'25

DOI 10.25205/1818-7935-2019-17-1-5-20

Комбинаторные варианты рифм в сонетах Рильке: квантитативно-типологический подход

Ю. Д. Григорьев¹, Г. Я. Мартыненко²

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

² Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

Аннотация

В рамках методик математического стиховедения осуществлен комбинаторный анализ корпуса немецких сонетов на материале творческого наследия Райнера Марии Рильке, построена классификация вариантов рифмовки сонетов на основе их симметричных отношений в катренной и терцетной частях. Классификация не учитывает так называемые нерегулярные сонеты. Согласно данной классификации вводятся понятия симметричных, частично симметричных и несимметричных сонетов и комбинаторным способом подсчитывается возможное количество сонетов каждого типа. С учетом этой классификации выявляется уникальный комбинаторный потенциал поэта, в частности исследуется эволюция структуры сонетов Рильке от «Новых стихотворений» к «Сонетам к Орфею». Вводятся понятия независимых и зависимых сонетов. На основе этих понятий показано, что «Новые стихотворения» содержат только независимые, т. е. с неповторяющимися способами рифмовки, сонеты, в то время как в «Сонетах к Орфею» этот принцип нарушается. Выделена ядерная часть сонетов Рильке, состоящая из тех независимых сонетов, которые порождают весь корпус сонетов Рильке, т. е. всю совокупность независимых и зависимых сонетов. В соответствующих подсчетах учтены также восемь сонетов из примыкающих к «Сонетам к Орфею». Отдельное внимание в статье уделено симметричным сонетам. Описана структура всех 18 симметричных сонетов Рильке, среди которых 11 являются независимыми («Новые стихотворения» – 4, «Сонеты к Орфею» – 7). В качестве примера проведен стилеметрический анализ ритмического поля одного из симметричных сонетов. Показано, что его внутренний ритм близок к структуре золотого сечения. Высказана гипотеза о влиянии вариантов рифмовки, присущих симметричным сонетам, на внешний и внутренний ритмы сонета.

Ключевые слова

математическое стиховедение, комбинаторика, немецкий сонет, Рильке, способы рифмовки, катрен, терцет, классификация рифм, симметрия, симметричные сонеты, несимметричные сонеты, независимые сонеты, зависимые сонеты, эволюция, гармония

Для цитирования

Григорьев Ю. Д., Мартыненко Г. Я. Комбинаторные варианты рифм в сонетах Рильке: квантитативно-типологический подход // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 5–20. DOI 10.25205/1818-7935-2019-17-1-5-20

Combinatorics of Rhyming Variants in Sonnets by R. M. Rilke: A Quantitative and Typological Approach

Yuri D. Grigoriev¹, Gregory Ya. Martynenko²

¹ St. Petersburg State Electrotechnical University

² St. Petersburg State University
St. Petersburg, Russian Federation

Abstract

Combinatorial analysis of the German sonnet corpus based on mathematical study of verse was carried out on poetic works by Rainer Maria Rilke. Sonnets rhyming methods classification based on their symmetry relations in quatrain and tercet parts was constructed. According to this classification, which does not take into consideration the so-called

© Ю. Д. Григорьев, Г. Я. Мартыненко, 2019

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Том 17, № 1
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2019, vol. 17, no. 1

irregular sonnets, the concepts of symmetric, partially symmetric and asymmetrical sonnets are introduced and the possible quantity of sonnets of each type is calculated by means of the combinatorial method. Taking into account these notions, the unique combinatorial potential of the poet has been revealed, and the evolution of Rilke's sonnets structure from "New poems" ("Neue Gedichte") to "Sonnets to Orpheus" ("Sonette an Orpheus") is also investigated. The notions of symmetrical and independent sonnets are introduced. Using these notions, it is shown that "New poems" contain just independent sonnets, which have non-recurring rhyming variants, while in "Sonnets to Orpheus" this principle is not respected. The nuclear part of sonnets, consisting of those independent sonnets that generate the whole amount of Rilke's sonnets (i. e., the entire collection of independent and dependent sonnets), has been determined. Eight sonnets adjacent to the "Sonnets to Orpheus" are also taken into account in the corresponding calculations, too. Special attention is paid to symmetrical sonnets. The structure of all 18 symmetric Rilke's sonnets is described, among which 11 are independent (four sonnets in "New poems", and seven sonnets in "Sonnets to Orpheus"). As an example, the stylometric analysis of the rhythmic field of one of the symmetrical sonnets was performed. It is shown that its internal rhythm is close to the structure of the golden section. A hypothesis is expressed concerning the influence of rhyming variants on the external and internal rhythms of the sonnet.

Keywords

mathematical study of verse, combinatorics, German sonnet, Rilke, rhyming methods, quatrain, tercet, rhyme classification, symmetry, symmetrical sonnets, asymmetrical sonnets, independent sonnets, dependent sonnets, evolution, harmony

For citation

Grigoriev Yuri D., Martynenko Gregory Ya. Combinatorics of Rhyming Variants in Sonnets by R. M. Rilke: A Quantitative and Typological Approach. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 5–20. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-5-20

Введение

Сонет относится к твердым поэтическим формам. Наряду с сонетом к этим формам принадлежат баллада, рондо, октава, триолет, ритурнель, вилланель и некоторые другие структуры [Рукавишников, 1925; Шенгели, 1960; Шапир, 2000]. Но наибольшую популярность приобрел сонет.

Предельная формализация организации сонета, упорядочивая и дисциплинируя стихию поэтического творчества, создает благоприятные условия для кристаллизации и концентрации поэтической мысли. В классическом (романском) варианте сонет включает два катрена и два следующих за ними терцета с определенной системой рифмовки в пределах строф и между строфами. Такая организация призвана обеспечить распределение смысла в сонетном «пространстве» на пути от центральной идеи (тезис, завязка, зачин) и сопутствующей идеи (антитезис) через синтез этих идей к завершительному выводу (концовке, развязке) [Гаспаров, 2001б].

Сонет родился в переходную, переломную эпоху, на гребне перехода от Средневековья к эпохе Возрождения. Первый сонет был написан нотариусом Якопо да Лентини (1215–1233), а канонические правила написания сонета были сформулированы спустя столетие, в 1332 г., юристом из Падуи Антонио да Темпо (конец XIII в. – 1339 г.). Любопытно, что оба изобретателя были правоведами, законниками, знатоками регламентов и правил. Сонет – это не только словесная, но и «наглядная» пространственная структура, которая воспринимается и по горизонтали – слово за словом, и по вертикали – строка за строкой, строфа за строфой – симультанно как целостность. Жесткая структура сонета, его компактность и экономность дает возможность убрать лишние детали, сохранив основной смысл. В сонете проявляется эффект ореховой скорлупы, выраженный в английской идиоме «*in a nutshell*» («в ореховой скорлупе»), которая означает «в двух словах», «не вдаваясь в подробности».

Спустя много столетий после изобретения сонета он претерпел многочисленные изменения, но основные требования к его организации сохранили силу. Возникли различные версии в строфике, ритмике сонета: итальянская, французская, английская. В основе этих версий лежат особенности конкретного языка, например место ударения в слове, особенности культурных традиций и некоторые субъективные моменты.

Определенные различия сложились и в системах рифмовки. Так, во французской поэзии предпочтение отдавалось сочетанию «закрытой» (охватной) рифмовки в катренах и «открытой» (перекрестной) в терцетах или наоборот, открытой в катренах и закрытой в терцетах [Гаспаров, 2001а. С. 224].

Когда из Франции сонет перекочевал в Англию, то рифмующая пара уже не начинала, а заканчивала собой терцеты (и весь сонет), а четверостишие, встав перед нею, принимало вид третьего катрена (подробнее об этом см.: [Гаспаров, 2001а. С. 225]).

В итальянском сонете из-за того, что ударение в слове чаще падает на предпоследний слог, строки имеют обычно женское окончание, поэтому поэт не должен заботиться о чередовании мужских и женских рифм. Итальянские поэты могли позволить себе более разнообразные способы рифмовки в терцетах. Как отмечает Гаспаров [Там же. С. 226], если для французского сонета характерна парная рифмовка в начале терцетов, а для английского – в конце, то для итальянского сонета парная рифмовка в терцетах невозможна.

Таким образом, структурные модификации в трех перечисленных странах коснулись и строфики, и ритмики, и способов рифмовки. Особые традиции структуризации сложились и в других европейских странах, например в Германии, Австрии и России.

Мы остановимся в данной работе на способах рифмовки преимущественно в немецком сонете. В качестве материала избран корпус сонетов Рильке.

С целью расширения возможностей характеристики сонета мы избрали принцип *симметрии* его катренной и терцетной частей. Изложению этого подхода и возникающим при этом выводам посвящена данная работа.

Сонет

Немецкий сонет согласно «первому впечатлению» строится по французскому образцу, хотя возможны и вариации в «итальянском стиле». Так ли это и в какой мере, покажет наше исследование.

Перечисление сонетов

В сонете содержатся два катрена (две строфы по четыре строки) и два терцета (две строфы по три строки). Катрены рифмуются независимо, а терцеты, наоборот, связаны между собой одной или тремя парными рифмами. Системы рифмования катренов: (12)(34) – смежная, (13)(24) – перекрестная, (14)(23) – кольцевая. Таким образом, возможны 9 комбинаций рифмовок катренов, скажем, (12)(34) и (12)(34), (12)(34) и (13)(24) и т. д.

В силу того что рифмы бывают *мужские* (а) и *женские* (б) (с ударением на последний слог в строке или его отсутствием соответственно), то таких пар различных рифм будет 16, скажем, (аа)(аа), (аа)(аb), (аа)(bа) и т. д. (перечень 16-ти 4-разрядных чисел в двоичной системе). Таким образом, приходим к $Q = 9 \times 16 = 144$ видам катренов.

Аналогично рассматриваем систему терцетов. Поскольку имеется $C_6^2 = 15$ способов рифмовок терцетов, скажем, (12)(34)(56), (12)(35)(46) и т. д., и 8 комбинаций мужских и женских рифм, скажем, (ааа), (ааb) и т. д. (запись 3-разрядного числа в двоичной системе), то в общей сложности имеем $T = 15 \times 8 = 120$ видов терцетов. Следовательно, общее число различающихся по структуре и виду рифм сонетов равно $N = Q \times T = 17\,280$. Даже если не считать перестановку катренов за новый сонет, все равно общее количество разных сонетов оказывается очень большим, и ни один поэт в одиночку не в силах в своем творчестве охватить всё их разнообразие.

Симметричные сонеты

Будем различать *симметричные*, *частично симметричные* и *несимметричные* сонеты. Каждый сонет содержит катренную (8 строк) и терцетную (6 строк) части. Катренную часть на-

зываем симметричной, если она имеет смежный (12)(34)(12)(34), перекрестный (13)(24)(13)(24) или кольцевой (14)(23)(14)(23) вид и на нее накладывается одинаковая для обоих катренов система мужских и женских рифм. Перечень симметричных катренов представлен в табл. 1. Симметричные части далее помечаем знаком плюс (+), а несимметричные – знаком минус (–).

Таблица 1

Симметричные катрены

Table 1

Symmetrical quatrains

(12)(34)	(13)(24)	(14)(23)
(12)(34)	(13)(24)	(14)(23)
(aa)(aa)	(aa)(aa)	(aa)(aa)
(ab)(ab)	(ab)(ab)	(ab)(ab)
(ab)(ba)	(ab)(ba)	(ab)(ba)
(ba)(ab)	(ba)(ab)	(ba)(ab)
(ba)(ba)	(ba)(ba)	(ba)(ba)
(bb)(bb)	(bb)(bb)	(bb)(bb)

Терцеты с симметричной структурой представлены на рисунке 1, а соответствующие им симметричные терцетные части – в табл. 2. Из нее видим, что всего их насчитывается 28. По количеству связывающих терцеты рифм их можно разбить на одно- и трехсвязные. Особо выделяется структура терцетной части (16)(25)(34), для которой любая из восьми систем рифмовок оставляет терцетную часть симметричной. Все остальные структуры этим свойством не обладают. Структуру (16)(25)(34) и соответствующие им терцетные части можно назвать *сильно симметричными*.

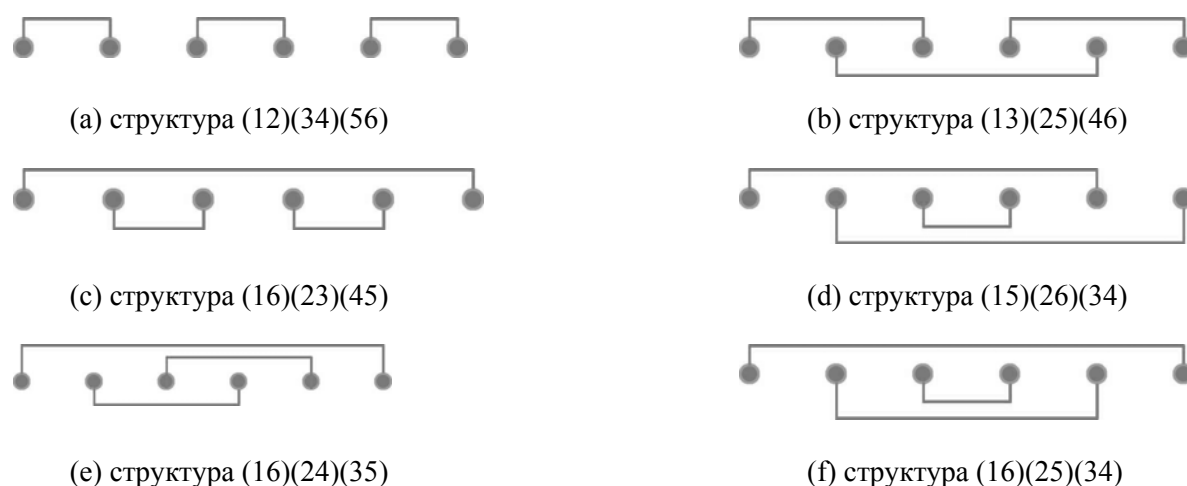


Рис. 1. Симметричные терцеты

Fig. 1. Symmetrical tercets

Таблица 2

Симметричные терцеты

Table 2

Symmetrical tercets

1-связные терцеты			3-связные терцеты		
(12)(34)(56)	(13)(25)(46)	(16)(23)(45)	(15)(26)(34)	(16)(24)(35)	(16)(25)(34)
(aaa)	(aaa)	(aaa)	(aaa)	(aaa)	(aaa)
(aba)	(aba)	(aba)	(aab)	(abb)	(aab)
(bab)	(bab)	(bab)	(bba)	(baa)	(aba)
(bbb)	(bbb)	(bbb)	(bbb)	(bbb)	(bbb)
					(baa)
					(bab)
					(bba)
					(bbb)

Сонет называется *симметричным* (+ +), если обе его части являются симметричными, *частично симметричным* (+ –) или (– +), если только одна из его частей является симметричной, и *несимметричным* (– –) в остальных случаях. Таким образом, любой из сонетов можно отнести к одному из четырех типов.

К симметричному типу, очевидно, относятся $18 \times 120 - (144 - 18) \times 28 = 504$ сонета, к частично симметричным – $18 \times (120 - 28) + (144 - 18) \times 28 = 5\,184$ сонета и к полностью несимметричным – $(144 - 18) \times (120 - 28) = 11\,592$ сонета. Эти соотношения представлены в табл. 3.

Таблица 3

Классификация общего корпуса сонетов по виду симметрии

Table 3

Sonnets corpus classification according to their symmetry forms

		Терцетная часть		Всего	Терцетная часть, %		Всего, %
		+	–		+	–	
Катренная часть	+	504	1 656	2 160	2,92	9,58	12,5
	–	3 528	11 592	15 120	20,42	67,08	87,5
Всего		4 032	13 248	17 280	23,34	76,66	100

Из табл. 3 следует, что сонетов типа (+ –) (симметричная катренная и несимметричная терцетная части соответственно) существует 9,58 %, типа (– +) – 20,42 % и т. д. Одна из наших дальнейших целей – построение на основе анализа корпуса сонетов Рильке таблицы, аналогичной табл. 3.

Характеризация сонетов Рильке

Поэзия Рильке с точки зрения формы его стихотворений крайне разнообразна. Мы остановимся только на сонетах. В поле нашего зрения будут два его сборника стихов – «Новые стихотворения» [Rilke, 1979] и «Сонеты к Орфею» [Рильке, 2002] (далее СО), а также небольшой цикл «Восемь сонетов из примыкающих к СО» [Rilke, 1987].

Сборники «Новые стихотворения» и СО состоят из двух частей каждый. В первом случае будем говорить о стихотворениях 1907 и 1908 гг., а во втором – о частях 1 и 2. Принадлеж-

ность сонетов к этим группам обозначаем как (1907), (1908) и СО-1, СО-2 соответственно. Примыкающие сонеты обозначаем как СО-3. Два других известных сборника Рильке «Часослов» [Рильке, 2004] и «Книга образов» [Rilke, 1976] сонетов не содержат. Очевидно, сонетов у Рильке значительно больше. Однако будем опираться только на упомянутые источники.

Сонеты Рильке разделим на две группы – *регулярные* и *нерегулярные*. К нерегулярным относятся сонеты, структура которых отлична от канонической, скажем, (aaa)(b) в катренах или (b)(b) в терцетах (123)(456). Имеющиеся у Рильке нерегулярные сонеты представлены в табл. 4 (далее они не рассматриваются).

Таблица 4

Нерегулярные сонеты Рильке

Table 4

The list of symmetric sonnets of Rilke

Сборник	Сонет	Катренная часть			Терцетная часть	
		катрен 1	катрен 2	рифма	терцеты	рифма
(1907)	Капиталь [Das Kapitol]	(13)(24)	(13)(24)	(ba)(ab)	(135)(246)	(ba)
	Последний вечер [Letzter Abend]	(13)(24)	(14)(23)	(ab)(bb)	(1245)(36)	(ba)
	Куртизанка [Die Courtisane]	(1)(234)	(1)(234)	(b)(a)	(16)(25)(34)	(aba)
	Римские фонтаны [Römische Fontäne]	(13)(24)	(13)(24)	(ba)(ba)	(1)(235)(46)	(bab)
	Остров, I [Die Insel]	(13)(24)	(14)(23)	(aa)(ab)	(145)(236)	(ba)
(1908)	Парк попугаев [Papageienpark]	(1234)	(1234)	(b)(b)	(135)(246)	(bb)
СО-2	26	(14)(23)	(13)(24)	(ab)(bb)	(146)(235)	(bb)

Примечание: (1907) – «Новые стихотворения», (1908) – «Новых стихотворений другая часть», СО-2 – «Сонеты к Орфею», часть 2.

Введем еще одну спецификацию сонетов Рильке. Возьмем какую-либо группу сонетов. Пусть N – их количество. Разобьем их на K непересекающихся множеств, каждое из которых содержит сонеты с совпадающими структурами катренов и терцетов и системой рифмовок. Очевидно, $K \leq N$. В случае $K = N$ такие сонеты назовем *независимыми*.

Пусть n_i – количество сонетов в i -м множестве, т. е.

$$N = \sum_{i=1}^K n_i, \quad n_i \geq 1.$$

Выберем из каждого множества по одному представителю и образуем из них множество независимых сонетов. Это множество будет *максимальным* в том смысле, что добавление к нему любого из оставшихся в K множествах сонетов приводит к тому, что оно перестает быть множеством независимых сонетов. Это множество всех представителей назовем *базисом* исходного множества сонетов. *Высотой* h_i ($i = 1, \dots, K$) каждого из векторов базиса назовем количество сонетов в том множестве, представителем которого он является. Все сонеты каждого из K множеств, кроме его представителя в базисе, назовем *зависимыми*. Очевидно, в качестве представителя данного множества в базисе можно выбрать любой зависимый сонет данного множества.

В нашем случае для выбора представителей используется лексикографический принцип – в качестве представителя конкретного множества всегда выбирается тот сонет, который в исходных списках сонетов является в хронологическом смысле наиболее ранним или, если речь идет об одном списке, занимает в нем более высокое место, т. е. имеет меньший порядковый номер.

Подсчет показывает, что рассматриваемый корпус сонетов Рильке содержит 79 независимых сонетов. Это его базис, а весь корпус сонетов (без учета нерегулярных) составляет $100 = 21 + 79$ сонетов.

Распределение сонетов по их структурным особенностям представлено в табл. 5 и 6.

Таблица 5

Распределение сонетов разных типов по сборникам

Table 5

Distribution of sonnets of different types in sonnets collections

Вид	«Новые стихотворения»		СО		Примыкающие к СО	Всего
	(1907)	(1908)	часть 1	часть 2		
Всего	23	21	26	29	8	107
Нерегулярные	5	1	–	1	–	7
Зависимые	–	–	6	11	4	21
Независимые	18	20	20	17	4	79

Таблица 6

Зависимые сонеты

Table 6

Dependent sonnets

i	Независимые сонеты		Зависимые сонеты	h_i
1	(1907; 3)	«Полуденный ангел» [Der Engel]	(CO-1; 1, 26)	3
2	(1908; 2)	«Артемид Критская» [Kretische Artemis]	(CO-1; 6, 7, 25), (CO-2; 13, 22, 23)	7
3	(1908; 13)	«Поздняя осень в Венеции» [Venezianischer Morgen]	(CO-3; 1)	2
4	(CO-1; 4)		(CO-2; 7, 14, 28), (CO-3; 2, 3)	6
5	(CO-1; 5)		(CO-1; 23)	3
6	(CO-1; 15)		(CO-1; 23)	2
7	(CO-1; 21)		(CO-2; 8)	2
8	(CO-2; 5)		(CO-2; 10)	2
9	(CO-2; 16)		(CO-2; 28)	2
10	(CO-2; 20)		(CO-2; 29)	2

Примечание: СО-1 – «Сонеты к Орфею», часть 1; СО-2 – «Сонеты к Орфею», часть 2; СО-3 – сонеты, примыкающие к «Сонетам к Орфею», $h_i \geq 2$ – высоты независимых сонетов, имеющих повторения, $\sum_{i=1}^{10} h_i - 10 = 21$ – количество зависимых сонетов.

Анализ табл. 5 и 6 приводит к следующим выводам.

1. В «Новых стихотворениях» содержится 6 нерегулярных сонетов, в то время как в СО – только 1.

2. В «Новых стихотворениях» *нет независимых сонетов*, в то время как в СО их насчитывается 21 (они перечислены в 4-м столбце табл. 6). Это наводит на мысль, что либо Рильке из всего разнообразия сонетов сознательно отбирал неповторяющиеся структуры, записывая уже использованные структуры на листочке бумаги или храня их в памяти, либо на чисто интуитивном уровне избегал повторения структур исходя из только ему ведомой мелодии каждой структуры. Ответ на этот вопрос мы, видимо, никогда не получим.

3. В «Сонетах к Орфею» и примыкающих к ним Рильке отступает от отмеченного выше принципа *независимости (неповторяемости)*. Таким образом, рассматриваемый корпус сонетов Рильке содержит 79 независимых сонетов. Это его базис, а весь корпус сонетов (без учета нерегулярных) составляет $100 = 21 + 79$ сонетов.

Теперь приступим к анализу 79 независимых сонетов Рильке. Полученная на основе их анализа табл. 7 по структуре аналогична табл. 3.

Таблица 7

Классификация независимых сонетов Рильке по виду симметрии

Table 7

Classification of independent Rilke's sonnets according to their symmetry forms

		Терцетная часть		Всего	Терцетная часть, %		Всего, %
		+	–		+	–	
Катренная часть	+	11	29	40	13,92	36,71	50,63
	–	12	27	39	15,19	34,18	49,37
Всего		23	56	79	29,11	70,89	100

Анализ табл. 7 позволяет сделать следующие выводы.

1. Рильке примерно в равной степени использует сонеты с симметричной и несимметричной катренными частями (50,63 и 49,37 % соответственно).

2. Большая часть сонетов Рильке имеет несимметричную структуру терцетной части (70,89 %).

3. Вывод 1 не согласуется с табл. 3, а вывод 2, наоборот, вполне согласуется. Это может говорить о том, что Рильке более сознательно конструировал катренную часть, в то время как терцетная часть сонетов создавалась в большей степени спонтанно. Говоря статистическим языком, в последнем случае «выборка» каждого сонета из генеральной совокупности, содержащей 17280 сонетов, осуществлялась целенаправленно и осознанно. При случайном отборе такое невозможно.

4. Соотношение между симметричными и полностью несимметричными сонетами в табл. 3 и 7 сохраняется (в первом случае 2,92 и 67,08 % соответственно, во втором 13,92 и 34,18 %). Тот же вывод сохраняется и для всего корпуса сонетов: всего сонетов 100 ($= 38 + 54 + 8$), среди них 18 являются симметричными (18,0 % симметричных и 82,0 % остальных).

Симметричные сонеты Рильке

Теперь перейдем к более конкретному анализу симметричных сонетов Рильке.

Классификация

Все симметричные сонеты Рильке представлены в табл. 8, которая является очевидным подмножеством табл. 6. Независимые сонеты представлены в табл. 9. Классификация симметричных сонетов представлена в табл. 10.

Таблица 8

Перечень симметричных сонетов Рильке

Table 8

Classification of the 18 symmetric sonnets

Независимые сонеты		Зависимые сонеты	Высота
(1907; 10)	«Пленник, II» [Der Gefangene]		1
(1907; 14)	«Синяя гортензия» [Blaue Hortensie]		1
(1908; 3)	«Леда» [Leda]		1
(1908; 5)	«Сивилла» [Eine Sibylle]		1
(CO-1; 2)			1
(CO-1; 4)		(CO-2; 7, 14, 18), (CO-3; 2, 3)	6
(CO-2; 3)			1
(CO-2; 5)		(CO-2; 10)	2
(CO-2; 16)		(CO-2; 28)	2
(CO-2; 25)			1
(CO-3; 8)			1
Всего			18

Примечание: (1907) – «Новые стихотворения», (1908) – «Новых стихотворений другая часть», CO-1, CO-2, CO-3 – «Сонеты к Орфею», части 1, 2 и примыкающие к ним соответственно, в скобках – порядковые номера сонетов в сборниках.

Таблица 9

Структуры независимых симметричных сонетов

Table 9

Structures of independent symmetrical sonnets

Сборник	Сонет	Катренная часть			Терцетная часть	
		катрен 1	катрен 2	рифма	терцеты	рифма
(1907, 10)	Пленник, II [Der Gefangene]	(14)(23)	(14)(23)	(ab)(ab)	(12)(34)(56)	(aba)
(1907, 14)	Синяя гортензия [Blaue Hortensie]	(14)(23)	(14)(23)	(bb)(bb)	(13)(25)(46)	(bab)
(1908, 3)	Леда [Leda]	(14)(23)	(14)(23)	(ab)(ab)	(13)(25)(46)	(aba)
(1908, 5)	Сивилла [Eine Sibylle]	(14)(23)	(14)(23)	(ab)(ab)	(16)(24)(35)	(abb)
CO-1	2	(14)(23)	(14)(23)	(ab)(ab)	(16)(25)(34)	(aba)
	4	(13)(24)	(13)(24)	(ba)(ba)	(13)(25)(46)	(bab)
CO-2	3	(13)(24)	(13)(24)	(ba)(ba)	(13)(25)(46)	(aba)
CO-2	5	(13)(24)	(13)(24)	(ba)(ba)	(13)(25)(46)	(bbb)
CO-2	16	(13)(24)	(13)(24)	(ba)(ba)	(16)(25)(34)	(abb)
CO-2	25	(13)(24)	(13)(24)	(ba)(ab)	(13)(25)(46)	(bbb)
CO-3	8	(13)(24)	(13)(24)	(aa)(aa)	(13)(25)(46)	(aaa)

Классификация 18-ти симметричных сонетов

Таблица 10

Table 10

Classification of the 18 symmetric sonnets

		Терцетная часть			Всего
		(13)(25)(46)	(13)(25)(46)	(13)(25)(46)	
Катренная часть	(14)(23)	(1907, 14)	(1907, 10)	(1908, 5)	5
	(14)(23)	(1908, 3)	–	(CO-1; 2)	
	(13)(24)	11	–	(CO-2; 16)	13
	(13)(24)			(CO-2; 28)	
Всего		13	1	4	18

Примечание: (1907, 10) – «Пленник, II»; (1907, 14) – «Синяя гортензия»; (1908, 3) – «Леда»; (1908, 5) – «Сивилла», 11 сонетов со структурой [(13)(24)(13)(24)][(13)(25)(46)] (6 независимых и 5 зависимых) перечислены в табл. 9.

Анализируя табл. 8–10, можно сделать следующие выводы.

1. Всего Рильке написал 18 симметричных сонетов: 11 независимых и 7 зависимых.

2. В катренной части структуры (14)(23)(14)(23) и (13)(24)(13)(24) хронологически сменяют одна другую. От первой из них, начиная с (CO-1; 4), Рильке отказывается и переходит ко второй, уже не возвращаясь к первой. Все сонеты со структурой (14)(23)(14)(23) являются независимыми, т. е. не имеют повторений, их всего 5, в то время как катренная часть (13)(24)(13)(24) встречается 13 раз, и среди таких сонетов есть зависимые.

3. В терцетной части Рильке использует только три из шести возможных структур: (12)(34)(56), (13)(25)(46) и (16)(25)(34), которые встречаются соответственно 1, 13 и 4 раза (в сумме 18). Таким образом, 4 сонета из 18 являются сильно симметричными.

4. Согласно табл. 10 наиболее излюбленной формой симметричного сонета у Рильке является структура (13)(24)(13)(24)(13)(25)(46), которой обладают 11 из 18 симметричных сонетов. Вот их перечень: (CO-1; 4), (CO-2; 3, 5, 7, 10, 14, 18, 25), (CO-3; 2, 3, 8).

Из этих 11 симметричных сонетов один, а именно (CO-2; 3), отнесен в [Rilke, 2013. S. 109] к числу наиболее совершенных. Это, конечно, не означает, что у Рильке нет других претендентов на это высокое звание. Скажем, особого внимания, наряду с сонетом (CO-2; 3), заслуживает симметричный сонет «Леда», написанный на исключительно популярный сюжет древнегреческого мифа о Зевсе и Леде.

Симметричные и гармонические отношения в сонете

Возможно, симметричные сонеты обладают какой-то повышенной мерой гармонии. Чтобы проверить эту гипотезу, проведем квантитативный анализ одного из симметричных сонетов, опираясь на технику измерения ритма сонетов, изложенную в работе [Григорьев, Мартыненко, 2017]. Кроме того, проанализируем этот же сонет в двух его различных переводах, обратившись к измерению его точности и вольности с помощью соответствующих коэффициентов [Гаспаров, 2001б].

Строфический ритм. Рассмотрим проявление гармонии строфического ритма в форме золотого сечения применительно к сонету. Пусть C – *слоговый* и $T = T_{\text{катр}} + T_{\text{терц}}$ – *тоновый* объемы сонета соответственно [Холшевников, 1984. С. 33]. Последний складывается из тоновых объемов катренной и терцетной частей сонета. Будем рассматривать две тройки чисел:

$$(T, C - T, C), (T, T_{\text{терц}}, T_{\text{катр}}). \quad (1)$$

Первая из троек чисел в (1) определяет *внешний ритм* сонета. Говорят (см. [Григорьев, Мартыненко, 2017]), что он подчиняется золотому сечению, если тоновый объем T близок к величине

$$T^* = \frac{3 - \sqrt{5}}{2} = 0,382 \times C. \quad (2)$$

Другими словами, величина $T_0 = [T^* + 0,5]$ – это тоновый объем ($[...]$ – выделение целой части числа, взятого в скобки), отвечающий золотому сечению при слоговом объеме C .

Положим

$$\Delta_{3c}(C, T) = \left| \frac{C - T}{T} - \frac{C}{C - T} \right|. \quad (3)$$

Разностный индикатор (3) – это *мера отклонения* структуры сонета от золотой пропорции $B/M = C/B$, где $C = M + B$ – деление целого на меньшую и большую части. Если величина $\Delta_{3c}(C, T)$ близка к нулю, то внешний ритм сонета близок к золотому. Аналогично вводится понятие *внутреннего ритма*, обусловленного второй тройкой чисел в (1). По той же схеме строится и индикатор внутреннего ритма типа (3).

Коэффициенты Гаспарова. Вопрос об измерении точности и вольности поэтического перевода с помощью специальных коэффициентов был поставлен в работе [Гаспаров, 2001б]. Затем этот подход рассматривался в ряде последующих работ [Портер, 2004; Андреев, 2012; Григорьев, 2018]. Кратко поясним суть коэффициентов Гаспарова.

Введем следующие обозначения:

- N – количество знаменательных слов (существительные, прилагательные, глаголы, наречия) подстрочника;
- M – количество знаменательных слов в переводе;
- x – количество слов подстрочника, опущенных в переводе;
- y – добавленные слова, т. е. слова, которых не было в подстрочнике.

Коэффициенты точности и вольности Гаспарова в форме, предложенной в работе [Григорьев, 2018], имеют вид

$$G = \frac{N - x}{N}, \quad W = \frac{y}{M} = \frac{y}{N - (x - y)}. \quad (4)$$

Чем выше значение W , тем более вольным является перевод. Основная трудность, возникающая при подсчете коэффициентов (4), заключается в подсчете величин x и y , так как в общем случае эта задача решается неоднозначно.

Пример: симметричный сонет (СО-2; 3)

Все сонеты из сборника СО не имеют отдельных названий. В данном случае речь пойдет о сонете 3 из второй части СО. В нем Рильке обращается к теме зеркал, которую он затрагивает и в других своих стихотворениях, например в «Даме перед зеркалом» («*Dame vor dem Spiegel*») [Rilke, 1979]. Эта тема подробно обсуждается во вступительной статье Ахтырской к переводам Пурина [Рильке, 2002].

Приведем текст сонета (СО-2; 3) и его примерный подстрочник:

Spiegel: noch nie hat man wissend beschrieben,
was ihr in euerem Wesen seid.
Ihr, wie mit lauter Löchern von Sieben
erfüllten Zwischenräume der Zeit.

Зеркала: известно, что еще никогда не было
описано, что вы есть в своем существе.
Вы словно громкие дыры от просеивания,
заполненных промежутками времени.

Ihr, noch des leeren Saales Verschwender – ,
wenn es dämmt, wie Wälder weit...
Und der Lüster geht wie ein Sechzehn-Ender
durch eure Unbetretbarkeit.

Вы также расточитель пустынной залы –
когда смеркается, словно леса вдали.
И люстра проходит как 16-конечник
сквозь вашу неприступность.

Manchmal seid ihr voll Malerei.
Einige scheinen in euch gegangen –,
andere schicktet ihr scheu vorbei.

Иногда вам, живописным,
некоторые кажутся вошедшими в вас –
других вы робко отсылаете прочь.

Aber die Schönste wird bleiben –, bis
drüben in ihre enthaltenen Wangen
eindrang der klare gelöste Narziß.

Но останется самое прекрасное –
вплоть до другой стороны ваших щек
проник ясный растворившийся Нарцисс.

Начнем с ритмико-гармонического анализа сонета (СО-2; 3) в переводе первого из авторов. Попутно отметим, что стихотворение «Дама перед зеркалом» по нашей классификации является сонетом типа (+ –), т. е. этот сонет несимметричен. Ритмическая структура сонета (СО-2; 3) представлена в табл. 11 (в ней $T_{стр}$ – ударность строки).

Катренная часть сонета (СО-2; 3) имеет 5-стопный размер, а терцетная – 4-стопный. Это уже неклассический случай, хотя в остальном Рильке выдерживает классический канон сонета. Согласно табл. 11 получаем $C = 124$, $T = 51$ и тройки гармонических чисел, описывающих внешний и внутренний ритмы сонета:

$$(T, C - T, C) = (51, 73, 124), (T_{терц}, T_{катр}, T) = (20, 31, 51).$$

Индикаторы

$$\Delta_{внеш}(124, 51) = \left| \frac{73}{51} - \frac{124}{73} \right| = 0,267, \quad \Delta_{внут}(51, 20) = \left| \frac{31}{20} - \frac{51}{31} \right| = 0,095$$

показывают, что внутренний ритм рассматриваемого сонета близок к золотому сечению. Возможно, это связано с тем, что Рильке меняет ритм сонета, переходя в терцетной части к 4-стопному размеру. Наконец, согласно (2) получаем те тоновые объемы, которые в данном случае являются золотыми:

$$T_{внеш} = 47, T_{внут} = 19.$$

Действительно, значение 47 значительно отличается от 51, в то время как значение 19, наоборот, близко к 20.

Теперь оценим вольность перевода сонета (СО-2; 3). Подсчеты для перевода, представленного в табл. 11, дают следующие значения нужных нам величин:

$$N = 42, M = 48, x = 23, y = 29.$$

Таким образом, при переводе из 42 слов подстрочника было опущено 23 и вместо них введено 29 новых слов. Поэтому общее количество знаменательных слов в переводе возросло на 6 слов. Следовательно,

$$G = 0,452, W = 0,604.$$

Таблица 11

Ритмическое поле сонета Рильке (СО-2; 3)

Table 11

Rhythmic field of Rilke's sonnet (SO-2; 3)

Сонет	Ритмическая структура						$T_{\text{стр}}$
	I	II	III	IV	V	VI	
О, зеркала! Кто мог бы потаенный	УУ	У*	У*	УУ	У*	У	3
смысл ваш насыщенный описать.	У*	У*	УУ	У*	УУ		3
Вы дыры, словно сито, в мир межзвездный,	У*	У*	У*	У*	У*	У	5
что там за время – не понять.	У*	У*	УУ	У*	У*		4
	$T_{1, \text{капр}} =$						15
Вы взор, чей лик в пустынной зале тает,	У*	У*	У*	У*	У*	У	5
когда чуть виден лес вдали.	У*	У*	У*	У*	УУ		4
И свет шестнадцати свечей пронзает	У*	У*	УУ	У*	У*	У	4
те чувства, что давно прошли.	У*	УУ	У*	У*	УУ		3
	$T_{2, \text{капр}} =$						16
Вам кажется в мечтах порой:	У*	УУ	У*	У*			3
вот эти вечно будут с вами,	У*	У*	УУ	У*	У		3
а тем назначен путь иной.	У*	У*	У*	У*			4
	$T_{1, \text{терц}} =$						10
Прекрасны вы, как легкий бриз, –	У*	У*	У*	У*			4
касаясь ваших щек губами,	УУ	У*	У*	У*	У		3
в вас растворился сам Нарцисс.	УУ	У*	У*	У*			3
	$T_{2, \text{терц}} =$						10

Примечание: $T_{\text{стр}}$ – тоновый объем строки.

Отсюда видим, что отношение $R = G/W = 0,749$ принадлежит отрезку $[\Phi^{-1}, \Phi] = [0,618, 1,618]$, где $\Phi^{-1} = (\sqrt{5} - 1)/2 = 0,618$ – золотое сечение [Григорьев, Мартыненко, 2017]. Это означает [Григорьев, 2018], что в системном отношении данный перевод можно считать вполне гармоничным, так как коэффициент точности G достаточно близок к гармонической константе $G^* = (5 - \sqrt{5})/6 = 0,461$, а коэффициент вольности W хотя и в меньшей степени – к гармонической константе $W^* = 0,618$. Не вдаваясь в детали, отметим, что адекватность перевода в содержательном плане (на уровне использованных лексем) можно также оценить гармоническим средним $Ha(G, 1-W)$ коэффициентов G и $1-W$ [Портер, 2004], которое в нашем случае таково:

$$Ha(G, 1-W) = \frac{2G(1-W)}{G + (1-W)} = \frac{19}{45} = 0,422.$$

Как уже отмечалось, роль коэффициентов Гаспарова не следует преувеличивать, так как уже само их вычисление не вполне однозначно. Тем не менее в одном случае их использование, видимо, оправданно, а именно при сравнении вольности разных переводов при условии, что вычисления проводятся на базе одного подстрочника. Так, соответствующие вычисления для перевода сонета (СО-2; 3), выполненного Пуриным [Рильке, 2002], дают следующий результат: $N = 42$, $M = 40$, $x = 29$, $y = 27$.

Отсюда $G = 0,309$, $W = 0,675$. Следовательно, вольность перевода Пурина выше первого перевода. Он также менее гармоничен и в системном смысле. Однако эксперты могут квалифицировать перевод Пурина, как более близкий к оригиналу, скажем, в ритмическом и (или) образном отношении. В то же время семантически спорные строки, в том числе неэквивалентные с оригиналом, можно обнаружить в обоих переводах.

Заключение

В статье рассмотрена классификация сонетов Рильке на основе введенного авторами принципа симметрии. Введено понятие базиса корпуса сонетов и выделен базис сонетов Рильке, насчитывающий 79 сонетов. Показано, что сонеты «Новых стихотворений» не содержат зависимых сонетов, причем независимые сонеты распределились между «Новыми стихотворениями» и «Сонетами к Орфею» примерно поровну (38 и 37 соответственно).

На основании проведенного исследования можно сделать следующие выводы.

Симметричные сонеты составляют 18 % от общего числа рассмотренных сонетов (100 сонетов), среди которых выделен сонет (СО-2; 3). Его анализ стилеметрическими методами показывает, что в его структуре проявляются черты гармонии, связанной с золотым сечением.

Отметим, что стихов в форме сонета у Рильке на самом деле значительно больше, чем было рассмотрено в статье. Все они, безусловно, специалистам известны. В подтверждение сказанному приведем еще 10 сонетов (табл. 12), представленных в 1-м томе его избранных произведений [Rilke, 1963].

Сонеты 1–3 принадлежат Микеланджело, т. е. являются переводами с итальянского. Первый из них «Auf den Tod der Vittoria Colonna» посвящен знаменитой итальянской поэтессе Виттории Колонна, маркизе де Пескара (Vittoria Colonna 1490/92 – 1547). Сонеты 4–7 принадлежат французской поэтессе Луизе Лабе (Louize Labe, 1522–1566) и представляют выборку из ее 24-сонетного цикла (сонеты 4, 8, 9 и 20). Наконец, сонеты 8–10 принадлежат английской поэтессе Элизабет Баррет-Браунинг (Elizabeth Barret-Browning, 1806–1861) и заимствованы Рильке из ее цикла сонетов «Sonette aus dem portugiesischen» (сонеты 6, 22 и 39). Таким образом, изначально последние три сонета, возможно, были написаны на португальском языке.

Таблица 12

Характеристики сонетов Рильке –
переводов с итальянского, французского и английского языков

Table 12

Characteristics of Rilke's sonnets translated from Italian, French and English

№	Сонет	Катренная часть			Терцетная часть		Тип
		катрен 1	катрен 2	рифма	терцеты	рифма	
1	Kein Wunder ist's	(13)(24)	(13)(24)	(aa)(aa)	(14)(26)(35)	(aab)	+ –
2	Des Todes sicher, nicht	(14)(23)	(14)(23)	(ab)(ab)	(16)(23)(45)	(bbb)	+ +
3	Schon angelangt ist	(14)(23)	(14)(23)	(aa)(ab)	(14)(25)(36)	(aaa)	– –
4	Seitdem der Gott zuerst	(14)(23)	(14)(23)	(bb)(ab)	(13)(25)(46)	(abb)	– –
5	Ich leb, ich sterb: ich	(14)(23)	(13)(24)	(bb)(bb)	(16)(25)(34)	a(ba)	– +
6	Gleich, wenn ich	(14)(23)	(13)(24)	(ab)(bb)	(14)(26)(35)	(bb)	– –
7	Mir ward gewahrsagt	(13)(24)	(13)(24)	(aa)(ab)	(15)(23)(46)	(abb)	– –
8	Geh fort von mir	(13)(24)	(14)(23)	(aa)(aa)	(13)(26)(45)	(aba)	– –
9	Wenn schweigend	(14)(23)	(13)(24)	(ab)(bb)	(12)(35)(46)	(bab)	– –
10	Angesich in Angesich	(13)(24)	(12)(34)	(ba)(ba)	(16)(25)(34)	(aaa)	– +

Сонеты из табл. 12 имеют отличительную особенность – все они по нашей классификации являются независимыми. Среди них есть один симметричный (+ +) и шесть полностью несимметричных (– –). Во втором катрене 10-го сонета мы впервые встречаем смежную рифмовку, которая согласно классическим требованиям, предъявляемым к сонету, вообще не должна использоваться в сонете, что и имеет место во всех сонетах Рильке.

Список литературы / References

- Андреев С. Н.** Точный и вольный перевод стихотворного текста: количественная оценка // Изв. СмолГУ. Филологические науки. 2012. Вып. 4 (20). С. 88–95.
Andreev, S. N. Exact and a free translation of the poetic text: a quantitative estimation. *Bulletin of Smolensk State University. Philological Sciences*, 2012, no. 4 (20), p. 88–95. (in Russ.)
- Гаспаров М. Л.** Подстрочник и мера точности // Гаспаров М. Л. О русской поэзии. Анализы. Интерпретации. Характеристики. СПб.: Азбука, 2001б. С. 361–372.
Gasparov, M. L. The word-for-word translation. In: *On a Russian poetry: Analyses, interpretations, characteristics*. St. Petersburg, Azbuka Publ., 2001, p. 361–372. (in Russ.)
- Гаспаров М. Л.** Русский стих начала XX века в комментариях. М.: Фортуна Лимитед, 2001а. 288 с.
Gasparov, M. L. Russian verse of the beginning of 20 centuries in comments. Moscow, Fortuna Limited, 2001. 288 p. (in Russ.)
- Григорьев Ю. Д.** Системная гармония перевода поэтических текстов // Научный вестник НГТУ. 2018. Вып. 1 (70). С. 79–102.
Grigoriev, Yu. D. System harmony of poetic text translation. *Science Bulletin of the Novosibirsk State Technical University*, 2018, no. 1 (70), p. 79–102. (in Russ.)
- Григорьев Ю. Д., Мартыненко Г. Я.** Последовательности типа Фибоначчи: теория и прикладные аспекты. СПб.: Изд-во «Лань», 2017. 516 с.
Grigoriev, Yu. D., Martynenko, G. Ya. Sequences of Fibonachchi type: the theory and applied aspects. St. Petersburg, Lan' Publ., 2017, 516 p. (in Russ.)
- Портер Л. Г.** Количественные критерии адекватности поэтического перевода // Мир перевода. 2004. № 1 (11). С. 39–54.
Porter, L. G. Quantitative criteria of the a poetic translation adequacy. *World of Translation*, 2004, no. 1 (11), p. 39–54. (in Russ.)
- Рильке Р. М.** Сонеты к Орфею / Пер. с нем. А. Пурина. СПб.: Азбука-классика, 2002. 192 с.

- Rilke, R. M.** "Sonnets to Orpheus". Transl. from German by A. Purin. St. Petersburg, Azbuka-klassika Publ., 2002, 192 p. (in Russ.)
- Рильке Р. М.** Часослов. Книга для чтения на немецком языке / Коммен. А. А. Верлинской. СПб.: КОРОНА принт, КАРО, 2004. 160 с.
- Rilke, R. M.** The Book of Hours. The book for reading in German. Comments of A. A. Verlinskaya. St. Petersburg, CORONA print CARO Publ., 2004, 160 p. (in Russ.)
- Рукавишников И. С.** Твердые формы // Литературная энциклопедия: Словарь литературных терминов: В 2 т. М.; Л.: Изд-во Л. Д. Френкель, 1925. Т. 2: П–Я. Стб. 907–910.
- Rukavishnikov, I. S.** Fixed forms of poetry. In: Literary encyclopedia: Dictionary of literary terms. In 2 vols. Moscow, Leningrad, L. D. Frenkel Publishing House, 1925, vol. 2: P–Ya, col. 907–910. (in Russ.)
- Холшевников В. Е.** Мысль, вооруженная рифмами. Поэтическая антология по истории русского стиха. Л.: Изд-во ЛГУ, 1984. 448 с.
- Kholshchevnikov, V. E.** Thought armed with rhymes. The poetic anthology on stories of Russian verse. Leningrad, Leningrad State University Publ., 1984, 448 p. (in Russ.)
- Шапир М. И.** На подступах к общей теории стиха (методы и понятия) // Шапир М. И. *Universum versus: Язык – стих – смысл в русской поэзии XVIII–XX веков*. М.: Языки русской культуры, 2000. С. 84.
- Shapir, M. I.** On approaches to the general theory of a verse (methods and concepts). In: Shapir M. I. *Universum versus: Language – a verse – sense in Russian poetry of the 18–20th centuries*. Moscow, Languages of Russian culture Publ., 2000, p. 84. (in Russ.)
- Шенгели Г.** Твердые формы // Шенгели Г. Техника стиха. М: Гослитиздат, 1960. С. 288–297.
- Shengeli, G.** Hard forms. In: Shengeli G. *Technics of verse*. Moscow, Goslitizdat, 1960, p. 288–297. (in Russ.)
- Rilke, R. M.** Werke. Auswahl in drei Bänden. Leipzig, IM Insel Verlag, 1963, Bd. 1: Gedichte.
- Rilke, R. M.** Das Buch der Bilder. Insel Taschenbuch, 1976.
- Rilke, R. M.** Neue Gedichte und der Neuen Gedichte anderer Teil. Insel Taschenbuch, 1979.
- Rilke, R. M.** Acht Sonette aus dem Umkreis der «Sonette an Orpheus». In: Rilke R. M. *Sämtliche Werke in sechs Bänden*. Leipzig: Insel Verlag, 1987, Bd. 2: Gedichte.
- Rilke, R. M.** Die schönsten Gedichte. Insel Verlag, 4. Auflag, Printed in Germany, 2013.

*Материал поступил в редколлегию
Date of submission
05.12.2018*

Сведения об авторах / Information about the Authors

Григорьев Юрий Дмитриевич, доктор технических наук, профессор кафедры математического обеспечения и применения ЭВМ Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» (ул. Профессора Попова, 5, Санкт-Петербург, 197376, Россия)

Yuri D. Grigoriev, D. Sc., Professor, St. Petersburg State Electrotechnical University (5 Prof. A. Popov Str., St. Petersburg, 197376, Russian Federation)

yuri_grigoriev@mail.ru
ORCID 0000-0002-5962-2395

Мартыненко Григорий Яковлевич, доктор филологических наук, профессор кафедры математической лингвистики Санкт-Петербургского государственного университета (Университетская наб., 11, Санкт-Петербург, 199034, Россия)

Gregory Ya. Martynenko, D. Sc., Professor, St. Petersburg State University (11 Universitetskaya Emb., St. Petersburg, 199034, Russian Federation)

g.martynenko@gmail.com

Частотное поведение прилагательных цвета в русских поэтических текстах

А. Ц. Масевич¹, В. П. Захаров²

¹ Санкт-Петербургский государственный институт культуры

² Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

Аннотация

Лингвистические корпуса и компьютерные технологии дают возможность осуществлять исследования и получать результаты на больших текстовых массивах, что прежде, до появления этих средств, было недостижимо. Настоящая статья описывает некоторые количественные данные использования прилагательных, обозначающих цвета, в русских поэтических текстах. Произведено сравнение частотности базовых прилагательных цвета (белый, черный, красный, зеленый, желтый, синий, голубой, коричневый, оранжевый, розовый, фиолетовый) в текстах четырех корпусов НКРЯ – основного, газетного, устного и поэтического. Сравнение осуществлялось двумя способами: в каждом корпусе определялась относительная частота (ipm) каждого из перечисленных выше прилагательных, а также ipm всех прилагательных, имеющих тег «цвет» семантической разметки НКРЯ. Оба способа определения показали, что в текстах поэтического корпуса частотность прилагательных цвета значительно выше, чем в текстах трех других корпусов (основном, газетном и устном). Установлено также, что убывание значений частоты прилагательных цвета в текстах всех четырех корпусов в значительной степени коррелирует с эволюционно-исторической схемой Берлина – Кея. Помимо этого, проведено исследование частоты встречаемости прилагательных цвета в поисковой системе Государственного Эрмитажа (массив аннотаций к произведениям искусства), где нам пришлось ограничиться абсолютным числом текстов, содержащих то или иное прилагательное цвета, поскольку данная поисковая система не содержит лингвистических характеристик. Корреляция полученных данных со схемой Берлина – Кея выражена в данном случае намного слабее, чем в корпусах НКРЯ. В статье детально описана методика извлечения прилагательных цвета из авторских подкорпусов поэтического корпуса НКРЯ и создания списков этих прилагательных средствами менеджера Sketch Engine из текстов авторов, отсутствующих в поэтическом корпусе НКРЯ. На основе анализа 180 лемм прилагательных цвета, извлеченных из текстов тридцати шести авторов, произведена категоризация прилагательных цвета. Для анализа и категоризации привлечены также и данные поисковой системы Эрмитажа. Приведены также количественные данные о представленности (репрезентации) выделенных категорий прилагательных в исследованных текстах. В ходе исследования выявлены некоторые недостатки семантической разметки в НКРЯ. Таким образом, мы выявили некоторые закономерности частотного поведения прилагательных цвета в русском языке и предложили один из возможных вариантов их классификации.

Ключевые слова

прилагательные цвета, теория базовых цветовых терминов Берлина – Кея, Национальный корпус русского языка, поэтический корпус, семантическая разметка

Для цитирования

Масевич А. Ц., Захаров В. П. Частотное поведение прилагательных цвета в русских поэтических текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 21–48. DOI 10.25205/1818-7935-2019-17-1-21-48

Frequency Behavior of Color Adjectives in Russian Poetic Texts

Andrey Ts. Masevich¹, Victor P. Zakharov²

¹ St. Petersburg State Institute of Culture

² St. Petersburg State University
St. Petersburg, Russian Federation

Abstract

Linguistic corpora and computer technologies make it possible to search and analyze large amount of unstructured texts. This paper describes in detail the method we used to extract adjectives of color from the poetic texts found in the Poetry corpus of the Russian National Corpus and from various Internet sources. Using a base of 180 lexical units extracted from the poetic texts of 36 authors, we devised a categorization scheme for adjectives of color; this scheme al-

so incorporated data obtained from the Hermitage Museum information system. It includes five categories, with the largest ("Derivatives") broken down into three subcategories. Our paper further provides quantitative data indicating the extent to which the elements of the different categories are represented in the texts; and from this data we were able to draw a preliminary conclusion vis-à-vis the use of adjectives of color by various authors. Specifically, we compared the frequency of use of basic adjectives of color (белый – white; чёрный – black; красный – red; зелёный – green; жёлтый – yellow; синий – blue; голубой – light blue; коричневый – brown; оранжевый – orange; розовый – pink; and фиолетовый – violet) in the texts of four corpora of the Russian National Corpus, i.e., the Basic corpus, Newspaper corpus, Oral corpus, and Poetry corpus.

Our paper describes some patterns of frequency behavior for adjectives denoting colors in Russian poetic texts. We arranged the adjectives retrieved from each corpus in order of decreasing frequency and found that, in all four corpora, the order of the adjectives of color largely correlates with the evolutionary theory of Berlin – Kay which describes the order of appearance of color adjectives in the historical development of different languages. The comparison showed that the frequency of adjectives of color in the Poetry corpus is significantly higher than in the other three corpora. In addition, we searched the information system of the State Hermitage Museum and established that the frequency correlation between adjectives of color and the Berlin – Kay evolutionary model is expressed there weaker than in the RNC corpora. Also, in the course of our study, we found a few semantic tagging errors in the Russian National Corpus. The patterns of frequency behavior of color adjectives revealed in the Russian language may become a reliable basis for further research. Their classification needs more attention too.

Keywords

color adjectives, Berlin – Kay's basic color terms theory, National corpus of Russian Language, poetic corpus, semantic tagging

For citation

Masevich Andrey Ts., Zakharov Victor P. Frequency Behavior of Color Adjectives in Russian Poetic Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 21–48. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-21-48

Введение

Обозначение цвета – исключительно популярная тема в лингвистике. В данной работе мы попытаемся ввести в рассмотрение пять принципиально новых аспектов проблемы.

1. Аспект лингвистической статистики. Все наши выводы базируются на количественных результатах, полученных на представительных корпусах.

2. Эволюционный аспект. Выявлена частичная корреляция между эволюционной теорией базовых цветовых терминов Брента Берлина и Пола Кея (Brent Berlin & Paul Kay) [Berlin, Kay, 1969] и статистикой частоты встречаемости прилагательных цвета в русском языке.

3. Предложен вариант классификации прилагательных цвета и показаны отличия в использовании прилагательных цвета разных типов.

4. Выявлены особенности использования прилагательных цвета в текстах различных жанров и в языке для специальных целей, а именно в описаниях произведений изобразительного искусства.

5. Описано частотное поведение единиц данной лексической группы на материале поэтических текстов.

В качестве материала исследования использовались:

1) поэтический корпус Национального корпуса русского языка (НКРЯ) [Гришина и др., 2009];

2) корпуса поэтических текстов собственной генерации, созданные специально для целей данного исследования в системе Sketch Engine;

3) основной, газетный и устный корпусы НКРЯ;

4) база данных описаний коллекций Государственного Эрмитажа.

В качестве инструментов исследования использовались:

1) поисковый аппарат (корпусный менеджер) НКРЯ, в том числе поиск по семантическому признаку;

2) корпусный менеджер Sketch Engine;

3) информационно-поисковая система Государственного Эрмитажа;

4) электронные таблицы MS Excel 2010.

Лингвистический корпус является моделью естественного языка. Очевидно, что, для того чтобы модель статистически достоверно отражала явления естественного языка, корпус должен быть репрезентативным (достаточно большим по объему) и сбалансированным (отражать различные литературные жанры и функциональные стили языка, причем в пропорции, которая соответствует доле данного жанра или стиля в реальном массиве текстов). Для русского языка больше всего этим требованиям удовлетворяет основной корпус НКРЯ (283 млн словоупотреблений). Исследование тех или иных лингвистических единиц или явлений применительно к стилевому многообразию требует, в свою очередь, корпусов определенных жанров (регистров). НКРЯ дает для таких исследований достаточный набор специальных корпусов. В нашем исследовании мы ограничили свой выбор такими специфическими подъязыками, как газетный язык, поэтический язык и устная речь, которым в НКРЯ соответствуют корпусы достаточного объема (228 521 421, 10 967 173 и 12 113 491 словоупотреблений, соответственно). Исследование проводилось в сравнении данных этих корпусов с данными основного корпуса.

Поскольку перечисленные корпусы содержат различное количество словоупотреблений, то абсолютные показатели количества тех или других единиц нельзя использовать как критерий для вывода о большей или меньшей частотности слова в одном корпусе по сравнению с другим. Поэтому для сравнений, как правило, используется относительная частота (частное от деления абсолютной частоты на объем соответствующего корпуса), которая «нормируется» путем преобразования количества случаев встречаемости слова по стандартной шкале, обычно в пересчете на 1 млн слов – *instances per million* (ipm) (относительная частота, умноженная на миллион). Нормированная частота обеспечивает основание для сравнения частотности каких-либо единиц в разных корпусах.

Одна из трудностей при анализе лексики в корпусной лингвистике – это проблема многозначности. И в нашем случае, нет, пожалуй, ни одного прилагательного цвета, по крайней мере из числа принадлежащих к основным категориям, которое не употреблялось бы в нескольких значениях, в том числе в переносном смысле – метафорически либо символически. Достаточно вспомнить выражения «белый свет», «чёрный список», «красной нитью», «оранжевая революция», «желтая пресса», «тоска зелёная», «синий чулок», «коричневая чума». Выскажем предположение: устойчивые словосочетания с изменением значения слова, как правило, образуют прилагательные, означающие базовые категории цвета, и, что также интересно, эти сочетания часто имеют негативную коннотацию.

Проблему учета многозначности можно было бы решить при наличии семантических корпусов с разметкой по значениям, однако таких корпусов мы не знаем. «Ручной» анализ при обработке больших объемов выходных данных нереален. Поэтому в данном исследовании мы попытались подключить к исследованию базу данных, где прилагательные цвета, как нам кажется, в подавляющем большинстве случаев используются в их прямом цветовом значении. Это база данных описаний предметов коллекции Государственного Эрмитажа. Однако поисковая система Эрмитажа на лексический запрос выдает только абсолютное число документов (описаний предметов коллекций), в которых искомая лексическая единица встретилась. Остается неизвестным суммарное число словоупотреблений в описаниях, образующих базу данных, из-за чего эти результаты поиска оказываются не сопоставимыми с характеристиками, получаемыми на основе корпусов НКРЯ и Sketch Engine. Тем не менее, сами по себе полученные результаты представляют интерес.

Одной из замечательных особенностей НКРЯ, отличающей его от многих других корпусов, является наличие семантической разметки [Кустова и др., 2005]. Ее назначение – выделение (поддержка поиска) групп лексических единиц, объединенных общими семантическими признаками. Семантические признаки приписываются всем словоформам знаменательных слов корпуса. К сожалению, в ходе нашей работы при использовании тегов семантической разметки в ней выявлено значительное число ошибок. Тем не менее, мы считаем,

что, несмотря на несовершенство существующей разметки, использование ее позволяет получить определенные результаты, которые нельзя получить никаким другим способом.

Корреляция между эволюционной теорией Берлина и Кея и частотой встречаемости прилагательных цвета в корпусах русского языка

Теория базовых цветовых терминов Берлина и Кея

Теория базовых цветовых терминов (категорий) Берлина – Кея состоит в том, что языки мира содержат полностью или частично общий набор слов цветообозначений, и эти слова появляются в языках в определенном порядке [Berlin, Kay, 1969]. В английском языке авторы выделяют 11 таких прилагательных, в славянских, в частности в русском, – 12.

Критерии принадлежности к базовым категориям следующие:

- 1) слово должно быть непроемным и не относиться к сложным словам;
- 2) его значение не должно быть уже значения другого имени цвета, указывающего на какой-либо близкий оттенок;
- 3) слово должно обладать широкой сочетаемостью;
- 4) для носителей данного языка слово должно быть психологически выделенным, значимым (salient).

Для английского языка Берлин и Кей указывают на цветовые термины, представленные на рис. 1. Для русского языка предлагается еще двенадцатое прилагательное – «голубой» (англ. light blue). «Исследовав десятки европейских языков, с одной стороны, и языков целого ряда примитивных культур, с другой, Берлин и Кей открыли “всеобщий” эволюционный закон, который выглядит так. Существует семь ступеней (стадий) развития цветовой терминологии, отражающих строгую последовательность появления в лексике языка каждого слова. Более поздняя стадия как бы включает в себя все предыдущие» [Василевич и др., 2005. С. 17].

«На самой “низшей” стадии, стадии I, во всех языках есть только два основных цветообозначения, и это во всех случаях оказываются слова, означающие белый и черный цвета. На стадии II к этим двум прилагательным цвета добавляется третье – и это всегда слово, обозначающее красный цвет. На стадиях III–IV добавляется каждый раз по одному слову из двух – “зелёный”, “жёлтый” (порядок может меняться). На стадии V появляется слово “синий”, на VI – “коричневый”, а высшая стадия VII характеризуется появлением сразу четырех прилагательных цветов – “розовый”, “оранжевый”, “фиолетовый” и “серый» [Там же] (рис. 2).

Концепция Берлина и Кея, сформулированная в 1969 г., не раз пересматривалась и уточнялась [Kay, McDaniel, 1978; Kay, Regier, 2006]. Есть языки, в которых последовательность появления слов, называющих цвета, не вполне соответствует закономерностям, зафиксированным у Берлина и Кея (например, китайский).

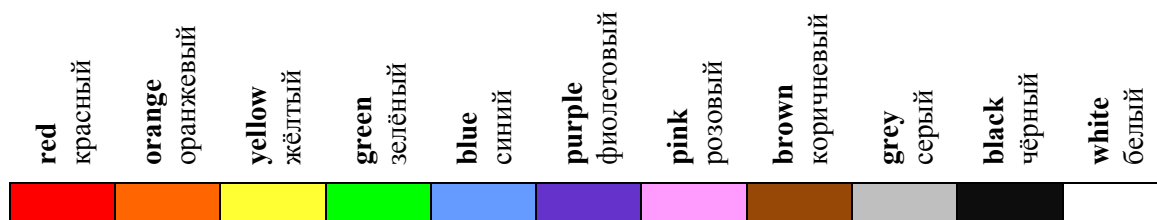


Рис. 1. Базовые цветовые категории по Берлину и Кею

Fig. 1. Basic color categories according to Berlin and Kay

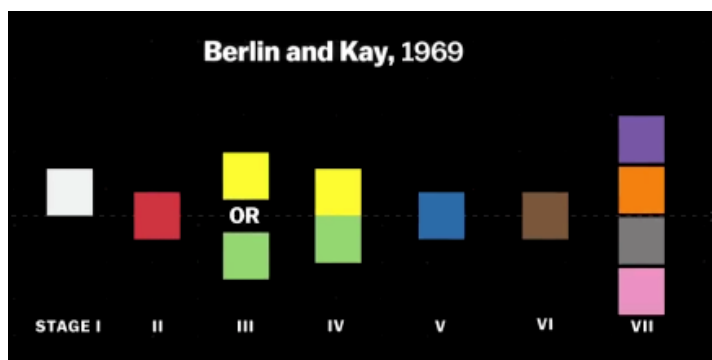


Рис. 2. Графическое изображение эволюции обозначения цветов в лексике английского языка по Берлину и Кею

Fig. 2. Graphic chart of evolution of color designations according to the theory of Berlin and Kay

В 1999 г. Пол Кей и Луиза Меффи опубликовали статью «Появление цвета и возникновение и эволюция лексикона основных цветов» [Kay, Maffi, 1999], в которой описывается серия поправок к теории на основе данных, полученных при реализации проекта «Мировой обзор цвета (World Color Survey – WCS)». Этот проект был предпринят с целью подтвердить, опровергнуть или, что вероятнее всего, модифицировать два основных положения теории Берлина – Кея: (1) наличие универсальных кросс-лингвистических закономерностей именования цветов и (2) фиксированный порядок возникновения системы цветовых терминов. В ходе реализации проекта в базу данных были собраны обозначения цветов из 110 языков. Кроме данных WCS, известна статья Стивена Левинсона, изучавшего вопрос на материале языка Yéî Dnye (Новая Гвинея, Папуа) [Levinson, 2000]. В этой работе отмечены некоторые отличия систем именования цветов от большинства закономерностей, выявленных WCS.

Мы высказали предположение о том, что между последовательностью стадий эволюции языка (по схеме Берлина – Кея, 1969), на которых появляются прилагательные цвета, и частотой их употребления в современном языке имеется корреляция.

Частота употребления базовых прилагательных цвета в корпусах НКРЯ

В данной части исследования мы хотим проверить предположение о том, что порядок появления 11 базовых «цветных» прилагательных (категорий по Берлину и Кею) в диахронии находит отражение в частоте их употребления в синхронном срезе языка, в данном случае русского.

Для подсчета частот мы выбрали корпуса НКРЯ, в первую очередь, основной, а также поэтический, газетный и устный (табл. 1). Дополнительно мы приводим ipm отобранных прилагательных по Новому частотному словарю русской лексики (НЧСРЛ) О. Н. Ляшевской и С. А. Шарова, созданному на базе основного корпуса НКРЯ 2007 г. [Ляшевская, Шаров, 2009].

Рассмотрим порядок снижения частотности прилагательных в корпусах НКРЯ (табл. 2). Как видим, из 11 прилагательных в четырех корпусах (включая НЧСРЛ) самым частотным оказывается «белый», в одном – «чёрный». Во всех пяти корпусах три прилагательных «белый», «чёрный» и «красный» составляют тройку самых частотных. Наименее частотными оказываются прилагательные «розовый», «коричневый», «оранжевый», «фиолетовый». Данные не позволяют говорить о полной корреляции убывания частоты употребления базовых «цветных» прилагательных в корпусах русского языка с эволюционной схемой Берлина и Кея, но, тем не менее, значительная степень корреляция, на наш взгляд, имеется.

Таблица 1

Частотность базовых прилагательных цвета в корпусах НКРЯ

Table 1

Frequencies of the basic adjectives of color in the corpora of the Russian National Corpus (RNC)

Лемма	ОК		ПК		ГК		УК		НЧСРЛ ipm
	c/y	ipm	c/y	ipm	c/y	ipm	c/y	ipm	
Белый	111 109	392,01	10 730	978,37	54 079	236,65	2 115	174,60	339,6
Чёрный	106 635	376,23	9 556	871,33	42 210	184,71	2 286	188,72	337,5
Красный	86 762	306,11	5 582	508,97	47 301	206,99	1 958	161,64	240,5
Синий	29 371	103,63	4 951	451,44	8 722	38,17	450	37,15	82,8
Зелёный	37 891	133,69	4 871	444,14	17 006	74,42	854	70,50	124,4
Голубой	23 412	100,30	4 221	384,88	13 967	61,11	441	36,41	71,5
Серый	38 297	135,12	2 692	245,46	7 753	33,93	453	37,40	96,1
Жёлтый	26 833	94,67	2 423	220,93	8 445	36,95	388	32,03	78,1
Розовый	16 436	57,99	1 883	171,69	3 938	17,23	216	17,83	49,3
Оранжевый	3 944	13,92	260	23,71	6 825	29,87	67	5,53	15,60
Коричневый	6 960	24,56	181	16,50	1 491	6,52	110	9,08	25,5
Фиолетовый	3 325	11,73	125	11,40	727	3,18	57	4,71	12,4
Среднее значение ipm		145,83		360,74		77,49		64,63	122,7

Условные обозначения (Legend):

ПК – поэтический корпус (poetic corpus)

ОК – основной корпус (basic corpus)

ГК – газетный корпус (newspaper corpus)

УК – устный корпус (oral corpus)

НЧСРЛ – Новый частотный словарь русской лексики (Frequency dictionary of the modern Russian language (FDMRL))

c/y – число словоупотреблений (число вхождений при поиске в корпусе) (number of matches (results of search in the corpus))

ipm – instances per million (относительная частота, умноженная на миллион)

Таблица 2

Порядок убывания частоты прилагательных цвета в четырех корпусах НКРЯ и НЧСРЛ
в сопоставлении с эволюционной схемой Берлина и Кея

Table 2

Order of decreasing frequencies of color adjectives in four corpora of the RNC
and in the Frequency dictionary of the modern Russian language (FDMRL) as compared with the evolutionary scheme of Berlin-Kay

Корпус (Corpus)	Ранг (Rank)										
	1	2	3	4	5	6	7	8	9	10	11
Поэтический (poetry)	белый	чёрный	красный	синий	зелёный	серый	жёлтый	розовый	оранжевый	коричневый	фиолетовый
Основной (basic)	белый	чёрный	красный	серый	зелёный	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
Газетный (newspaper)	белый	красный	чёрный	зелёный	синий	жёлтый	серый	оранжевый	розовый	коричневый	фиолетовый
Устный (oral)	чёрный	белый	красный	зелёный	серый	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
НЧСРЛ (FDMRL)	белый	чёрный	красный	зелёный	серый	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
Стадия (Phase)											
	I		II	III–IV		V	VI		VII		
Берлин – Кей (Berlin – Kay)	белый	чёрный	красный	зелёный	жёлтый	синий	коричневый	серый	розовый	оранжевый	фиолетовый

Три самых частотных прилагательных соответствуют первой и второй стадиям схемы. Правда, полное соответствие наблюдается не во всех корпусах. В одном случае – газетный корпус – прилагательное «красный» оказывается вторым по частотности, а прилагательное «чёрный» третьим. Если рассматривать седьмую стадию, то частотность трех прилагательных из четырех коррелирует со схемой. Полное отсутствие корреляции со схемой наблюдается в частотном поведении прилагательных «серый», «жёлтый» и «коричневый». Частотность первого из них занимает 4–6 ранги в разных корпусах, что, по нашему предположению, соответствует III–V стадиям эволюционной схемы. Напротив, прилагательное «жёлтый» «отстает» от своей позиции в эволюционной схеме. Его частотность в корпусах соответствует V–VI стадиям, а по Берлину – Кею прилагательное «жёлтый» появляется на III–IV стадиях. Частотность прилагательного «коричневый» соответствует, по нашим результатам, VII стадии, а по схеме – VI. Частотность прилагательного «синий» в трех случаях коррелирует со схемой – в основном, устном корпусах и в НЧСРЛ.

Таким образом, мы можем утверждать, что между частотностью прилагательных цвета в разных корпусах и эволюционной схемой Берлина – Кея наблюдается частичная корреляция. В естественном языке, особенно в семантике, вряд ли можно найти явление, которое описывается единой универсальной формулой. Обратим внимание, что порядок убывания частотности прилагательных цвета в разных корпусах хоть и отличается, но отличия в целом не слишком велики, что говорит о наличии закономерностей в частотном распределении прилагательных этой группы.

Корреляция между эволюционной теорией обозначения цвета Берлина и Кея и частотой встречаемости прилагательных цвета в описаниях коллекций Государственного Эрмитажа

Мы предположили, что в описаниях произведений искусства прилагательные цвета будут встречаться чаще в своем основном значении. Поисковая система Эрмитажа не является лингвистическим корпусом. Мы не можем сравнивать относительную частоту встречаемости лексических единиц (irm), поскольку у нас нет данных об общем числе словоупотреблений в системе. Однако можно и интересно сопоставить порядок убывания абсолютного числа найденных описаний по запросам на «цветные» прилагательные. Приведем список результатов поиска по запросу на прилагательные цвета в БД Государственного Эрмитажа (табл. 3). Материал отсортирован по убыванию числа найденных описаний.

Обратим внимание, что малочастотное в корпусах НКРЯ прилагательное «коричневый» в данном случае оказывается третьим по частотности, а частота прилагательного «белый» (самого частотного в корпусах НКРЯ) намного ниже. Как и в приведенных выше данных, можно, пожалуй, говорить о некоторой корреляции со схемой Берлина и Кея. При этом корреляция, против нашего ожидания, выражена слабее, чем в текстах из НКРЯ. Возможно, это особенность коллекций Эрмитажа, возможно, цветовая гамма «мира вещей» чем-то принципиально отличается от цветовой картины «мира языка».

Частотное поведение прилагательных цвета в поэтических текстах

Перейдем к описанию поэтических текстов. Сравним частотное поведение прилагательных цвета в поэтических текстах с текстами других типов.

Уже при анализе данных табл. 1 обращает на себя внимание тот факт, что в целом irm прилагательных цвета в поэтическом корпусе превышает irm прилагательных в остальных трех корпусах, что хорошо видно по средним значениям irm. Лишь в двух случаях (прилагательные «коричневый» и «фиолетовый») значения irm в основном корпусе незначительно превышают irm в поэтическом корпусе, и в еще одном случае irm прилагательного «оранже-

вый» в газетном корпусе превышает частотность этого прилагательного в поэтическом корпусе.

Повышенная частотность прилагательных цвета в поэтическом корпусе особенно наглядно видна, если взять в рассмотрение все «цветные» прилагательные во всех корпусах (табл. 4, условные обозначения см. в табл. 1). Поиск прилагательных цвета производился с использованием семантического тега «цвет».

Таблица 3

Результаты поиска по прилагательным цвета
в поисковой системе Эрмитажа

Table 3

Representation of color adjectives
in the Information System of the State Hermitage Museum

№	Лемма	Число описаний, содержащих данное прилагательное
1	Чёрный	1 186
2	Красный	1 087
3	Коричневый	966
4	Зелёный	917
5	Белый	834
6	Серый	819
7	Жёлтый	367
8	Голубой	326
9	Синий	320
10	Розовый	219
11	Фиолетовый	43

Таблица 4

Частотность прилагательных, имеющих тег семантической разметки «цвет»,
в четырех корпусах НКРЯ

Table 4

Frequency of adjectives supplied with the tag “color”
in four corpora of the RNC

ПК		ОК		ГК		УК	
с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
85 286	7776,48	77 1638	2722,48	280 630	1228,02	12 682	1046,93

Анализ данных табл. 4 подтверждает сделанные ранее наблюдения, что частотность прилагательных цвета в текстах поэтического корпуса выражена выше, чем в остальных трех корпусах.

Методика исследования

Для того чтобы получить статистические данные, характеризующие употребление прилагательных цвета в русском поэтическом языке и в языке отдельных авторов, была выполнена следующая работа.

1. В поэтическом корпусе НКРЯ на первом шаге созданы авторские подкорпуса из текстов 15 авторов. Были выбраны наиболее известные авторы XIX–XX вв. из числа представленных в НКРЯ. В созданных корпусах производится поиск прилагательных цвета по семантическому тегу «цвет». В конце каждой страницы результатов поиска корпусный менеджер НКРЯ выводит таблицы частоты встречаемости заданных в запросе лексических единиц в подмножестве документов данной страницы результатов, отдельно для словоформ и лемм. Нас интересует таблица частоты лемм.

2. Таблицы словоформ и лемм указывают частоты лексем только по данной странице результатов, а таких страниц может оказаться немало. Поэтому каждая таблица лемм для каждой очередной страницы результатов поиска переносится в книгу MS Excel, где данные суммируются средствами Excel («Сводная таблица»), т. е. для каждой леммы подсчитывается суммарное количество словоупотреблений.

3. Средствами Excel для каждой леммы данного авторского подкорпуса подсчитываются дополнительные параметры (величина *ipm* по отношению к объему всего авторского подкорпуса и доля данного прилагательного (леммы) среди всех прилагательных цвета в текстах у данного автора) (рис. 3).

4. Наконец, таблицы по отдельным авторам сливаются в общую таблицу, в которой выполняются сортировки и анализ данных по различным критериям.

5. Также для вычисления показателей использования прилагательных цвета каждым автором запоминаются (вычисляются) объем каждого авторского корпуса, число цветных прилагательных, число всех прилагательных в этом корпусе.

1	Лемма	автор	Число с/у	IPM	%
2	белый	Блок	225	1982,5	13,69
3	темный	Блок	193	1700,55	11,74
4	черный	Блок	182	1603,62	11,07
5	светлый	Блок	159	1400,97	9,67
6	красный	Блок	131	1154,26	7,97
7	синий	Блок	118	1039,71	7,18
8	бледный	Блок	103	907,54	6,27
9	голубой	Блок	97	854,68	5,9
10	зеленый	Блок	73	643,21	4,44
11	серый	Блок	52	458,18	3,16
12	алый	Блок	41	361,26	2,49
13	розовый	Блок	41	361,26	2,49
14	лазурный	Блок	38	334,82	2,31
15	желтый	Блок	33	290,77	2,01
16	сумрачный	Блок	26	229,09	1,58
17	пестрый	Блок	17	149,79	1,03
18	золотистый	Блок	16	140,98	0,97

Рис. 3. Фрагмент таблицы частотности прилагательных цвета в текстах А. А. Блока

Fig. 3. Fragment of the frequency table of color adjectives in the texts by Alexander Block

В поэтическом корпусе НКРЯ отсутствуют тексты многих известных русских поэтов второй половины XX в., таких как Евтушенко, Вознесенский, Ахмадулина, Высоцкий, Окуджава, Матвеева и др. В связи с этим на базе различных массивов текстов, взятых из Интернета (Библиотека Максима Мошкова, персональные сайты поэтов) средствами менеджера Sketch Engine были созданы авторские корпуса текстов других русских поэтов (еще 21 автор). В этой системе нет семантической разметки, поэтому в этих новых корпусах поиск производился по сводному списку прилагательным цвета, полученному на основе НКРЯ.

Далее по корпусу каждого автора выполняются операции 3–5, описанные выше.

Частотное поведение прилагательных цвета в поэтических текстах отдельных авторов

В конечном счете, все таблицы для 36 авторов были сведены в единую таблицу Excel общим объемом строк 1 571. Таблица 5 представляет собой фрагмент сводной таблицы для 50 лемм с наибольшим *ipm* в корпусах отдельных авторов (сортировка по *ipm*).

Таблица 5

Фрагмент сводной таблицы частотности прилагательных,
обозначающих цвет, в текстах 36 авторов (часть 1)

Table 5

Fragment of the summary frequency table of color adjectives
in the texts by 36 authors (part 1)

Ранг и лемма	Автор	Число с/у	<i>ipm</i>	Доля, %
1. Белый	Хлебников	223	4041,10	22,57
2. Черный	Хлебников	149	2700,11	15,08
3. Белый	Ахматова	122	2159,14	18,63
4. Черный	Набоков	130	2079,87	13,50
5. Черный	Ахматова	117	2070,65	17,86
6. Синий	Хлебников	111	2011,49	11,23
7. Белый	Блок	225	1982,50	13,69
8. Белый	Ахмадулина	103	1842,64	28,77
9. Белый	Вознесенский	52	1822,45	25,00
10. Синий	Есенин	105	1789,61	14,73
11. Белый	Есенин	101	1721,43	14,17
12. Темный	Блок	193	1700,55	11,74
13. Белый	Тарковский	63	1696,01	18,81
14. Черный	Мандельштам	102	1689,05	16,59
15. Светлый	Бальмонт	179	1655,90	13,81
16. Белый	Набоков	102	1631,90	10,59
17. Черный	Блок	182	1603,62	11,07
18. Темный	Хлебников	88	1594,69	8,91
19. Белый	Бальмонт	166	1535,64	12,81
20. Белый	Кривулин	44	1513,48	18,64
21. Белый	Заболоцкий	31	1464,27	21,09

Окончание табл. 5

Ранг и лемма	Автор	Число с/у	ipm	Доля, %
22. Черный	Тарковский	54	1453,72	16,12
23. Черный	Вознесенский	41	1436,93	19,71
24. Красный	Цветаева	245	1435,59	15,88
25. Белый	Цветаева	240	1406,30	15,55
26. Светлый	Блок	159	1400,97	9,67
27. Белый	Рубцов	35	1353,70	15,56
28. Черный	Есенин	78	1329,42	10,94
29. Светлый	Тютчев	45	1325,24	24,46
30. Темный	Рубцов	34	1315,03	15,11
31. Черный	Цветаева	222	1300,82	14,39
32. Черный	Кривулин	36	1238,3	15,25
33. Зеленый	Хлебников	68	1232,26	6,88
34. Златой	Державин	115	1229,91	18,34
35. Голубой	Хлебников	67	1214,14	6,78
36. Голубой	Вертинский	16	1196,53	16,00
37. Белый	Аронзон	34	1179,33	24,29
38. Красный	Блок	131	1154,26	7,97
39. Красный	Маяковский	156	1153,09	26,17
40. Бледный	Бальмонт	122	1128,61	9,41
41. Голубой	Есенин	65	1107,85	9,12
42. Черный	Пастернак	86	1104,43	21,39
43. Зеленый	Тарковский	41	1103,75	12,24
44. Белый	Пастернак	81	1040,22	20,15
45. Синий	Блок	118	1039,71	7,18
46. Темный	Мандельштам	62	1026,68	10,08
47. Черный	Симонов	84	1020,97	21,65
48. Светлый	Рубцов	26	1005,61	11,56
49. Темный	Бальмонт	107	989,84	8,26
50. Красный	Есенин	58	988,55	8,13

Из 25 наиболее частотных словоупотреблений (ранги 1–25) 12 прилагательных «белый», 7 – «черный», 2 – «синий», 1 – «красный». Еще три словоупотребления: прилагательные «темный» и «светлый», являются в нашей терминологии «квазицветами» (т. е. в семантической разметке НКРЯ имеют тег «цвет» и потому входят в массив обнаруженных слов). Здесь можно заметить некоторую корреляцию между частотностью прилагательных цвета и стадиями эволюции цветообозначений. В данном частотном ранге преобладают слова, обозначающие цвета I эволюционной стадии. На 24-м месте появляется «красный» – цвет, отнесенный Берлином и Кеем к стадии II. Несколько нарушает корреляцию прилагательное «синий» у Есенина и Хлебникова, но это, по-видимому, особенности авторских идиолектов.

В следующем фрагменте (ранги 26–50) преобладает прилагательное «чёрный» (5 позиций), появляются цвета, отсутствовавшие в первом фрагменте таблицы: «зелёный» (2 раза), «голубой» (3 раза), появилось аналоговое прилагательное «золотой».

Сформируем табл. 6, показывающую порядок появления основных цветовых категорий в сводной таблице (1 571 строка).

Таблица 6
Ранг первого появления основных цветовых категорий
в сводной таблице

Table 6
Frequency ranking of the first appearance of basic color categories
in the summary table

Лемма	Автор	Ранг первого появления в списке	ipm
Белый	Хлебников	1	4041,1
Черный	Хлебников	2	2700,11
Синий	Хлебников	6	2011,49
Красный	Цветаева	23	1435,59
Зеленый	Хлебников	33	1232,26
Голубой	Хлебников	35	1214,14
Желтый	Ахматова	82	725,61
Серый	Хлебников	101	634,25
Розовый	Набоков	133	511,97
Оранжевый	Вознесенский	252	280,38
Коричневый	Мандельштам	582	66,24
Фиолетовый	Набоков	593	64,00

Можно говорить о некоторой (неполной) корреляции между рангами прилагательных в табл. 6 и их принадлежностью к эволюционным стадиям. Обращает внимание, что из 12 прилагательных 6 впервые встречаются в текстах В. Хлебникова. Это, по-видимому, говорит о каких-то существенных особенностях текстов этого автора. Однако анализ текстов отдельных авторов не входит в нашу задачу.

Классификация имен прилагательных цвета

Как мы уже говорили, таблицы 36 авторов были слиты в единую таблицу (1 571 строка). Для 180 лемм, представленных в этой общей таблице, было подсчитано число авторов, у которых каждая из лемм встретила (табл. 7).

Для обозначения цвета естественный язык использует разные части речи. В данной статье рассматриваются только прилагательные. Но и для прилагательных цвета существуют различные классификации [Герасимов, 1969; 1978; Василевич и др., 2005]. Рискнем предложить еще одну. Как часто бывает, в рамках одной классификации присутствуют разные основания выделения классов. Этот недостаток свойственен и нашей. Но главная наша задача – дать статистическую картину широкой или ограниченной распространенности различных прилагательных и их типов в текстах разных поэтов и в поэтическом языке в целом.

Таблица 7

Наиболее часто встречающиеся
прилагательные цвета

Table 7

Most frequent color adjectives

Лемма	Число авторов
1. Белый	36
2. Голубой	36
3. Желтый	36
4. Зеленый	36
5. Красный	36
6. Светлый	36
7. Синий	36
8. Темный	36
9. Черный	36
10. Бледный	35
11. Серый	34
12. Розовый	33
13. Рыжий	31
14. Цветной	31
15. Багровый	30
16. Пестрый	30
17. Алый	29
18. Мутный	29
19. Белоснежный	28
20. Золотистый	26
21. Лиловый	24
22. Сизый	24
23. Сумрачный	24
24. Серебристый	23
25. Багряный	21

Вот предлагаемая нами классификация.

1. Прилагательные, означающие собственно цвета
2. Прилагательные, означающие квазицвета
3. Прилагательные, означающие специфические цвета определенных объектов
4. Прилагательные цвета-дериваты
 - 4.1. Диминутивы
 - 4.2. Прилагательные подобия
 - 4.3. Составные прилагательные цвета
5. Аналоговые цвета

Первая группа – собственно цвета. Таких слов мы обнаружили 29 из 180 (15,7 %) (табл. 8).

Таблица 8

Прилагательные, означающие собственно цвет

Table 8

Adjectives, meaning color per se

Лемма	Число авторов	Лемма	Число авторов
1. Белый	36	16. Бурый	18
2. Голубой	36	17. Оранжевый	18
3. Желтый	36	18. Пурпурный	16
4. Зеленый	36	19. Коричневый	13
5. Красный	36	20. Лазоревый	13
6. Синий	36	21. Фиолетовый	12
7. Черный	36	22. Пунцовый	10
8. Серый	34	23. Пурпуровый	10
9. Розовый	33	24. Сивый	10
10. Багровый	30	25. Палевый	9
11. Алый	29	26. Бордовый	5
12. Лиловый	24	27. Червлёный	5
13. Сизый	24	28. Бежевый	3
14. Багряный	21	29. Бордовый	3
15. Лазурный	20	Среднее число авторов	21,1

В эту группу вошли все 12 основных цветов. В пределах одного частотного ранга слова отсортированы по алфавиту. И в этом случае можно говорить о некоторой корреляции между степенями эволюции по Берлину и Кею.

Среди прилагательных этой группы встречаются синонимы. Вряд ли можно сделать уверенное различие между значениями прилагательных «красный», «алый», «пунцовый», «бордовый» или «фиолетовый», «лиловый», а также, например, «бурый» и «коричневый». Подтвердим этот тезис с помощью поисковой системы Государственного Эрмитажа – вот некоторые образцы синонимичных цветов (рис. 4–6).

Представленные образцы демонстрируют похожесть и различия между цветами, обозначаемыми одним прилагательным, и особенности каждого оттенка, которые, кстати говоря, трудно поддаются вербализации, особенно с точки зрения рядового носителя языка. Так, цвета предметов на изображениях, полученных по запросу «красный», субъективно воспринимаются как различные, что, возможно, обусловлено различием фактуры материалов. С другой стороны, цвета предметов на изображениях, полученных по запросу «алый» и «бордовый», субъективно воспринимаются скорее как сходные, несмотря на различную фактуру материала.

Нет сомнения, что авторы описаний коллекций Эрмитажа, являясь профессиональными искусствоведами, дают предельно точные наименования цветов. В этом случае, видимо, нужно говорить об особенностях языков для специальных целей. Кроме того, прилагательное «красный» обозначает не просто цвет, а широкую цветовую категорию, содержащую большое число значений – оттенков красного цвета. А категории «алый» и «бордовый» значительно уже, индивидуальнее.



Рис. 4. Поисковый термин «красный» (результат поиска – 1 087 описаний):

a – фрагмент: Эннер, Жан-Жак (1829–1905), Этюд женщины в **красном**, Франция, начало 1890-х гг., холст, масло, 55 × 38 см; *б* – обломок фриза из **красного** мрамора с гравировкой, Урарту, VIII–VII вв. до н. э., мрамор, 6 × 2,2 см; *в* – ткань с крупным стилизованным растительным узором по **красному** фону, Испания (?), середина XVII в., шелк, серебряная позолоченная нить, 55 × 53 см

Fig. 4. Search term “red”:

a – fragment: Henner Jean-Jacques (1829–1905), Study of woman in **red** dress, France, beginning of 1890-s; oil, canvas 55 × 38 cm; *б* – piece of frieze made of **red** marble with engraving; Urartu; 9th – 7th century B.C.; marble; 6 × 2,2 cm; *в* – textile with large-sized floral pattern against **red** background; Spain (?), 17th century; silk; silver gilded thread; 55 × 53 cm



Рис. 5. Поисковый термин «алый» (результат поиска 46 описаний):

a – фрагмент: Икона – Святой Николай, место создания Россия, Новгородская школа, XIII–XIV вв., дерево, темпера, 107 × 82 × 3 см (из аннотации: «**алый (киноварный)** фон (воплощение негаснущего света вечности), характерный для новгородской школы иконописи...»); *б* – мундирное платье Екатерины II, по форме армейской пехоты, **шелк алый**, рукава и спина зеленого шелка, подкладка шелковая (на лифе – белая, на юбке – алая), галуны и пуговицы золоченые (сзади и спереди разрезы для езды верхом по-мужски), 1772 г., шелк

Fig. 5. Search term “scarlet”:

a – fragment: Icon Saint Nicolas; Russia; the Novgorodian school; 13th – 14th century; wood; tempera; 107 × 82 × 3 cm (from annotation: “**scarlet (cinnabar)** background (symbol of inextinguishable light of eternity) typical of the Novgorodian school”); *б* – uniform dress of Catharina II, to the uniform of army infantry, **scarlet silk**, sleeves and back of green silk, backing of silk (white on the bodice, scarlet on the skirt) gallons and buttons gilded, 1772; silk



а



б

Рис. 6. Поисковый термин «бордовый» (результат поиска 8 описаний):

а – фрагмент: Винтерхальтер, Франц Ксавер. 1806–1873, Портрет С. П. Нарышкиной, Франция, 1858 г., холст, масло 150 × 114 см (из аннотации: «...изящный жест холеных рук, смелое сочетание темно-**бордовых** и лимонно-желтых тонов одежды придают облику модели претенциозную элегантность»); б – передвижная отопительная печь, Франция, конец XIX в., чугун, эмаль, металл, слюда литые, ковка, чеканка 72 × 68 × 33 см (Печь, покрытая эмалью бордового цвета, по форме и рельефному декору относится к эпохе модерна)

Fig. 6. Search term “vinous”:

а – fragment: Winterhalter Franz Xaver (1806–1873), Portrait of S. P. Naryshkina; France 1858, oil, canvas (from annotation: “daring combination of dark **vinous** and lemon yellow colors of clothes enhance to the image pretentious elegance...”); б – movable heating stove; France; end of 19th century, cast, iron, metal, mica, molding, forging, embossed work; 72 × 68 × 33 cm (The stove covered with enamel of **vinous** color...)

Вторую группу слов мы условно назвали «квазицвета» (19 лемм, 10,6 %) (табл. 9).

Таблица 9

Прилагательные «квазицвета»

Table 9

“Quasi colors” adjectives

Лемма	Число авторов
1. Светлый	36
2. Тёмный	36
3. Бледный	35
4. Цветной	31
5. Пёстрый	30
6. Мутный	29
7. Сумрачный	24
8. Разноцветный	21
9. Бесцветный	20
10. Клетчатый	10
11. Самоцветный	10

Окончание табл. 9

Лемма	Число авторов
12. Пятнистый	9
13. Цветастый	8
14. Линялый	6
15. Трёхцветный	3
16. Двухцветный	1
17. Крапчатый	1
18. Одноцветный	1
19. Потускнелый	1
Среднее число авторов	16,4

В число прилагательных, обозначающих цвета (т. е. имеющих тег «цвет» в НКРЯ), попадают слова «тёмный», «светлый», «бледный», «пёстрый», «сумрачный», «цветной», которые, собственно, означают не цвет, а его интенсивность, оттенок (тёмный, светлый, бледный), просто наличие цвета (цветной) – как оппозиция черному и белому, множество цветов (пестрый) или даже настроение (сумрачный), которое могут вызвать цвет или комбинация цветов. Эти прилагательные описывают различные визуальные характеристики объекта – кроме, пожалуй, его формы.

Часть этих слов не обозначает, а характеризует цвет. Некоторые из них образуют сложные прилагательные типа «бледно-зелёный», «светло-серый», «тёмно-красный» и т. д. (см. далее, табл. 15).

Как видно из табл. 9, представленность в текстах некоторых из этих прилагательных достаточно высока. Особенно это касается прилагательных «светлый» и «тёмный». Некоторые авторы утверждают, что именно смена светлого и темного времени суток обусловила появление цветообозначений «белый» и «чёрный» на самой ранней стадии эволюции (см. рис. 2).

Сравним результаты поиска этого вида прилагательных в разных корпусах НКРЯ (табл. 10).

По результатам анализа отметим следующее.

1. Средний *ipm* названий квазицветов ниже, чем названий цветов (см. табл. 2):

- по поэтическому корпусу в 3,4 раза;
- по основному корпусу в 5,1 раза;
- по газетному корпусу в 11,4 раза;
- по устному корпусу 8,4 раза.

2. Как и в случае с основными цветами среднее *ipm* квазицветов в поэтическом корпусе значительно выше, чем в трех других корпусах.

Третью группу составляют прилагательные, которые относятся к определенному, чаще всего одушевленному, предмету. Это может быть цвет волос или глаз, масть или окрас шерсти животного и т. п. (28 лемм, 15,5 %) (табл. 11, 12).

Тенденция соотношений *ipm* у этого типа прилагательных примерно та же, что у прилагательных собственно цвета и прилагательных «квазицветов», т. е. среднее *ipm* в поэтическом корпусе значительно выше, чем в остальных трех корпусах. При этом среднее *ipm* этого типа прилагательных во всех корпусах ниже, чем для прилагательных собственно цвета и квазицветов.

Таблица 10

Сравнение частотности прилагательных «квазицвета» в текстах четырех корпусов НКРЯ

Table 10

Comparison of frequency of the “quasi colors” adjectives in texts of four corpora the RNC

Лемма	ПК		ОК		ГК		УК	
	с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
Тёмный	7310	666,53	60395	213,08	8899	38,94	820	67,69
Светлый	6631	604,62	33797	119,24	8097	35,43	515	42,51
Бледный	3910	356,52	20243	71,42	1482	6,49	90	7,43
Пёстрый	1100	100,30	7581	26,75	913	4,00	34	2,81
Сумрачный	942	85,89	2972	10,49	180	0,79	11	0,91
Мутный	835	76,14	7116	25,11	964	4,22	41	3,38
Цветной	585	53,34	9060	31,97	5918	25,90	166	13,70
Разноцветный	276	25,17	5272	18,60	1508	6,60	24	1,98
Бесцветный	166	15,14	2580	9,10	250	1,09	17	1,40
Самоцветный	86	7,84	289	1,02	14	0,06	3	0,25
Пятнистый	65	5,93	964	3,40	249	1,09	8	0,66
Клетчатый	46	4,19	1881	6,64	334	1,46	27	2,23
Линялый	41	3,74	306	1,08	41	0,18	2	0,17
Трёхцветный	41	3,74	520	1,83	268	1,17	9	0,74
Цветастый	14	1,28	623	2,20	260	1,14	4	0,33
Одноцветный	14	1,28	256	0,90	39	0,17	2	0,17
Потускнелый	5	0,46	12	0,04	0	0,00	0	0,00
Крапчатый	4	0,36	122	0,43	9	0,04	2	0,17
Двухцветный	3	0,27	124	0,44	88	0,39	4	0,33
Среднее значение ipm		105,9		28,62		6,8		7,73
Среднее значение для прилагательного собственно цвета (см. табл. 1)		360,74		145,83		77,48		64,63

Таблица 11

Прилагательные,
относящиеся к определенному предмету

Table 11

Object-related adjectives

Лемма	Число авторов
1. Рыжий	31
2. Румяный	21
3. Русый	21
4. Гнедой	15
5. Карий	14
6. Пегий	10
7. Вороной	9
8. Поседель	9
9. Рыжеватый	7
10. Седенький	7
11. Сизокрылый	7
12. Буланный	6
13. Порыжелый	4
14. Рыжеволосый	4
15. Рыженький	4
16. Рябоватый	3
17. Саврасый	3
18. Чалый	3
19. Черноволосый	3
20. Черномазый	3
21. Седоватый	2
22. Соловый	2
23. Каурый	1
24. Муругий	1
25. Мухортый	1
26. Рыжебородый	1
27. Рябенький	1
28. Чубарый	1
Среднее число авторов	6,9

Таблица 12

Частотность прилагательных, относящихся к определенному предмету, в четырех корпусах НКРЯ

Table 12

Frequency of the object-related adjectives in four corpora of the RNC

Лемма	ПК		ОК		ГК		УК	
	с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
Румяный	992	90,45	4159	14,67	442	1,93	23	1,90
Рыжий	931	84,89	14326	50,54	1780	7,79	234	19,32
Русый	311	28,36	2919	10,30	1326	5,80	29	2,39
Карий	136	12,40	2164	7,63	720	3,15	28	2,31
Вороной	117	10,67	1467	5,18	246	1,08	32	2,64
Пегий	113	10,30	772	2,72	22	0,10	2	0,17
Рыжебородый	95	8,66	450	1,59	11	0,05	1	0,08
Гнедой	70	6,38	907	3,20	95	0,42	18	1,49
Поседлый	60	5,47	459	1,62	80	0,35	7	0,58
Буланый	39	3,56	221	0,78	4	0,02	1	0,08
Соловый	38	3,46	197	0,70	15	0,07	0	0,00
Сизокрылый	33	3,01	50	0,18	10	0,04	1	0,08
Рыжеватый	31	2,83	1242	4,38	59	0,26	4	0,33
Седенький	24	2,19	477	1,68	25	0,11	3	0,25
Рыженький	22	2,01	671	2,37	71	0,31	21	1,73
Чалый	21	1,91	304	1,07	221	0,97	0	0,00
Черноволосый	20	1,82	1086	3,83	130	0,57	6	0,50
Саврасый	18	1,64	121	0,43	2	0,01	0	0,00
Рыжеволосый	17	1,55	375	1,32	210	0,92	0	0,00
Черномазый	16	1,46	410	1,45	19	0,08	9	0,74
Каурый	13	1,19	132	0,47	12	0,05	2	0,17
Порыжелый	12	1,09	219	0,77	0	0,00	0	0,00
Седоватый	9	0,82	410	1,45	16	0,07	2	0,17
Рябоватый	6	0,55	218	0,77	6	0,03	0	0,00
Муругий	6	0,55	27	0,10	0	0,00	0	0,00
Рябенький	3	0,27	89	0,31	9	0,04	3	0,25
Чубарый	3	0,27	56	0,20	3	0,01	0	0,00
Мухортый	1	0,09	29	0,10	0	0,00	0	0,00
Среднее значение IPM		10,28		4,28		0,86		1,26

Некоторая часть этих слов также не означает собственно цвета, но цветовую неоднородность, наличие пятен («пегий», «рябенький», «чубарый», «чалый»), в других случаях указывается на результат некоторого процесса («поседелый», «порыжелый»). Иногда образуется прилагательное, содержащее одновременно объект и цветовую характеристику, причем и то, и другое присуще одному лицу (рыжебородый, черноволосый).

Значительное место в этой группе занимает масть лошадей или окрас собак. Известно, что масть или окрас – это не столько цвет, сколько сочетание цветов, генетически детерминированный тип распределения пигментов в шерсти животного.

Четвертая группа прилагательных цвета, дериваты (83 леммы, 46 %), может быть разделена на три подгруппы, отражающие некоторые формы прилагательных. Так, в первую подгруппу четвертой группы вошли прилагательные-диминытивы (13 лемм, 7,2 %) (табл. 13).

Таблица 13

Прилагательные-диминытивы

Table 13

Adjectives diminutive

Лемма	Число авторов
1. Беленький	13
2. Серенький	13
3. Синенький	9
4. Аленький	6
5. Голубенький	5
6. Чёрненький	5
7. Зелёненький	3
8. Жёлтенький	2
9. Красненький	2
10. Бледненький	1
11. Темнѐхонький	1
12. Чернѐхонький	1
13. Чернѐшенький	1
Среднее число авторов	4,8

В их число вошли не только собственно прилагательные цвета, но и прилагательные «темнѐхонький» и «бледненький», исходные формы которых «тёмный» и «бледный» мы отнесли к квазицветам. В табл. 12 мы находим три прилагательных, от которых образованы диминытивы, – «рыженький», «рябенький» и «седенький». Здесь нам пришлось столкнуться с типичной проблемой любой классификации – один объект имеет два значимых классификационных признака.

Интересно, что в поисковой системе Эрмитажа цветоозначающие прилагательные-диминытивы отсутствуют. Это связано с тем, что эта форма придает слову экспрессивность, ненужную в научных описаниях произведений искусства.

Прилагательные второй подгруппы условно назовем прилагательными подобия (10 лемм, 5,5 %) (табл. 14).

Прилагательные подобия

Таблица 14

Adjectives of likeness

Table 14

Лемма	Число авторов
1. Розоватый	13
2. Синеватый	11
3. Зеленоватый	9
4. Красноватый	8
5. Голубоватый	7
6. Лиловатый	6
7. Желтоватый	4
8. Сероватый	4
9. Белесоватый	2
10. Беловатый	1
Среднее число авторов	6,5

В этой форме встретились прилагательные, исходные формы которых мы отнесли к третьей группе (см. табл. 12), – «рябоватый», «рыжеватый» и «седоватый». Обращает внимание, что от этих же исходных слов образованы диминутивы (см. табл. 13).

Третья подгруппа – это сложные прилагательные, имеющие обычно два корня (62 леммы, 35 %) (табл. 15). Они разными способами описывают интенсивность или оттенок цвета.

Число лемм сложных прилагательных значительно превышает число членов других групп и подгрупп. При этом частотность каждой леммы ниже, чем у лемм в остальных группах и подгруппах (среднее число авторов на одну лемму 2,26).

Чаще всего в составные прилагательные этого типа входят комбинации с корнями *тёмно-* (9 лемм), *ярко-* (9 лемм), *чёрно-* (7 лемм), *светло-* (6 лемм), *нежно-* (5 лемм), *бледно-* (4 леммы). Отметим, что прилагательные этой группы присутствуют и в поисковой системе Эрмитажа (табл. 16).

Обращает на себя внимание низкая частотность сложных прилагательных по сравнению с прилагательными основных цветов. Наибольшее число описаний найдено на сочетания с прилагательным «коричневый».

Наконец, пятая группа содержит прилагательные, которые мы называли аналоговыми. Они передают цвета и оттенки через цвета и оттенки хорошо известных предметов. Герасимов [1969] называет их относительными прилагательными. Семантическая разметка НКРЯ плохо учитывает слова такого типа, поэтому при поиске по семантическому признаку в корпусе найдено всего 8 таких прилагательных (5 %) (табл. 17). Однако в текстах НКРЯ, как показывают другие наши исследования, их значительно больше, просто большинству из них не приписан тег «цвет». Следует отметить, что частотность каждого из найденных прилагательных довольно высока.

Таблица 15

Сложные прилагательные цвета

Table 15

Compound adjectives of color

Лемма	Число авторов	Лемма	Число авторов
1. Тёмно-синий	14	32. Красно-синий	1
2. Тёмно-голубой	9	33. Кроваво-красный	1
3. Чёрно-белый	9	34. Медно-зелёный	1
4. Тёмно-зелёный	8	35. Небесно-голубой	1
5. Бледно-голубой	7	36. Нежно-зелёный	1
6. Чёрно-синий	6	37. Нежно-золотой	1
7. Светло-зелёный	5	38. Нежно-розовый	1
8. Ярко-жёлтый	4	39. Потёрто-белый	1
9. Молочно-белый	3	40. Пречёрный-чёрный	1
10. Нежно-белый	3	41. Прозрачно-синий	1
11. Светло-серый	3	42. Розовато-белый	1
12. Светло-синий	3	43. Светло-пегий	1
13. Серо-зелёный	3	44. Светло-сиреневый	1
14. Серо-синий	3	45. Серо-голубой	1
15. Ярко-красный	3	46. Сине-красный	1
16. Бледно-зелёный	2	47. Тёмно-бирюзовый	1
17. Бледно-лиловый	2	48. Тёмно-вишневый	1
18. Медно-красный	2	49. Тёмно-лиловый	1
19. Нежно-голубой	2	50. Чёрно-бархатный	1
20. Светло-голубой	2	51. Чёрно-бурый	1
21. Сине-зелёный	2	52. Чёрно-золотой	1
22. Синий-синий	2	53. Чёрно-красной	1
23. Тёмно-жёлтый	2	54. Чёрно-серебряный	1
24. Тёмно-красный	2	55. Электродово-синий	1
25. Тёмно-серый	2	56. Ярко-белые	1
26. Бледно-розовый	1	57. Ярко-голубой	1
27. Буйно-красный	1	58. Ярко-голубой	1
28. Грязно-белый	1	59. Ярко-зелёный	1
29. Грязно-розовый	1	60. Ярко-золотистый	1
30. Жёлто-зелёный	1	61. Ярко-оранжевый	1
31. Зеленовато-голубой	1	62. Ярко-розовый	1
		Среднее число авторов	2,3

Таблица 16

Результаты поиска сложных слов в поисковой системе Эрмитажа

Table 16

Representation of compound color adjectives
in the Information System of the State Hermitage Museum

Лемма	Число описаний, содержащих данное прилагательное	Лемма	Число описаний, содержащих данное прилагательное
1. Тёмно-коричневый	43	15. Ярко-синий	3
2. Светло-коричневый	42	16. Бледно-зелёный	2
3. Тёмно-синий	33	17. Тёмно-голубой	2
4. Тёмно-зелёный	18	18. Тёмно-розовый	2
5. Светло-зелёный	15	19. Бледно-жёлтый	1
6. Ярко-красный	12	20. Бледно-голубой	1
7. Светло-жёлтый	10	21. Бледно-розовый	1
8. Светло-серый	10	22. Бледно-серый	1
9. Тёмно-красный	9	23. Светло-розовый	1
10. Светло-голубой	8	24. Светло-синий	1
11. Светло-фиолетовый	5	25. Ярко-голубой	1
12. Тёмно-серый	4	26. Ярко-зелёный	1
13. Тёмно-фиолетовый	3	27. Ярко-розовый	1
14. Ярко-жёлтый	3	28. Бледно-зелёный	2

Таблица 17

Аналоговые прилагательные цвета

Table 17

Analog adjectives of color

Лемма	Число авторов
1. Золотистый	26
2. Серебристый	23
3. Златой	21
4. Изумрудный	18
5. Бирюзовый	10
6. Дымчатый	8
7. Агатовый	3
8. Аметистовый	2
Среднее число авторов	13,9

Выводы

Исследование прилагательных цвета в поэтических текстах 36 авторов показывает следующее.

1. Число прилагательных, означающих собственно цвет, составляет 29 единиц. Эти прилагательные наиболее частотны.

2. Что касается частотного поведения прилагательных цвета вообще, то здесь можно обозначить два аспекта. С одной стороны, частота встречаемости прилагательных в разной степени коррелирует с эволюционной теорией цветов Берлина и Кея, т. е. прилагательные, относящиеся согласно этим авторам к более ранней стадии эволюции, являются более частотными. С другой стороны, эта корреляция не является абсолютной.

3. Степень корреляции частотности прилагательных цвета в разных текстах НКРЯ и стадий эволюционной модели Берлина и Кея примерно одинакова. Корреляция между моделью Берлина и Кея и абсолютной частотой прилагательных цвета в поисковой системе Эрмитажа выражена слабее.

4. Частота встречаемости прилагательных цвета любого типа значительно выше в поэтических текстах, чем в устных речевых произведениях, газетных статьях и т. п.

5. Отмечаются выраженные различия частотности прилагательных, относящихся к разным классам нашей классификации.

6. Частотность прилагательных собственно цвета всегда выше частотности прилагательных других классов нашей классификации.

7. Сравнение частотности прилагательных цвета по корпусам поэтических текстов разных авторов показывает большой разброс значений ipm .

8. При расположении лемм прилагательных по убыванию частотности (по числу словоупотреблений / лемм в авторских корпусах либо по числу авторов, в текстах которых встречается данная лемма) видно, что основные категории более частотны, и при этом существует очень большое число сложных цветовых терминов единичного употребления (закон Ципфа подтверждается и здесь).

9. Наблюдается корреляция между использованием прилагательных различными авторами.

10. Анализ использования прилагательных цвета (50 самых частотных лемм) в текстах 36 авторов (см. табл. 5) показывает, что в текстах авторов XX в. наблюдается гораздо более высокая частотность прилагательных цвета. Так, в число первых пятидесяти самых больших значений ipm (для пары «лемма – автор») из авторов XIX в. вошли только Державин (златой) и Тютчев (светлый).

11. Прилагательные основных цветов часто имеют большое число синонимов, различия между которыми трудно поддаются вербализации. Однако на основании сравнения образцов, полученных через поисковую систему Эрмитажа, можно предположить, что различия в восприятии основных цветов более выражены, чем различия в восприятии цветов, обозначаемых синонимами.

Заключение

Лингвистические корпуса и компьютерные технологии дают возможность осуществлять исследования и получать результаты на больших текстовых массивах, что прежде, до появления этих средств, было недостижимо.

Наличие в семантической разметке НКРЯ тега «цвет» делает возможным объемное статистическое исследование русских прилагательных, обозначающих цвета. Однако в ходе исследования мы обнаружили, что в ряде случаев этот признак не приписывается цветообозначающему прилагательному (например, аналоговому).

Кроме того, невозможно автоматически установить прямое или переносное значение прилагательного. Существуют, правда, интересные попытки решить такую задачу [Чекменева 2018], но результаты носят пока предварительный характер. Есть и другие нерешенные проблемы семантической разметки, которые в настоящей публикации мы не затрагиваем. В наши дальнейшие планы входит как раз описание и систематизация этих проблем.

Среднее значение *ipm* в текстах наших 36 авторов (7054,74) близко к *ipm* «цветных» прилагательных по поэтическому корпусу НКРЯ в целом (7776,48), что является косвенным доказательством репрезентативности нашего корпуса и позволяет говорить о достоверности полученных результатов.

Достоверно выявлено, что в поэтических текстах наблюдается более высокая частотность прилагательных цвета, чем в текстах других типов. Выборочный просмотр результатов поиска в поэтическом корпусе показал вероятность того, что в поэтическом языке преобладает употребление изучаемых прилагательных в метафорическом значении. Изложенная выше проблема разделения значений не позволяет количественно подтвердить это утверждение.

Обращение к базам данных художественных музеев (не только Эрмитажа, но и других), дает некоторые дополнительные возможности, поскольку в описаниях художественных произведений прилагательные цвета употребляются почти исключительно в прямом значении.

Список литературы / References

- Василевич А. П., Кузнецова С. Н., Мищенко С. С.** Цвет и названия цвета в русском языке / Под общ. ред. А. П. Василевича. М.: КомКнига, 2005. 216 с.
Vasilevich, A. P., Kuznecova, S. N., Mishchenko, S. S. Color and color names in the Russian language. Ed. by A. P. Vasilevich. Moscow, KomKniga Publ., 2005, 216 p. (in Russ.)
- Герасимов Г. И.** Прилагательные, обозначающие цвет // Русский язык за рубежом. 1969. № 3. С. 26–35.
Gerasimov, G. I. Adjectives denoting color. *Russkij yazyk za rubezhom* [Russian Language Abroad], 1969, no. 3, p. 26–35. (in Russ.)
- Герасимов Г. И.** Снежно-белый, изумрудно-зеленый: о сложных прилагательных, обозначающих цвет // Русская речь. 1978. № 2. С. 79–84.
Gerasimov, G. I. Snow-white, emerald-green: on complex adjectives denoting color. *Russkaya rech'* [Russian Speech], 1978, no. 2, p. 79–84. (in Russ.)
- Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В.** Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 71–113.
Grishina, E. A., Korchagin, K. M., Plungyan, V. A., Sichinava, D. V. The poetic corpus within the framework of the RNC: General structure and prospects of use. In: *Nacional'ny korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [National corpus of the Russian language: 2006–2008. New results and prospects]. St. Petersburg, 2009, p. 71–113. (in Russ.)
- Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.** Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М., 2005. С. 155–174.
Kustova, G. I., Lyashevskaya, O. N., Paducheva, E. V., Rakhilina, E. V. Semantic markup of vocabulary in the national corpus of the Russian language: principles, problems, prospects. In: *Nacional'ny korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* [The National corpus of the Russian language: 2003–2005. Results and prospects]. Moscow, 2005, p. 155–174. (in Russ.)
- Ляшевская О. Н., Шаров С. А.** Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. 1087 с.
Lyashevskaya, O. N., Sharov, S. A. The Frequency dictionary of the modern Russian language (the Russian National Corpus). Moscow, Azbukovnik Publ., 2009, 1087 p. (in Russ.)

- Чекменева А. В.** Оригинальные метафорические употребления цветообозначений (корпусный анализ) // Компьютерная лингвистика и вычислительные онтологии (Тр. XXI Международной объединенной конференции «Интернет и современное общество», IMS-2018): Сб. науч. ст. СПб., 2018. Вып. 2. С. 131–147.
- Chekmeneva, A. V.** Original metaphorical use of color terms (corpus analysis). In: *Komp'yuternaya lingvistika i vychislitel'nye ontologii* (Trudy XXI Mezhdunarodnoj ob"edinennoj konferencii «Internet i sovremennoe obshchestvo, IMS-2018) [Computational linguistics and computational ontologies (Proceedings of the XXI International Joint Conference "Internet and modern society", IMS-2018). St. Petersburg, University ITMO, 2018, issue 2, p. 131–147. (in Russ.)
- Berlin, B., Kay, P.** Basic Color Terms: Their Universality and Evolution. Berkeley, University of California Press, 1969.
- Kay, P., Maffi, L.** Color Appearance and the Emergence and Evolution of Basic Color Lexicons. *American Anthropologist*, 1999, vol. 101 (4), p. 743–760.
- Kay P., McDaniel Ch. K.** The Linguistic Significance of Meanings of Basic Color Terms. *Language*, 1978, vol. 54 (3), p. 610–646.
- Kay P., Regier T.** Language, thought and color: recent developments. *Trends in Cognitive Sciences*, 2006, vol. 10 (2), p. 51–54.
- Levinson, Stephen C.** Yeli Dnye and the Theory of Basic Color Terms. *Journal of Linguistic Anthropology*, 2000, vol. 10 (1), p. 3–55.

Материал поступил в редколлегию
Date of submission
20.09.2018

Сведения об авторах / Information about the Authors

Масевич Андрей Цезаревич, старший преподаватель кафедры информационного менеджмента Санкт-Петербургского государственного института культуры (Дворцовая наб., 2, Санкт-Петербург, 191186, Россия)

Andrey Ts. Masevich, St. Petersburg State Institute of Culture (2 Dvortsovaya Emb., St. Petersburg, 191186, Russian Federation)
andmasev@mail.ru

Захаров Виктор Павлович, кандидат филологических наук, доцент Санкт-Петербургского университета, ведущий научный сотрудник Института лингвистических исследований РАН (Университетская наб., 7-9, Санкт-Петербург, 199034, Россия)

Victor P. Zakharov, St. Petersburg State University (7–9 Universitetskaya Emb., St. Petersburg, 199034, Russian Federation)
v.zakharov@spbu.ru
ORCID 0000-0003-0522-7469

Исследование акустической вариативности *s* методом главных компонент

И. А. Зибер, Г. А. Мороз

*Национальный исследовательский университет
«Высшая школа экономики»
Москва, Россия*

Аннотация

Статья посвящена оценке степени акустической вариативности альвеолярного сибиланта *s* в разных языках. В работе привлекается материал пяти неродственных языков народов России с различными фонетическими системами: адыгейского, нанайского, русского, удмуртского и чукотского. Рассматривается в первую очередь вариативность единиц в одном и том же контексте в условиях контроля максимального числа параметров. Анализируются полевые записи авторов. Произнесения были размечены вручную, затем информация о спектре автоматически извлекалась и трансформировалась методом линейного предиктивного кодирования. Полученные спектральные срезы анализировались по десяти параметрам: частота первого пика (в герцах и барках), угловой коэффициент, отражающий крутизну подъема огибающей до первого пика, спектральный центр масс (в герцах и барках), стандартное отклонение (в герцах и барках), асимметрия и эксцесс. Поскольку анализировать все эти меры одновременно не представляется возможным, для уменьшения числа переменных был использован метод главных компонент. Кумулятивная доля дисперсии, объясненная двумя первыми главными компонентами, составила 65 %. В статье показано, как с помощью метода главных компонент можно измерить и сравнить вариативность произнесений *s* в разных языках. В результате исследования: 1) был создан алгоритм анализа вариативности, который может использоваться и в других областях акустики; 2) можно заключить, что у отдельных дикторов различия в реализации *s* могут быть более резкими, чем между языками в целом; 3) выяснилось, что произнесения на нанайском и чукотском более вариативны, чем на других языках, что соотносится с объемом подсистем сибилантов выбранных языков. Полученные данные могут быть аргументом в пользу того, что в больших подсистемах единицы менее вариативны, и чем меньше подсистема, тем больше вариативность входящих в нее единиц.

Ключевые слова

акустическая вариативность, сибиланты, метод главных компонент, фонетическая типология, консонантизм

Благодарности

Авторы признательны редакции журнала и двум рецензентам за внимание к нашему исследованию и полезные замечания. И. А. Зибер работала над статьей в ходе проведения исследования в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2017–2018 гг. (№ 17-05-0043) и в рамках государственной поддержки ведущих университетов России «5-100»

Для цитирования

Зибер И. А., Мороз Г. А. Исследование акустической вариативности *s* методом главных компонент // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 49–64. DOI 10.25205/1818-7935-2019-17-1-49-64

Estimating the Acoustic Variation of *s* via Principal Component Analysis

Inna A. Sieber, George A. Moroz

*National Research University Higher School of Economics
Moscow, Russian Federation*

Abstract

Articulatory and acoustic variation has recently become one of the most prominent spheres in phonetics. The acoustics of fricative sounds, sibilants in particular, however, is known to be very difficult to study. As far as we are concerned, no algorithm has been created yet to estimate the degree of acoustic variation of fricatives, neither between the speak-

ers of one language, nor between languages. In this article, we try to estimate the degree of acoustic variability of the alveolar sibilant *s*. We were interested in creating and evaluating algorithm for estimation interlanguage variability and make data from different languages comparable. We also were interested in estimation correlation between the phonological complexity of sibilant subsystem and variability of *s* sound. We analyze *s* in similar contexts pronounced by several female speakers of five unrelated languages of Russia: Adyghe, Nanai, Russian, Udmurt and Chukchi. All pronunciations were manually annotated, and then spectral information were automatically extracted and transformed via Linear Predicting Coding. The obtained spectral slices were analyzed and ten different features were extracted: frequency of the first peak (in Hertz and Bark), slope of the linear regression based on values from global minimum before peak (annotated manually) to peak itself, center of gravity (in Hertz and Bark), standard deviation (in Hertz and Bark), skewness and kurtosis. Since it is hard to analyze all these features separately, we used Principal Component Analysis transformation for reducing number of variables. Cumulative percentage of the dispersion explained by first and second Principal Components is equal to 65 %. At the end we show how it is possible to use obtained Principal Components for measuring variability and comparing different utterance of alveolar sibilants. As a result we achieve some goals we planned: 1) we developed the algorithm for variability analysis that could be used in any other field of acoustics; 2) our analysis shows that some speakers could be more variable then the whole languages; 3) the analysis of our data using this algorithm shows that Nanai and Chukchi is more variable comparing to other variables. This also corresponds to the least complex sibilant subsystems.

Keywords

acoustic variation, sibilants, Principal Component Analysis, phonological typology, consonants

Acknowledgements

We are very grateful to the editorial board of the Journal and two our reviewers for their attention to our study and for useful advice. Inna Sieber's work is a part of collective study (№ 17-05-0043) within a program "Scientific fund of the National Research University "Higher School of Economics" (NRU HSE)" in 2017–2018 and due to government support of the leading universities of Russia "5-100".

For citation

Sieber, Inna A., Moroz, George A. Estimating the Acoustic Variation of *s* via Principal Component Analysis. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 49–64. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-49-64

Введение

Открытие бесконечной вариативности звуковых единиц в речи называют одним из важнейших достижений лингвистики XX в. [Ohala, 1993]. В современной фонетической науке исследования вариативности – одна из наиболее активно развивающихся областей (см., например: отдельная глава в [The Oxford Handbook..., 2012]; специальные выпуски журналов и сборников¹; диссертации [Chodroff, 2017]). При этом до сих пор остается открытым вопрос о том, как измерить степень вариативности фонетических единиц. Кроме того, акустика фрикативных в языках народов России изучена недостаточно: со времени, когда проводилось большинство исследований акустической фонетики отдельных идиом (см. [Андреева, 1988] по эвенкийскому языку, [Асиновский, 1999; 2003] по чукотско-камчатским языкам), инструментальная фонетика сильно продвинулась вперед, а современных экспериментальных работ, посвященных отдельным подсистемам консонантизма (см. [Белова, 2000] об удмуртских аффрикатах), по-прежнему мало. В настоящей статье представлены результаты работы в обоих направлениях. В работе используются транскрипционные символы МФА в варианте 2015 г.

Сибиллянты

Принято считать, что турбулентный шум фрикативных согласных образуется при прохождении воздушной струи через сужение между артикуляторами. Так образуются фрикативные несибиллянты, такие как *f*, *x*, *ɸ*, *θ*. При произнесении сибиллянтов сужение в ротовой по-

¹ См., например: Variation and Gradience in Phonetics and Phonology. F. Kügler, C. Féry, R. van de Vijver (eds.). Phonology and Phonetics 14, A. Lahiri (ed.). Berlin, New York, Mouton de Gruyter. 2009; Linguistics (Berlin, De Gruyter, 2017, volume 55, issue 5: Special issue: The Sociophonetics of /s/. Guest Editors: Erez Levon, Marie Maegaard, and Nicolai Pharao).

лости увеличивает скорость воздушного потока, который при столкновении со сближенными зубами порождает специфический шум – высокочастотный и высокоинтенсивный. Согласные, фрикативные и аффрикаты, соответствующие приведенному описанию, и называются чаще всего сибилантами [Ladefoged, Maddieson, 1996. P. 138, 150]. Таковы, к примеру, *s*, *z*, *ʃ*, *ʒ*. Исследования подтверждают, что сибиланты существуют как общность в сознании наивных носителей разных языков – говорящие склонны группировать соответствующие буквы и звуки [Nartey, 1982. P. 13].

Вариативность

В отличие от многих типов звуков для сибилантов чрезвычайно трудно выделить такие акустические параметры, которые надежно отличали бы их друг от друга на уровне языка [Newman et al., 2001; Haley et al., 2010]. Кроме того, иногда сделать обобщения о качестве *s* в языке оказывается невозможным, потому что данные носителей слишком различаются [Gordon et al., 2002]. В лингвистической среде даже возникло мнение, что для сибилантов вовсе невозможно выделить акустический инвариант, а есть только относительное расположение на спектральной картине (см. обзор: [Haley et al., 2010]).

Известны различия и в том, как реализуются сибиланты одного места и способа образования (в частности, *s*) в разных языках [Heffeman, 2004; Li et al., 2007; Fuchs, Toda, 2010]. Межъязыковая вариативность в реализации сибилантов, особенно альвеолярного, лучше всего отражена в исследованиях на материале многих языков, принадлежащих разным ареалам и языковым семьям. Это, главным образом, [Gordon et al., 2002] (семь языков) и [Nartey, 1982] (14 языков). Впечатляющая межъязыковая вариативность альвеолярного сибиланта демонстрирует те широчайшие артикуляционные возможности, которыми он располагает и многими из которых пользуется, по-видимому, в каждом языке. Что же вызывает столь сильную вариативность *s* и других сибилантов, проявляющуюся даже в одном языке в одной и той же позиции? В каких условиях она обнаруживается?

Частично причины вариативности кроются в физиологии. На артикуляцию *s* влияют вид прикуса говорящего, взаимное расположение верхних зубов и нижней челюсти, величина альвеолярного бугра, размеры и форма языка, резцов, всей ротовой полости [Ladefoged, Maddieson, 1996. С. 146. Chodroff, 2017. С. 110]. С учетом множества параметров говорящий находит ту конфигурацию, которая позволяет достигать необходимого акустического эффекта [McCutcheon et al., 1980].

Частично вариативность формируют социолингвистические параметры, такие как пол или гендер говорящего (подробный обзор см. в статье [Fuchs, Toda, 2010] и в сборнике «Linguistics» (2017)), сексуальная ориентация [Munson et al., 2006; Campbell-Kibler, 2011], возраст, социально-экономические характеристики [Levon, Holmes-Elliott, 2013; Podesva, Van Hofwegen, 2014].

Значительная доля вариативности, конечно же, обусловлена собственно языковыми процессами и явлениями. Наибольшее влияние на звук оказывает ближайший фонетический контекст, прежде всего последующий гласный ([Hughes, Halle, 1956; Lindblad, 1980; Shadle, Scully, 1995]). Отмечается, что коартикуляция гласному альвеолярных фрикативных типа *s* сильнее, чем постальвеолярных типа *ʃ* [Nittrover et al., 1989; Tabain, 2001]. Сообщается о влиянии на качество сибиланта слоговой позиции [Silbert, de Jong, 2008], стиля произнесения [Maniwa et al., 2009] и др.

Очень важным представляется вопрос о том, какую роль в формировании вариативности играет фонологическая система языка. Тогда как гласные звуки при заполнении вокалического пространства стремятся сохранять максимальные дистанции, не похоже, чтобы организация систем согласных происходила так же – при этом перцептивная дифференциация, несомненно, имеет значение для этого процесса [Lindblom, Maddieson, 1988]. Каково это значение, не совсем понятно. В старых грамматиках можно найти утверждение о том, что «об-

щевропейское современное *s* стало таким острым благодаря развившемуся наряду с ним звуку *sch*» (/) [Rapp, 1855], иными словами, что «фонологическая оппозиция имеет следствием более четкие акустико-артикуляционные различия» [Гордина, 2006. С. 256]. Современные лингвисты высказываются более осторожно, однако подтверждения этому можно найти в имеющихся типологических данных. Так, в английском языке, где имеется два артикуляционных жеста для фрикативных сибилантов, альвеолярные согласные свободно варьируют положение кончика языка – они бывают апикальными и ламинальными и разделяют говорящих примерно поровну [Bladon, Nolan, 1977; Dart, 1991]. В то же время в языке тода, имеющем четыре артикуляционных жеста для сибилантов, положение кончика языка помогает различать согласные смежных мест образования, и альвеолярный сибилант всегда реализуется как ламинальный [Ladefoged, Maddieson, 1996; Gordon et al., 2002]. Противоположный полюс иллюстрирует гавайский язык (восемь согласных), в котором отмечается свободная вариативность *t* и *k*, но в условиях ограниченного инвентаря: по сути, любой фрикативный и любая аффриката, а также любой не губной и не гортанный взрывной могут функционировать как *k* [Schütz, 1994]. Таким образом, мы видим более высокую вариативность единиц там, где их число меньше, а потому не можем исключать того, что система, различающая большее число единиц внутри некоторого класса, будет накладывать более сильные ограничения на произношение этих единиц, как это происходит в тода.

Вопросы

Несмотря на то что на сегодняшний день обнаружены и описаны многочисленные факторы и параметры вариативности фрикативных, сибилантов, и альвеолярных сибилантов в частности, никто, насколько нам известно, не предпринимал попыток измерить степень этой вариативности с учетом нескольких параметров и представить ее численно. Без измерения вариативности нет возможности ни сравнить то, насколько она велика в разных фонетических контекстах, ни оценить участие многочисленных факторов, о которых говорилось выше. В первую очередь это важно выяснить для альвеолярного сибиланта, который является и самым распространенным сибилантом в языках мира, и, насколько можно судить, самым вариативным.

В цели настоящего исследования, таким образом, входил поиск ответов на следующие проблемные вопросы.

1. В равной ли степени вариативны *s* носителей одного языка?
2. В равной ли степени вариативны *s* в разных языках?
3. Зависит ли степень вариативности *s* от числа сибилантов в языке?
4. Какие акустические измерения отражают вариативность?
5. Как измерить и сравнить степень вариативности?

Исходные данные

Исследование проводилось на материале пяти неродственных языков народов России: адыгейского (абхазо-адыгские, Адыгея и Краснодарский край), удмуртского (уральские, Удмуртия), нанайского (тунгусо-маньчжурские, Хабаровский край и Приморский край), русского (индоевропейские) и чукотского (чукотско-камчатские, Чукотский АО). Говоры, выбранные для исследования, время и место записи и другая информация представлены в приложении 1.

Исследуемые языки различаются по составу подсистемы фрикативных и, в частности, подсистемы сибилантов². Поскольку в противопоставления бывают вовлечены различные

² В настоящей статье под подсистемой сибилантов имеется в виду в первую очередь подсистема фонемного инвентаря. Но поскольку в исследуемых идиомах набор звукотипов сибилантов и фонем сибилантов совпадает, а стимулы содержат /s/ в сильной позиции, для решения наших задач разделять фонетический и фонологический

признаки, формирующие в больших инвентарях серии единиц, мы учитываем в оценке объема подсистемы три числа. Первое – это абсолютное число сибилантов (например, девять в русском языке). Второе – число мест основной артикуляции (например, свистящий и шипящий ряды сибилантов в русском); третье – то же число с учетом палатализации, которая существенно влияет на место сужения в речевом тракте, так что в некоторых случаях палатализованные согласные даже могут рассматриваться в качестве отдельного локального ряда (например, для русского можно постулировать наличие свистящего, свистящего палатализованного, шипящего и шипящего палатализованного рядов сибилантов). Данные об объеме подсистем приводятся в таблице.

Объем подсистемы сибилантов в исследуемых языках
The size of sibilant subsystems in the languages under analysis ³

Язык	Семья	Число сибилантов	Число мест основной артикуляции	Число мест артикуляции с учетом палатализации
Адыгейский	абхазо-адыгские	31	4	4
Нанайский	тунгусо-маньчжурские	3	2	2
Русский	индоевропейские	9	2	4
Удмуртский	уральские	8	3	3
Чукотский	чукотско-камчатские	1	1	1

Единственный сибилант, общий для всех исследуемых языков, – альвеолярный *s*, характеризующийся сужением на передней части альвеолярного бугра ⁴. Он и является объектом нашего исследования.

Исследуемые идиомы имеют различную структуру слога, морфемы и слова. Для сравнения привлекались такие позиции, которые имеются и являются частотными во всех исследуемых идиомах. Таким образом, не рассматривались, в частности, геминированные произнесения типа *рассада* и словоформы с конечным *s* типа *матрас*.

В выборку были включены слова, содержащие *s* перед *a* в начале слова и в середине слова после гласного; таким образом, исследовался согласный в последовательностях *sa*, *asa*, *asa* (например, чук. *sajkok* ‘чайник’, нан. *giasa* ‘изгородь’). Были отобраны 3–5 дикторов для каждого языка, 3–4 слова от каждого диктора, 3–4 повторения каждого слова (изолированные и в рамочной фразе), итого в среднем по 12 произнесений каждого диктора. Все дикторы – женщины.

Список стимулов и число произнесений для каждого диктора приводятся в приложении 2.

уровни нет необходимости, и сибиланты-фонемы и их конкретные реализации в дальнейшем называются одинаково «сибилантами» или «единицами».

³ List of languages in lines: Adyghe, Nanai, Russian, Udmurt, Chukchi.

In columns: language family, number of sibilants, number of places of the basic articulation (regardless of additional articulations), number of places of articulation including palatalization as a separate place.

⁴ В русскоязычной традиции *s* и схожие с ним звуки других языков часто называются зубными, но в действительности очень трудно понять, является тот или иной согласный альвеолярным или зубным. Предполагается, что собственно (язычные) зубные, или дентальные, согласные в языках встречаются нечасто; таков, в частности, тогда, в котором зубной сибилант противопоставлен альвеолярному [Ladefoged, Maddieson, 1996. P. 145].

Измерения

Все произнесения были размечены вручную, измерения производились для 100 % длительности фрикативной части без учета возможных динамических изменений. Разметка и измерения выполнены в программе Praat⁵. Визуализация и анализ методом главных компонент производились с использованием языка программирования R⁶ и пакетов «ggplot2»⁷ и «lingtypology»⁸. Код, созданный для анализа и визуализации данных, доступен по адресу https://github.com/agricolamz/paper_on_fricatives.

Так как человеческое ухо по-разному оценивает высоту звука в зависимости от интенсивности на разных частотах, в исследовании мы использовали не только абсолютные значения в герцах, но и психоакустическое представление на шкале барков, предложенное в работе [Zwicker, 1961].

Поскольку спектральные характеристики звука необходимо было обрабатывать автоматически, для удобства использовалась сглаженная огибающая, полученная методом линейного предиктивного кодирования (ЛПК, LPC, см. [Markel, Gray, 1976]). Был выбран порядок предсказания 20, при котором корректно отображаются максимумы интенсивности, выявляемые на спектральных срезах (быстрое преобразование Фурье), на самом вариативном промежутке – до 10 кГц.

Пример таких огибающих, на которых и производились все измерения, представлен на рис. 1. Это типичный спектр альвеолярного сибиланта, характеризующийся выраженным пиком в области 4–5 кГц с участками малой интенсивности до и после него. Значение первого пика может варьироваться в широком промежутке частот от 3 до 10 кГц, но именно выраженность пика обычно отличает сибиланты от несибилантов [Gordon et al., 2002]. Тем не менее, среди анализируемых произнесений есть и такие, которые характеризует плоский спектр без выраженного пика; такие произнесения воспринимаются на слух как более низкие, менее резкие и менее «компактные». Если произнесения с выраженным пиком являются прототипическими сибилантными произнесениями, то произнесения с плоским спектром – периферийные.

Использованные меры

Для количественной оценки амплитудно-частотного спектра чаще всего используются первые четыре спектральных момента: среднее, дисперсия, асимметрия и эксцесс [Forrest et al., 1988; Jongman et al., 2000]. Средняя частота спектра в герцах, взвешенная по амплитуде, или *спектральный центр масс*, – самая часто используемая статистическая мера [Gordon et al., 2002; Heffeman, 2004; Jones, McDougall, 2009; Haley, 2010].

Стандартное отклонение – это еще одно измерение, которое мы производили для каждой спектральной огибающей. Вместе со спектральным центром масс эта мера, как показали исследования (в частности, [Haley, 2010]), полнее характеризует распределение, показывая рассеивание наблюдаемых значений.

Коэффициент асимметрии (“skewness”) отражает то, насколько нормально распределена спектральная энергия: сдвиг в сторону 0 Гц от среднего дает отрицательное значение [Jones, McDougall, 2009]. Эта мера часто используется наряду с другими спектральными моментами (см., например: [Forrest et al., 1988; Jongman et al., 2000]). Предполагается, что все они так или иначе коррелируют с местом образования шумных согласных [Chodroff, 2017. Р. 106] и эффективно различают разные фрикативные, причем наилучшие результаты дают спектральное среднее и асимметрия [Haley et al., 2010].

⁵ Boersma P., Weenink D. Praat: doing phonetics by computer [Computer program]. Версия 6.0.37. URL: <http://www.praat.org/>. 2018.

⁶ R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>. 2018.

⁷ Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, Springer-Verlag, 2009.

⁸ Moroz G. Lingtypology: easy mapping for Linguistic Typology. 2017. URL: <https://CRAN.R-project.org/package=lingtypology>.

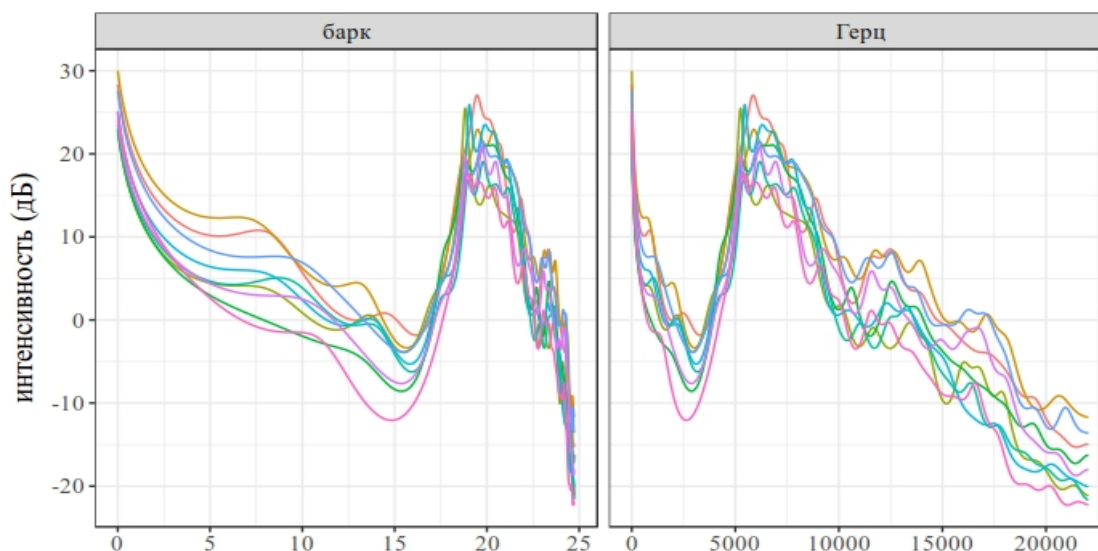


Рис. 1. Спектральные огибающие (ЛПК) в барках (слева) и в герцах (справа) для восьми произнесений; одна линия – одно произнесение. Нанайский язык, диктор – женщина. Предусиление частот до 50 Гц. Порядок предсказания 20

Fig. 1. Spectral slices (LPC) in barks (left, at the x-axis) and Herz (right, at the x-axis) for eight utterances; each line for an utterance. Intensity is at the y-axis. A female speaker for Nanai. Pre-emphasizing from 50 Hz. Number of peaks – 20

Коэффициент эксцесса (“kurtosis”) показывает степень выраженности пика спектральной огибающей, ее форму [Jones, McDougall, 2009] и используется наряду с другими спектральными моментами [Forrest et al., 1988; Jongman et al., 2000].

Поскольку частота первого пика сильно влияет на воспринимаемую высоту звука, мы включили ее в наши измерения. Автоматическое выделение максимумов производилось в пределах частотных промежутков, которые задавались вручную.

Так как в наших данных мы обнаружили как типичный для *s* частотный пик в промежутке 4–5 кГц, так и его отсутствие, мы решили измерить крутизну подъема огибающей. Для этого мы вычисляли локальный минимум перед пиком, строили линейную регрессию на всех точках от локального минимума до локального максимума, а потом брали из получившейся регрессионной формулы угловой коэффициент.

Результаты измерений

В ходе исследования использовались десять переменных: частота первого пика (в герцах и барках), угловой коэффициент (в герцах и барках), спектральный центр масс (в герцах и барках), стандартное отклонение (в герцах и барках), коэффициент асимметрии и коэффициент эксцесса. Когда они используются по одной, результаты измерений удобно сравнивать между собой, но они не могут дать ответа на наши вопросы, так как отражают только отдельные грани акустической картины, каждая из которых не может считаться значимой без участия остальных. При этом громоздкие графики, которые можно получить для каждого типа измерения, не дают никакой возможности сравнивать одновременно несколько переменных, а именно это входит в наши задачи.

Чтобы анализировать сразу несколько переменных, мы решили использовать один из методов уменьшения размерностей.

Методы уменьшения размерностей

Все полученные переменные – числовые, так что к ним можно применять метод главных компонент (МГК, PCA, см., например, [Husson et al., 2017]). В результате работы данного алгоритма ищется такое линейное преобразование n -мерного пространства в другое n -мерное пространство, которое минимизирует корреляцию между отдельными переменными. В результате получается новое n -мерное пространство, которое описывается n компонентами. Компоненты упорядочены на основании доли дисперсии изначальных данных, которую каждая из компонент объясняет. Благодаря этому часть получившихся компонент с высоким порядком можно не рассматривать, так как они в сумме объясняют малое количество дисперсии. Достаточными обычно признаются две-три первые компоненты, они и организуют новое пространство, которое отражает максимум дисперсии и вместе с тем удобно для визуализации и анализа.

Мы применили метод главных компонент (МГК) к трем разным наборам переменных: отдельно к переменным в барках, переменным в герцах и всем переменным вместе. Перед применением МГК все переменные были нормализованы (z-нормализация), для того чтобы единицы, в которых измеряются значения каждой из переменных, были сопоставимы.

Если использовать все переменные, две первые компоненты объясняют 60 % дисперсии, если только барковые – 65 %, если только герцевые – 63 %. Мы приняли решение использовать для визуализации две первые компоненты (ГК1 и ГК2). Так как доля дисперсии, объясненной двумя компонентами, выше всего в барковых переменных, мы будем использовать только барковый вариант.

Доля объясненной дисперсии, которая приходится на каждую из компонент, полученных в ходе применения МГК к нашим данным, представлены на рис. 2.

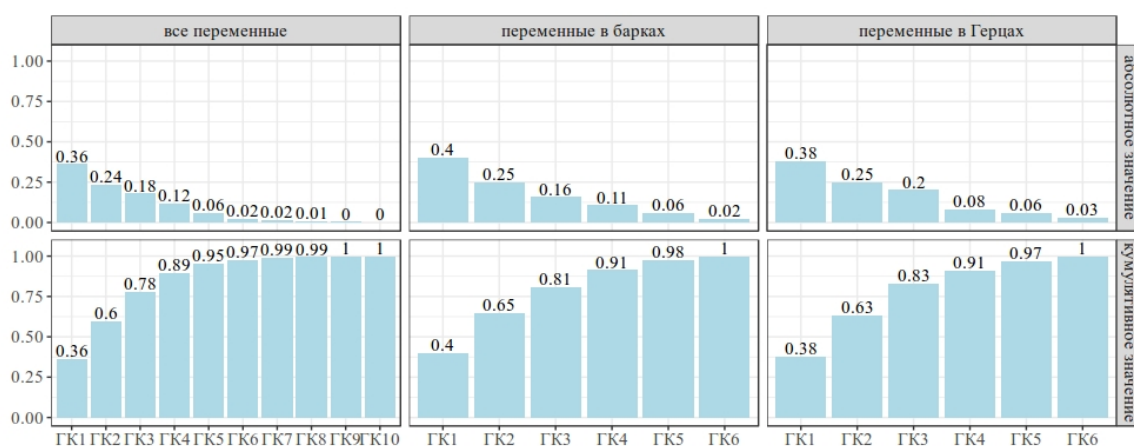


Рис. 2. Доля объясненной дисперсии всех измерений, измерений в барках и в герцах: для каждой компоненты по отдельности (вверху) и кумулятивная (внизу)

Fig. 2. Impact of each PC (Rus. ГК1, ГК2 etc.) in different models: for all variables (right), bark variables (center) and Hertz variables there are cumulative proportion (above) and proportion of variance (below)

На рис. 3 изображены все наблюдения в новой системе координат, где ось ординат представляет собой первую главную компоненту (ГК1), а ось абсцисс – вторую (ГК2). Каждой точке соответствует одно произнесение одного диктора, транспонированное в новое про-

странство. Координаты каждой точки обусловлены одновременно всеми измерениями в барках: спектральным центром масс, стандартным отклонением, коэффициентами асимметрии и эксцесса, частотой первого пика и угловым коэффициентом. Оси, соответствующие этим измерениям и транспонированные в новую систему координат, на рис. 3 обозначены черными стрелками с соответствующими подписями. Чем меньше угол между двумя стрелками, тем больше они скоррелированы по сравнению с остальными переменными; если угол превышает 90° , то коэффициент корреляции будет отрицательным. На рис. 3 мы видим, что такие сильно скоррелированные переменные – коэффициент асимметрии и центр масс ($r = -0,84$), стандартное отклонение и коэффициент эксцесса ($r = -0,52$). Овалы представляют собой доверительные эллипсы для каждого языка (алгоритм расчета см. в [Fox, Weisberg, 2011]). Мы видим, что доверительный эллипс для нанайского имеет большую площадь в сравнении с другими, т. е., предположительно, отражает большую вариативность *s*.

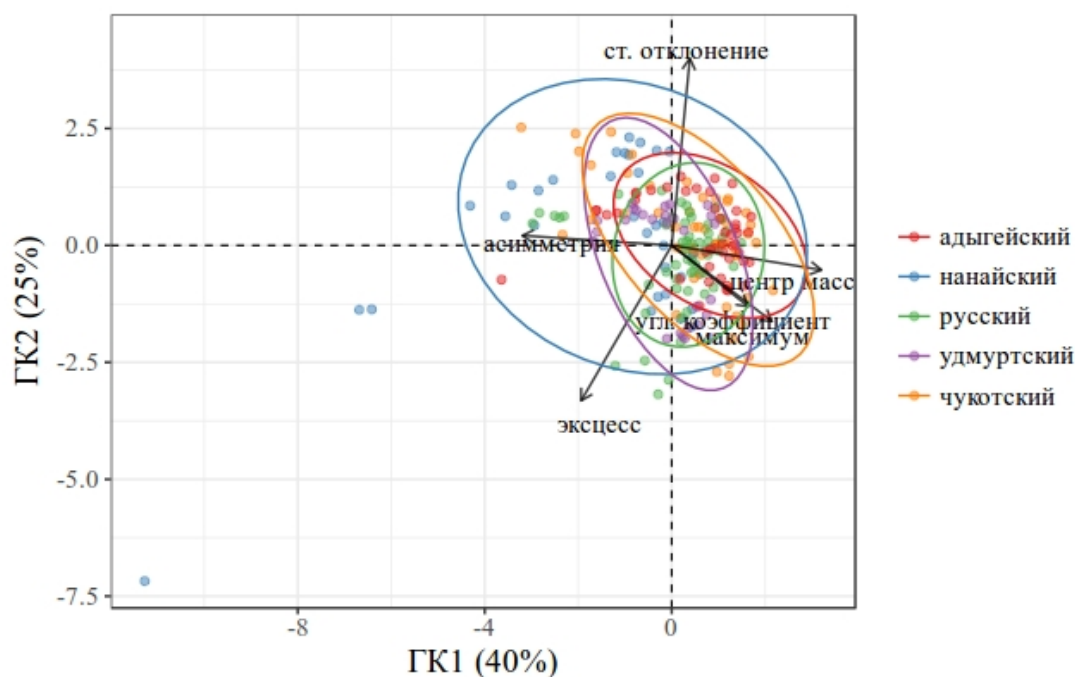


Рис. 3. Наблюдаемые значения и переменные в пространстве первых двух компонент.

Точки – произнесения каждого языка обведены доверительным эллипсом

Fig. 3. PCA (PC 1 and PC 2) with bark variables from the top clockwise: standard deviation, center of gravity, spectral slope, kurtosis, scwness. A point for each [s] utterance, an ellipse for each language: red for Adyghe, blue for Nanai, green for Russian, purple for Udmurt, orange for Chukchi

Чтобы количественно измерить степень вариативности сибиланта в выбранных нами языках, мы решили найти центроиды для каждого из них. В новом пространстве мы нашли точку, расстояния до которой от всех наблюдений по одному языку будет наименьшим. После этого мы измерили среднее расстояние от наблюдаемых точек до данного центроида. Полученные значения являются усредненными значениями евклидова расстояния между точками в новом пространстве, полученном после нормализации переменных и применения МГК. Результаты можно увидеть на рис. 4. На графике также приводятся среднее и стандартное отклонение расстояний для центроида каждого языка.

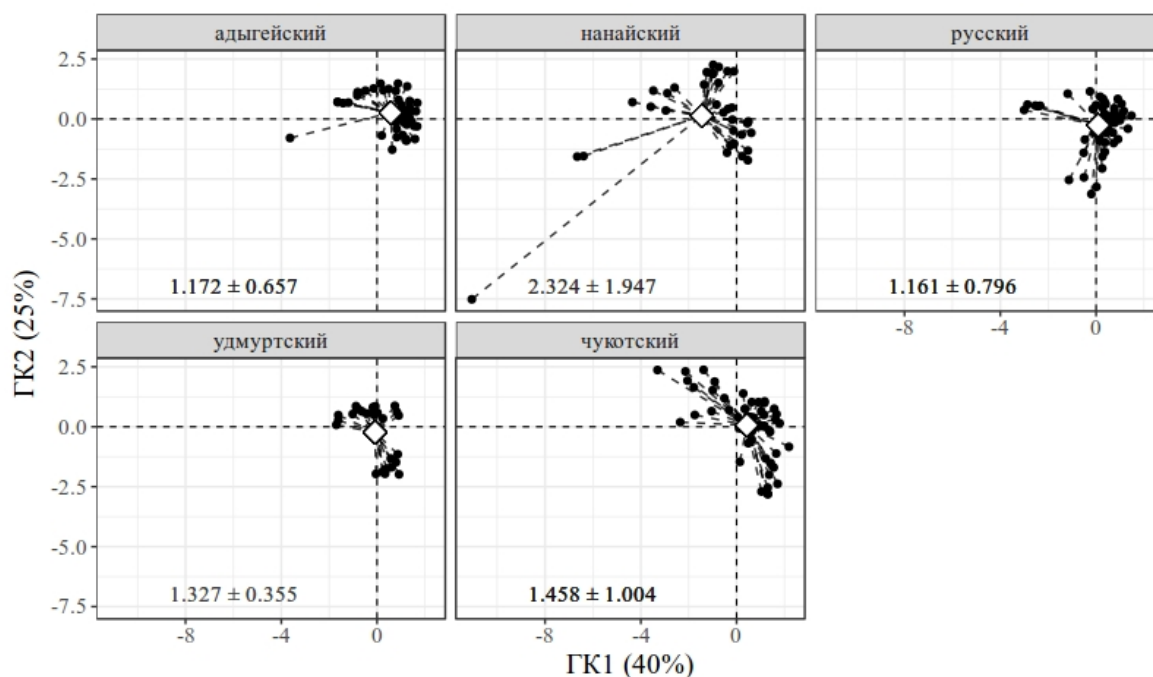


Рис. 4. Наблюдаемые значения, центроиды (обозначены ромбом) для каждого языка и расстояния до них в пространстве первых двух главных компонент (ГК1 и ГК2). Внизу подписаны среднее и стандартное отклонение расстояния до центроида для каждого из языков

Fig. 4. PCA (PC 1 and PC 2) with bark variables, from top left: Adyghe, Nanai, Russian, Udmurt, Chukchi.

A point for each utterance, a rhombus marks a centroid. For each language mean distance from the centroid to the utterances and standard deviation of the distance are estimated

Стандартное отклонение в данном случае иллюстрирует то, насколько сильно наблюдения отклоняются от среднего. Так, если стандартное отклонение достаточно велико, значит, можно подозревать в наших данных выбросы – наблюдения, сильно отличающиеся от основной выборки. Меньшее стандартное отклонение сигнализирует более равномерное распределение вокруг центроида.

Интерпретация результатов

Языки с наибольшим средним расстоянием до центроида – нанайский и чукотский. Как видно из графиков и значений стандартного отклонения, своим первенством эти языки – и особенно нанайский – во многом обязаны отдельным произнесениям, которые отличаются от остальных. Мы видим такие произнесения во всех языках, кроме удмуртского, но у нанайского и чукотского их больше всего. По-видимому, это важная особенность структуры вариативности ζ в этих языках.

Структура вариативности в русском и адыгейском языках в целом сходная. Часть значений расположена компактно, при этом присутствуют отдельные выбросы. В русском таких выбросов значительно больше, чем в адыгейском, что видно на графике и отражено в значениях стандартного отклонения, которое для русского языка выше. Таким образом, хотя среднее расстояние примерно равно для русского и адыгейского (у адыгейского даже выше

на 1 %), высокое значение стандартного отклонения для русского языка вкупе с формой графиков дает основания говорить о том, что русский *s* более вариативен, чем адыгейский.

На графике произнесений для удмуртского языка видно, что точки-значения формируют две компактные группы, не сообщающиеся между собой (см. рис. 4). Такую картину может давать контекстное распределение двух типов реализаций, но в данном случае, как представляется, дело именно в вариативности между дикторами. Это заметно по отдельным измерениям. В частности, значения спектрального центра масс у половины удмуртских дикторов расположены в широком промежутке до 4 кГц, тогда как у другой половины все они выше 5 кГц и даже 5,5 кГц. Такой картины в остальных языках мы не наблюдаем. Можно сказать, что для удмуртского языка с его компактным расположением значений различия между говорящими играет ключевую роль в формировании структуры вариативности.

Таким образом, на основании среднего расстояния до центроида и с учетом стандартного отклонения можно, имея в виду сделанные выше оговорки, выстроить языки в следующем порядке от наиболее вариативного до наименее вариативного *s*: нанайский – чукотский – удмуртский – русский – адыгейский.

Сопоставим с полученным порядком данные об инвентаре сибилантов в этих языках. Поскольку подсистемы сибилантов в языках устроены по-разному, место в иерархии объема подсистем может зависеть от способа его оценки. Число мест основной артикуляции (в скобках), вопреки ожиданиям, не повторяет порядка нашей иерархии степени вариативности: нанайский (2) – чукотский (1) – удмуртский (3) – русский (2) – адыгейский (4).

Ближе к этому порядку подходит оценка числа мест образования, учитывающая палатализованные согласные и корректирующая место русского языка: нанайский (2) – чукотский (1) – удмуртский (3) – русский (4) – адыгейский (4).

В целом соответствует иерархии степени вариативности абсолютное число сибилантов: нанайский (3) – чукотский (1) – удмуртский (8) – русский (9) – адыгейский (31).

Сопоставление иерархии вариативности и объема подсистем сибилантов для языков не дает однозначного ответа на вопрос, какой из параметров инвентаря влияет на степень вариативности больше всего, но показывает некоторые закономерности. С одной стороны, мы видим, что самыми вариативными оказываются нанайские и чукотские согласные. Нанайский и чукотский имеют меньше всего сибилантов по всем трем оценкам, что особенно заметно по абсолютному числу, которое различается в разы (три нанайских сибиланта и один чукотский против восьми в удмуртском, девяти в русском и 31 в адыгейском). С другой стороны, адыгейский, который мы признали наименее вариативным, по всем оценкам имеет больше всего сибилантов. Это дает основания полагать, что, действительно, существует связь между объемом подсистемы сибилантов и степенью вариативности *s* в языке.

В то же время интерпретация данных русского языка неоднозначна. Причина этого в дополнительной артикуляции палатализации, которая наличествует и имеет функциональную нагрузку в русском, но отсутствует в остальных исследуемых языках. Дело в том, что палатализация оказывает большое влияние на место основного сужения язычных согласных. Для палатализованных постальвеолярных согласных (или альвеоло-палатальных) даже приняты отдельные символы в Международном фонетическом алфавите (*ɕ* *ʑ*). В лингвистической литературе ведутся споры о том, стоит ли считать альвеоло-палатальные согласные палатализованными постальвеолярными или же они имеют свое особое основное место образования (ср. МФА 2005 г.). Для палатализованных альвеолярных согласных этот вопрос не стоит так остро, но некоторые лингвисты утверждают, что русские *s* и *s'* отличаются друг от друга не меньше, а то и больше, чем *f* и *ɕ* (С. В. Князев, личное сообщение), таким образом, расщеплять альвеолярное место образования имеет столько же смысла, сколько постальвеолярное. Оказывается, что в зависимости от того, как оценивается объем подсистемы сибилантов, русский язык по числу мест образования может приравниваться к нанайскому, где сибилантов втрое меньше (в обоих выделяются свистящие и шипящие), адыгейскому, где их втрое больше (в обоих можно выделить четыре целевые артикуляции для фрикативных) или бе-

сермянскому, где их почти столько же (если принимать альвеоло-палатальные согласные за отдельный локальный ряд и учитывать бесермянские «шепелявые» и русские альвеоло-палатальные одинаково). При этом по совокупности признаков сибилантов, да и по абсолютному числу сибилантов, русская подсистема является несколько более дробной, чем бесермянская, и может, таким образом, занимать то же место в иерархии объема подсистем, какое она занимает в иерархии степени вариативности.

Таким образом, наши результаты могут быть некоторым аргументом в пользу того, что степень вариативности зависит от объема подсистемы. Это требует дальнейшей проверки на большем материале.

В предыдущих разделах мы продемонстрировали алгоритм оценки вариативности произнесения s в разных языках. Преимуществом предложенного метода является наглядная визуализация, которая позволяет отобразить с некоторой потерей данные по десяти переменным по любому числу языков. Более того, данный метод вовсе не ограничен исключительно сибилантом s , а может быть распространен и на другие фрикативные. Важно подчеркнуть, что наш подход позволяет использовать и большее число переменных, не ограничивая исследования каким-то определенным набором признаков. Так, во многих работах исследователи пытаются найти одну или две переменные, которые позволили бы им различать те или иные типы сибилантов, тогда как наш подход позволяет использовать столько переменных, сколько представляется необходимым.

Недостатком нашего метода является неоднозначность интерпретаций самих компонент. Для настоящего исследования было важно унифицировать число произнесений, повторений и дикторов, записанных в полевых условиях, поэтому общее количество данных неизбежно было небольшим. Любые новые данные могут изменить картину. Кроме того, если другой исследователь приложит метод к своим данным, он также может получить совсем другие компоненты.

Заключение

В нашей работе мы попытались количественно оценить вариативность произношения s в адыгейском, нанайском, русском, удмуртском и чукотском языках. Наши данные позволяют нам ответить на поставленные вопросы следующим образом.

1. Альвеолярный сибилант носителей одного языка может быть вариативен в разной степени. Некоторым дикторам свойственна большая вариативность произнесений s , чем другим дикторам – носителям того же языка, и даже целым языкам.

2. В разных языках s вариативен в разной степени. Из пяти исследуемых языков наиболее вариативными по итогам применения нашего метода оказались нанайский и чукотский, а наименее вариативным – адыгейский.

3. Наши результаты могут быть аргументом в пользу того, что структура фонетической системы и ее подсистем, в частности абсолютное число сибилантов, влияет на степень артикуляционной и акустической вариативности отдельных единиц. Вариативность альвеолярного сибиланта выше в тех фонетических системах, где мало сибилантов, и ниже в тех системах, где системы сибилантов богаче. Этот вопрос, тем не менее, требует дальнейшего изучения.

4. Мы оценивали произнесения при помощи следующих измерений: частота первого пика (в герцах и барках), угловой коэффициент (в герцах и барках), спектральный центр масс (в герцах и барках), стандартное отклонение (в герцах и барках), коэффициент асимметрии и коэффициент эксцесса. Вариативность эффективно отражают все выбранные нами меры.

5. Измерения производились автоматически, после чего анализировались методом главных компонент, который позволил расположить их все на новой двумерной плоскости. Максимально информативной оказалась модель, где использовались лишь барковые переменные. В новом пространстве мы нашли центроид для каждого языка и оценили расстояние от всех

наблюдаемых произнесений до языкового центроида. Среднее и стандартное отклонение расстояний от наблюдаемых значений до центроида позволили оценить вариативность произношения в каждом языке и сравнить по этим параметрам все языки.

Список литературы / References

- Андреева Т. Е.** Звуковой строй томмотского говора эвенкийского языка. Экспериментально-фонетические исследования. Новосибирск: Наука, 1988.
Andreeva, T. E. The sound system of Tommot Evenki. Experimental phonetic studies. Novosibirsk, Nauka, 1988. (in Russ.)
- Асиновский А. С.** Консонантизм чукотского языка (на основе экспериментально-фонетических данных). Л.: Наука, 1991.
Asinovskij, A. S. The consonantal system of Chukchi (based on experimental data). Leningrad, Nauka, 1991. (in Russ.)
- Асиновский А. С.** Сопоставительная фонетика чукотско-камчатских языков: Дис. ... д-ра филол. наук: 10.02.02. Санкт-Петербург, 2003.
Asinovskij, A. S. Comparative phonetics of the Chukotko-Kamchatkan languages. Doctoral dissertation. St. Petersburg, 2003. (in Russ.)
- Белова Е. Б.** Аффрикаты удмуртского языка в экспериментальном освещении: Дис. ... канд. филол. наук: 10.02.07. Ижевск, 2000.
Belova, E. B. Udmurt affricates in the experimental study. PhD thesis. Izhevsk, 2000. (in Russ.)
- Гордина М. В.** История фонетических исследований (от античности до возникновения фонологической теории). СПб.: Филологический факультет СПбГУ, 2006.
Gordina, M. V. The history of phonetic research (from antiquity to the rise of phonological theory). St. Petersburg, The faculty of philology, Saint Petersburg State University, 2006. (in Russ.)
- Bladon, R. A. W., Nolan, F.** A videofluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, 1977, no. 5, p. 185–93.
- Campbell-Kibler, K.** Intersecting variables and perceived sexual orientation in men. *American Speech*, 2011, vol. 86 (1), p. 52–68.
- Chordoff, E. R.** Structured variation in obstruent production and perception. A PhD dissertation. Baltimore, Maryland, 2017, 281 p.
- Dart, S.** Articulatory and Acoustic Properties of Apical and Laminal Articulations. *UCLA Working Papers in Phonetics*, 1991, no. 79, p. 1–155.
- Forrest, K., Weismer, G., Milenkovic, P., Dougall, R. N.** Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 1988, vol. 84 (1), p. 115–123.
- Fox, J., Weisberg, S.** An R Companion to Applied Regression. Second Edition. Thousand Oaks CA: Sage, 2011. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
- Fuchs, S., Toda, M.** Do differences in male versus female /s/ reflect biological or sociophonetic factors? In: *Turbulent Sounds. An Interdisciplinary Guide*, 2010, p. 281–302.
- Gordon, M., Barthmaier, P., Sands, K.** A cross-linguistic acoustic study of voiceless fricatives. *Journal of International Phonetic Association*, 2002, no. 32, p. 141–174.
- Haley, K. L., Seelinger, E., Mandulak, K. C., Zajac, D. J.** Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics*, 2010, vol. 38 (4), p. 548–554.
- Heffeman, K.** Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics*, 2004, vol. 23 (2), p. 71–84.
- Hughes, G. W., Halle, M.** Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, 1956, vol. 28 (2), p. 303–310.

- Husson, F., Lê, S., Pagès, J.** Exploratory multivariate analysis by example using R. Chapman and Hall/CRC, 2017.
- Jones, M. J., McDougall, K.** The Acoustic Character of Fricated /t/ in Australian English: A Comparison with /s/ and /esh/. *Journal of the International Phonetic Association*, 2009, vol. 39.3, p. 265–289.
- Jongman, A., Wayland, R., Wong, S.** Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 2000, vol. 108 (3), p. 1252–1263.
- Ladefoged, P., Maddieson, I.** The Sounds of the World's Languages. Blackwell, 1996, 414 p.
- Levon, E., Holmes-Elliott, S.** East end boys and west end girls: /s/-fronting in Southeast England. *University of Pennsylvania Working Papers in Linguistics*, 2013, vol. 19 (2), p. 111–120.
- Li, F. F., Edwards, J., Beckman, M.** Spectral Measures for Sibilant Fricatives of English, Japanese and Mandarin Chinese. In: Proceedings of the 16th International Congress of Phonetic Sciences, 6–10 August 2007. Saarbrücken, Germany. 2007, p. 917–920.
- Lindblad, P.** Svenskans sje- och tje-ljud i ett Allmänfonetisk Perspektiv. Lund, C. W. K. Gleerup, 1980.
- Lindblom, B., Maddieson, I.** Phonetic Universals in Consonant Systems. In: Li, C., Hyman, L. M. (eds.). Language, Speech and Mind. London, Routledge, 1988, p. 62–78.
- Maniwa, K., Jongman, A., Wade, T.** Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, 2009, vol. 125 (6), p. 3962–3973.
- Markel, J. D., Gray, A. J.** Linear prediction of speech. Berlin, Springer-Verlag, 1976.
- McCutcheon, M. J., Hasegawa, A., Fletcher, S. G.** Effects of palatal morphology on /s, z/ articulation. *The Journal of the Acoustical Society of America*, 1980, vol. 67 (S1), p. S94–S94.
- Munson, B., Jefferson, S. V., McDonald, E. C.** The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America*, 2006, vol. 119 (4), p. 2427–2437.
- Nartey, J. N. A.** On Fricative Phones and Phonemes. PhD Dissertation. UCLA, 1982, 192 p.
- Newman, R. S., Clouse, S. A., Burnham, J. L.** The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 2001, vol. 109 (3), p. 1181–1196.
- Nittrouer, S., Studdert-Kennedy, M., McGowan, R. S.** The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 1989, vol. 32 (1), p. 120–132.
- Ohala, J. J.** The phonetics of sound change. In: Historical linguistics: Problems and perspectives. Ed. by C. Jones. London, New York, Longman, 1993, p. 237–278.
- Podesva, R. J., Van Hofwegen, J.** How conservatism and normative gender constrain variation in inland California: The case of /s/. *University of Pennsylvania Working Papers in Linguistics*, 2014, vol. 20 (2), p. 128–137.
- Rapp, K. M.** Vergleichende Grammatik als Naturwissenschaft. Stuttgart, Tübingen, 1855.
- Schütz, A. J.** The Voices of Eden: A History of Hawaiian Language Studies. Honolulu, University of Hawai'i Press, 1994.
- Shadle, C. H., Scully, C.** An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, 1995, vol. 23 (1), p. 53–66.
- Silbert, N., de Jong, K.** Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *The Journal of the Acoustical Society of America*, 2008, vol. 123 (5), p. 2769–2779.

- Tabain, M.** Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production. *Language and Speech*, 2001, vol. 44 (3), p. 57–94.
- The Oxford Handbook of Laboratory Phonology. Eds. A. C. Cohn, C. Fougerson, M. K. Huffman. Oxford University Press, 2012.
- Zwicker, E.** Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, 1961, vol. 33 (2), p. 248–248.

Приложение 1
Supplement 1

В таблице приводится основная информация об исследуемых аудиозаписях.
The following table contains information on the audio recordings under study.

Язык	Диалект	Сведения о проведении аудиозаписи		
		место	год	исследователь
Адыгейский	бжедугский	аул Нешукай, Теучежский р-н, Адыгея	2014	Г. Мороз
Нанайский		с. Джуен, Амурский р-н, Хабаровский край	2013	
Русский	литературный, младшая норма	г. Москва	2017	И. Зибер, О. Волков, Г. Мороз
Удмуртский	бесермянский	д. Шамардан, Юкаменский р-н, Удмуртия	2010–2016	коллектив исследователей из МГУ и НИУ ВШЭ
Чукотский		с. Амгуэма, Иультинский р-н; г. Анадырь, Чукотский АО	2016–2017	

Приложение 2
Supplement 2

В таблице представлены стимулы на пяти языках с переводом и указанием числа произнесений для каждого из дикторов. Дикторы анонимны и пронумерованы от одного до пяти для каждого языка. Если использовались произнесения трех или четырех дикторов, в графах для остальных дикторов стоит отметка «нет». Если слово не записано от некоторого диктора, в соответствующей графе стоит «0».

The table contains stimuli in five languages with translation to Russian and the number of utterances for each speaker (speaker 1 is “Д. 1” etc.).

Язык	Стимул	Перевод	Д. 1	Д. 2	Д. 3	Д. 4	Д. 5
Адыгейский	sabəj	ребенок	4	4	4	нет	нет
	sapɐ	пыль	4	4	4	нет	нет
	ħjəsap	математика	4	4	4	нет	нет
Нанайский	qiasa	изгородь	3	3	3	3	3
	sajna	наверное	0	3	3	3	3
	sain	отметина	0	3	3	0	0
	saji	отметина	0	0	0	0	3
	sajrakta	бусинка	0	0	0	0	3

Окончание таблицы

Язык	Стимул	Перевод	Д. 1	Д. 2	Д. 3	Д. 4	Д. 5
Русский	сахар		4	4	4	4	4
	коса		4	4	4	4	4
	осанка		4	4	4	4	4
Удмуртский	sajkanə	протрезветь	3	0	0	3	нет
	sajkaz	протрезвел	0	0	0	3	нет
	sajkatənə	заставить протрезветь	0	3	0	0	нет
	sajkət	ясный, светлый	3	0	0	3	нет
	sajkiəkənə	очнуться, просветлеть	0	0	3	0	нет
	sajkatekənə	очнуться, просветлеть	0	3	0	0	нет
	sajkateke	очнулся, просветлел	0	0	3	0	нет
Чукотский	nəsasaqen	вкусный	2	6	4	3	3
	sajkok	чайник	3	3	3	4	0

Материал поступил в редколлегию
Date of submission
14.10.2018

Сведения об авторах / Information about the Authors

Зибер Инна Арнольдовна, преподаватель Школы лингвистики Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

Inna A. Sieber, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)
innasieber@gmail.com

Мороз Георгий Алексеевич, младший научный сотрудник Международной лаборатории языковой конвергенции Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

George A. Moroz, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)
agricolamz@gmail.com

Использование мер семантической близости для распознавания кореференции в русском языке

И. Л. Азеркович

*Национальный исследовательский университет
«Высшая школа экономики»
Москва, Россия*

Аннотация

Статья посвящена описанию серии экспериментов по исследованию роли семантической информации в разрешении кореферентных связей для русского языка, ее использованию в системах автоматического анализа текстов и оценке результатов их работы. Целью первого этапа экспериментов было определить, какие метрики семантической близости между референциальными выражениями больше соответствуют кореферентным связям между ними. Подсчет метрик производился на материале русской Википедии и тезауруса RuThes. На втором этапе была разработана система автоматического распознавания кореферентности, использующая метрики семантической близости в качестве признаков для машинного обучения, и оценено качество ее работы. Результаты проведенных экспериментов позволяют установить метрики семантической близости, подходящие для использования в системах разрешения кореферентности, а также демонстрируют повышение качества работы подобных систем при использовании семантической информации.

Ключевые слова

автоматическая обработка естественного языка, распознавание кореферентности, метрики семантической близости, машинное обучение, русский язык

Для цитирования

Азеркович И. Л. Использование мер семантической близости для распознавания кореференции в русском языке // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 65–77. DOI 10.25205/1818-7935-2019-17-1-65-77

Using Semantic Relatedness Measures in Coreference Resolution for Russian

Ilya L. Azerkovich

*National Research University Higher School of Economics
Moscow, Russian Federation*

Abstract

The paper is devoted to the role of semantic information (in the form of semantic relatedness measures) in coreference resolution for the Russian language. It describes a series of experiments in calculating metrics of semantic relatedness based on Russian material and evaluating the possibility of using them in systems of natural language processing, as well as the performance of such systems.

The goal of the first stage of experiments was to find out, which semantic relatedness measures better correspond to coreference relations between referential expressions. For this purpose, several metrics calculated from different parameters were chosen and evaluated on the test set, derived from the Russian coreference corpus RuCor. Semantic data for the metrics was obtained from two sources: Russian Wikipedia and RuThes thesaurus. The results showed that while RuThes provided more reliable data for common nouns, Wikipedia data correlated better with named entities. Based on the obtained results, metrics that corresponded to coreference relations the most were chosen to be implemented during the next stage of experiments.

For the second stage of experiments a machine-learning based coreference resolution system that could use semantic relatedness measures as features was developed, based on the decision trees classification algorithm. Four versions of the system were tested: without any features derived from semantic information, with features derived from only one

of the sources, and with features derived from both sources. Tests were performed on the subset of RuCor corpus that already included gold standard mark-up as the base for evaluation. The tests showed noticeable improvement for the version that was using semantic information from both data sources.

The experiments made demonstrate the increase of quality of coreference resolution with the implementation of features based on semantic information. The results obtained are comparable to or exceed the ones described in similar papers on the topic of Russian coreference resolution.

Keywords

natural language processing, coreference resolution, semantic relatedness measures, machine learning, Russian language

For citation

Azerkovich, Ilya L. Using Semantic Relatedness Measures in Coreference Resolution for Russian. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 65–77. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-65-77

Введение

Распознавание кореферентных связей является одной из задач, решение которой актуально для различных систем автоматического анализа текста, таких как, например, извлечение информации из текста, машинного перевода и др. Два языковых выражения находятся в кореферентных отношениях, если они обозначают одну и ту же сущность реального мира. Таким образом, задача распознавания кореферентных отношений состоит в выделении в тексте цепочек упоминаний одной и той же сущности. Кореферентную цепочку, например, составляют выделенные референциальные выражения в следующем примере:

- (1) [Facebook] уже несколько лет делает ставку на видео. <...> К февралю 2017 года наконец стало понятно, насколько амбициозны планы [компании], – [она] хочет заменить собой телевизор.

Классические алгоритмы для решения такой задачи в основном опираются на данные синтаксического и морфологического разбора текста с использованием базовых семантических признаков, например информации об именованных сущностях или о семантической совместимости кореферентных выражений в терминах таксономической иерархии (ср. семантические классы, описанные в [Kameyama, 1997] или [Soon et al., 2001]). В качестве источников семантической информации во многих исследованиях используются внешние источники, такие как тезаурусы (WordNet) или онлайн-энциклопедии (Википедия) [Harabagiu et al., 2001; Ponzetto, Strube, 2006; Rahman, Ng, 2011].

Разработка систем автоматического разрешения кореферентности для русского языка началась позже, чем для английского, но ведется достаточно активно (см. описания систем в [Толпегин, 2008; Kamenskaya et al., 2014; Kutuzov, Ionov, 2014], а также [Toldova et al., 2014] о результатах соревнования между системами). Из-за морфологических и синтаксических особенностей русского языка стандартные методы, применяемые для английского языка, могут работать не так хорошо ([Toldova et al., 2016]), но, тем не менее, как показали работы [Toldova, Ionov, 2017; Azerkovich, 2018], использование признаков, основанных на семантической информации, позволяет повысить их эффективность.

Данная работа развивает результаты, достигнутые в статьях [Toldova, Ionov, 2017; Azerkovich, 2018]. В ней излагаются результаты двух серий экспериментов, в которых исследуется роль семантической информации в автоматическом распознавании кореферентности. Первая из них была призвана ответить на вопрос, какие метрики семантической близости между референциальными выражениями больше соответствуют кореферентным связям между ними. Вторая серия экспериментов описывает работу системы автоматического распознавания кореферентности, использующей выбранные на первом этапе метрики в качестве признаков для машинного обучения.

Результаты проведенных экспериментов демонстрируют повышение качества распознавания кореферентности с использованием семантической информации. Полученное улучшение превышает описанное С. Толдовой и М. Ионовым в статье [Toldova, Ionov, 2017], что

говорит о лучшем качестве использованных признаков. Кроме того, в рамках исследования проведена оценка различных мер семантической близости между референциальными выражениями и создан рабочий прототип системы автоматического распознавания кореферентности, использующий данные метрики.

Предыдущие работы

Ранние системы автоматического разрешения кореферентности были основаны на правилах, но с начала 2000-х гг. наблюдается активное развитие систем, полностью основанных на машинном обучении либо использующих комбинированный статистически-правильный подход [Soon et al., 2001; Ng, Cardie, 2002]. Как показали результаты конференции CONLL-2012 (см. [Pradhan et al., 2012]), комбинированные системы позволяют получить лучшие результаты в сравнении с системами с другими архитектурами. Система, описанная в работе [Soon et al., 2001], считается классической благодаря простоте реализации и высокому качеству анализа и часто воспроизводится в статьях других исследователей.

Многие системы разрешения кореферентности, как ранние, так и более поздние, используют семантическую информацию в основном в виде сведений о семантической однородности (semantic consistency) референциальных выражений (например, совпадении по роду и числу или принадлежности к одному и тому же таксономическому классу, ср. [Kameyama, 1997; Harabagiu et al., 2001]). Эта информация может быть выделена как другими модулями обработки текста (например, система извлечения именованных сущностей Стэнфордского университета в [Rahman, Ng, 2011]), так и модулями, непосредственно встроенными в систему (информация о семантических ролях в [Kamenskaya et al., 2014] или извлечение именованных сущностей в [Soon et al., 2001]). С появлением электронных ресурсов, содержащих экспертную и онтологическую информацию, все большее число исследователей использует в качестве источников именно их. Одними из самых часто используемых ресурсов являются тезаурус WordNet (см. [Harabagiu et al., 2001; Ponzetto, Strube, 2006]) и онлайн-энциклопедия Википедия (см. [Ponzetto, Strube, 2006; Haghighi, Klein, 2009; Rahman, Ng, 2011]).

На материале русского языка исследований по автоматическому разрешению кореферентности и по использованию для этого семантической информации на момент написания статьи существует меньше, чем для английского. В рамках форума RuEval, прошедшего в 2014 г., было проведено первое соревнование систем такого рода [Toldova et al., 2014]. В соревновании были отдельно представлены задачи автоматического разрешения анафоры и кореферентности между именами собственными, именными группами и местоимениями, и из восьми систем – участников трека по разрешению анафоры три также приняли участие в треке по разрешению кореферентности. Однако из трех систем лишь одна [Bogdanov et al., 2014] использовала семантическую информацию, причем в качестве ее источника выступала проприетарная онтология.

Предоставление публичного доступа к корпусу текстов с кореферентной разметкой, использовавшемуся при проведении соревнования, значительно облегчило задачу разработки подобных систем, однако проблема использования семантической информации до недавнего времени была изучена недостаточно. В то же время за последний год были опубликованы две работы, описывающие улучшение качества работы систем автоматического разрешения кореферентности при использовании семантической информации: [Toldova, Ionov, 2017; Azerkovich, 2018].

Первая статья описывает эксперимент по улучшению качества распознавания кореферентности с использованием таких семантических средств, как векторы представления слов, списки именованных сущностей и тезаурус RuThes. Несмотря на сравнительно простые признаки, использовавшиеся в работе, результаты исследования демонстрировали повышение качества в пределах 0,2–0,3 %. Это позволило предположить, что использование других способов представления семантической информации позволит еще больше повысить качество распознавания. В настоящем исследовании были проведены эксперименты по подбору мет-

рик семантической близости, наиболее полезных при идентификации кореферентных связей. Их результаты показали, что семантические признаки, основанные на подобранных метриках, дают значительно более ощутимое улучшение, чем семантические признаки, описанные в статье С. Толдовой и М. ИONOва.

Вторая статья описывает результаты эксперимента по использованию информации из русской Википедии для автоматического распознавания кореферентности. Положительные результаты, полученные на небольшом корпусе с применением признаков, основанных на пересечениях между текстами статей, позволили продолжить эксперименты с данным видом информации, описанные далее.

Эксперименты

Для исследования различных параметров интеграции семантической информации в систему распознавания кореферентных связей была проведена серия экспериментов. Эксперименты были призваны решить следующие задачи: оценить различные метрики семантической близости двух референциальных выражений и выделить метрики, использование которых дает наилучшие результаты, а также оценить эффективность использования энциклопедической информации для обсуждаемой задачи. Данными для экспериментов, описываемых ниже, послужил корпус RuCor, подготовленный в рамках соревнования Ru-Eval-2014.

Использованные данные

Эксперименты проводились на материале корпуса RuCor [Toldova et al., 2016], собранного для соревнования систем автоматического распознавания кореферентности, проводившегося в рамках форума RuEval-2014. В корпус входит 180 текстов различных жанров, содержащих в общей сложности 3 638 кореферентных цепочек и 16 557 кореферентных групп. Все тексты токенизированы, морфологически и синтаксически размечены с помощью инструментов, описанных в [Sharoff, Nivre, 2001]. В соревновании корпус использовался в том числе в качестве «золотого» стандарта разметки, в связи с чем в нем выделены и снабжены соответствующими тегами все кореферентные выражения.

Первый этап экспериментов

Целью первого этапа экспериментов была разработка оптимального представления используемой для анализа семантической информации. Для этого в качестве альтернативного источника информации был проанализирован тезаурус RuThes, и подвергнуты сравнению различные метрики семантической близости между референтами.

Альтернативным источником энциклопедических знаний, использованным на данном этапе, стал тезаурус русского языка RuThes, описанный в работе [Лукашевич, 2011]. Его опубликованная часть RuThes-lite включает в себя 55 тысяч сущностей, которым соответствует 158 тысяч лексических входов, слов и выражений. Сущности внутри онтологии соединены между собой отношениями *выше/ниже*, *часть/целое*, а также ассоциативными связями.

В качестве альтернативы Википедии был выбран именно этот ресурс, потому что по многим признакам, как структурным, так и содержательным, представлял ее полную противоположность:

1) Википедия является «открытой энциклопедией», т. е. она доступна для редактирования любому пользователю, но ее пополнение не является чьей-либо обязанностью. В связи с этим информация в Википедии часто может быть неполной или устаревшей, а иногда и неверной. RuThes же создавался лингвистами, и он содержит экспертную информацию, которая организована в соответствии с теоретическими принципами построения онтологий и, следовательно, должна быть непротиворечивой, последовательной и корректной;

2) категории Википедии не полностью соответствуют иерархической структуре онтологии, которой является RuThes. Кроме того, они являются намного более дробными по содер-

жанию; страницы Википедии часто могут содержать техническую информацию или метаинформацию, не относящуюся к содержанию статьи, что затрудняет их анализ;

3) в RuThes, в отличие от Википедии, у многих ключей отсутствуют словарные статьи, вместо которых даны только списки синонимов.

В системе распознавания кореферентности использовалась информация о семантической близости двух именных групп. Под семантической близостью понимается то, насколько тесно сущности связаны между собой связями в онтологии. Меры семантической близости успешно используются для снятия семантической неоднозначности, создания вопросно-ответных систем, в том числе для разрешения кореферентности. С одной стороны, эта метрика достаточно прозрачно представляет степень сходства, вплоть до совпадения, между объектами, с другой стороны, она по определению представляется в численном виде, а потому легко может быть интегрирована в автоматические системы в качестве параметра.

Существует большое количество работ, предлагающих методы подсчета семантической близости между сущностями (см. [Budanitsky, Hirst, 2006; Крюков и др., 2010] для описания и сравнения различных метрик). В рамках эксперимента был проведен сравнительный анализ того, насколько меры семантической близости, подсчитанные несколькими различными методами, удовлетворяют задачам автоматического разрешения кореферентности.

Отобранные для сравнения меры разбивались на три больших класса в зависимости от принципа подсчета, на котором они основаны. К первому относились меры, основанные на иерархической структуре онтологии – опирающиеся на расстояние между сущностями. Основной из них является предложенная в статье [Rada et al., 1989] длина пути $path(c_1, c_2)$ между вершинами онтологии c_1 и c_2 , соответствующими искомым сущностям (далее в тексте *rada*).

В работе [Wu, Palmer, 1994] предлагается нормализовать данную меру с учетом максимальной глубины онтологии D :

$$wp(c_1, c_2) = -\log \frac{path(c_1, c_2)}{2D}.$$

Другой вариант нормализации, с учетом глубины узлов онтологии описан в статье [Leacock, Chodorow, 1998]:

$$lc(c_1, c_2) = \frac{2 * depth(lcs)}{depth(c_1) + depth(c_2)}, \quad (1)$$

где $depth(n)$ – глубина узла n ; lcs – ближайший общий родительский узел узлов c_1 и c_2 .

Второй класс составили меры, основанные на подсчете информационной содержательности (information content) понятия. Под информационной содержательностью понимается частота встречаемости понятия, которая подсчитывается как вероятность его появления в корпусе текстов. Термин впервые встречается в статье [Resnik, 1995], метод же подсчета семантической близости на основе этой величины описан в работе [Seco et al., 2004]. Авторы предлагают оценивать близость между узлами онтологии c_1 и c_2 по внутренней информационной содержательности (intrinsic information content) ближайшего к ним общего родительского узла:

$$res(c_1, c_2) = 1 - \frac{\log(hypo(lcs) + 1)}{\log(\max_{wn})},$$

где lcs – общий родительский узел (см. (1)); $hypo(n)$ – число гипонимов узла n ; $\max(w_n)$ – число узлов онтологии.

Третий класс был представлен мерой текстовых пересечений (text overlaps), предложенной М. Леском [Lesk, 1986] для подсчета близости между словарными определениями сущностей. Модификация этой меры для расширенных текстовых пересечений (extended text

overlaps), предназначенная для подсчета семантической близости между статьями Википедии, описана в статье [Banerjee, Pedersen, 2003].

Ее вычисление основано на количестве пересечений различной длины между текстами, в случае Википедии между первым абзацем статьи и ее полным текстом: $overlap(t_1, t_2) = \sum_n m^2$, для m n -словных пересечений между текстами t_1 и t_2 . В статье [Ponzetto, Strube, 2006] предлагается метод нормализации данной метрики с учетом длины сравниваемых текстов:

$$lesk(t_1, t_2) = \tanh\left(\frac{overlap(t_1, t_2)}{length(t_1) + length(t_2)}\right).$$

После того, как были выбраны источники и способ представления данных, было необходимо оценить, насколько полученная информация может быть использована для разрешения кореферентности. Оценка репрезентативности пяти выбранных мер (*rada*, *wp*, *lc*, *res*, *lesk*) была проведена на материале подкорпуса корпуса RuCor [Toldova et al., 2016], созданного в рамках проведения упоминавшегося выше форума RuEval-2014. Корпус RuCor использовался в качестве золотого стандарта для систем, участвовавших в соревновании, в связи с чем в нем вручную выделены все выражения, являющиеся кореферентными, что позволило использовать данные из него, не прибегая к дополнительной разметке. Целью эксперимента было сравнение того, насколько значения мер, полученные для пар сущностей, являющихся или не являющихся кореферентными, коррелируют с разметкой и, следовательно, могут использоваться для автоматического анализа.

Для анализа были взяты кореферентные цепочки из 55 текстов, входивших в тестовую выборку при проведении соревнования RuEval-2014, в которых в общей сложности было выделено 1016 кореферентных цепочек. В экспериментальный корпус вошло 200 пар кореферентных выражений из цепочек и такое же количество референциальных выражений, не входивших ни в одну кореферентную цепочку. Затем, используя Википедию и RuThes в качестве двух различных источников данных, для каждой пары именных групп мы подсчитали значения пяти описанных выше метрик.

Как нижний порог оценки качества для каждой пары (i, j) была подсчитана мера Жаккара. Она вычислялась на основе количества результатов (*hits*) в выдаче поисковой системы Google по запросам, содержащим элементы пары i и j по отдельности и вместе:

$$jaccard = \frac{hits(i \text{ and } j)}{hits(i) + hits(j) - hits(i \text{ and } j)}.$$

В табл. 1 приведен пример значений всех метрик, подсчитанных для пары кореферентных и пары некорреферентных сущностей.

Наконец, для всех значений каждой из описанных выше метрик, вычисленных на материале обоих источников, был определен коэффициент корреляции Пирсона с корпусной разметкой. Для того чтобы осуществить сравнение было возможно, кореферентные пары условно считались имеющими максимальное значение соответствующей метрики, а некорреферентные – минимальное.

В табл. 2 приведены полученные значения коэффициента корреляции для оценивавшихся мер, включая базовое значение коэффициента Жаккара. Меры сгруппированы по авторам статей, в которых они были впервые предложены, и источнику семантических данных. Для Википедии приведены два набора мер: для всех сущностей в выборке и отдельно только для стимулов, включавших в себя именованные сущности. Мера *lesk* не могла быть подсчитана на данных из RuThes, поскольку глоссы в нем приведены лишь для очень небольшого числа сущностей.

Таблица 1

Пример подсчета метрик семантической близости
для пар референциальных выражений

Table 2

Example of calculating semantic relatedness measures for pairs
of referential expressions

Пара	<i>jaccard</i>	RuThes					Википедия				
		<i>rada</i>	<i>wp</i>	<i>lc</i>	<i>res</i>	<i>lesk</i>	<i>rada</i>	<i>wp</i>	<i>lc</i>	<i>res</i>	<i>lesk</i>
такса – собака	0,07	2	0,07	0,93	0,8	–	3	1,47	1	0,27	0,025
здание – каби- нет	0,06	5	0,03	0,89	0,74	–	5	1,17	0,77	0,73	0,031

Таблица 2

Значения коэффициента корреляции метрик семантической близости
с кореферентной разметкой

Table 2

Values for coefficient of semantic relatedness measures correlation
and coreference annotation

Источник	<i>jaccard</i>	<i>rada</i>	<i>wp</i>	<i>lc</i>	<i>res</i>	<i>lesk</i>
RuThes	0,34	0,56	0,51	0,59	0,30	n/a
Википедия	0,34	0,05	0,58	0,35	0,23	0,03
Википедия (именованные сущности)	0,6	0,7	0,08	0,6	0,2	0,2

Как видно из табл. 2, меры, подсчитанные на основе данных RuThes, в целом оказались более репрезентативными, чем меры, основанные на данных Википедии. Это может быть связано с особенностями структуры дерева категорий онлайн-энциклопедии, которое содержит не только иерархические отношения между сущностями, но и большое количество специальных и метакатегорий. Такая структура может приводить к тому, что путь в дереве между двумя некорреляционными сущностями оказывается значительно короче благодаря существованию общей для них метакатегории (например, статьи «Барак Обама» и «Билл Клинтон» будут иметь общую родительскую категорию «Президенты США» притом, что соответствующие сущности явно не являются корреляционными).

Меры, основанные на расстоянии между сущностями, как обычная мера *rada*, так и ее нормализованные варианты, оказались наиболее репрезентативными из рассмотренных. Мера информационной содержательности оказалась менее репрезентативной для данных из обоих источников. Одной из причин этого может являться то, что узлы онтологии более высокого уровня могут иметь меньше непосредственных гипонимов, чем узлы более низкого уровня. Представляется возможным бороться с этим, используя для подсчета метрики количество не только прямых гипонимов узла, но и всех узлов, лежащих ниже в онтологии. Таким образом, для сущностей, общий родительский узел которых лежит ниже в онтологии, значение меры будет выше, чем если такой узел будет находиться ближе к ее вершине.

Мера семантической близости, основанная на текстовых пересечениях, оказалась наименее репрезентативной из всех рассмотренных и не использовалась в дальнейшем. Низкие значения меры, свидетельствующие о небольшом объеме текстовых совпадений между статьями, могут быть связаны с различными особенностями стиля написания. Данный вопрос

выходит за рамки настоящей работы, но представляется интересным для тематических исследований.

Несмотря на более высокую степень корреляции мер, основанных на данных тезауруса, с кореферентной разметкой, из табл. 2 также видно, что меры близости, подсчитанные на данных Википедии отдельно только для тех стимулов, которые содержат именованные сущности, оказываются выше базового значения и вполне сравнимы со значениями мер, подсчитанных на данных RuThes. Этот факт оказывается важен в связи с тем, что в тезаурусе полностью отсутствуют данные о персоналиях, и доступная семантическая информация о них может быть взята только из Википедии. Таким образом, наилучшим способом использовать рассмотренные источники данных было совместить информацию из обоих для подсчета мер семантической близости, используя один из них в зависимости от конкретного типа референта.

Второй этап экспериментов

Заключительным этапом в описываемой в данной работе серии экспериментов стало создание системы автоматического разрешения кореферентности, имплементирующей рассмотренные выше меры семантической близости в качестве одного из параметров. Разработанная система использовала подход, основанный на машинном обучении, поскольку, как правило, именно этот класс систем показывает лучшие результаты на соревнованиях (см. [Pradhan et al., 2012; Toldova et al., 2014]). Система была создана на основе описанной в статье [Kutuzov, Ionov, 2014] модели, использующей в качестве алгоритма классификации дерева решений (decision trees).

Использованная система основана на алгоритме попарной классификации и включает в себя признаки, основанные на расстоянии между членами пары и их морфологическими характеристиками, но кроме того учитывает и синтаксические и простейшие семантические признаки, такие как проверки на то, является ли одна из групп в паре аппозитивной к другой или являются ли члены пары именами собственными. Итоговый набор признаков совпадает с 11 основными признаками, описанными в статье [Toldova, Ionov, 2017], кроме того, были имплементированы дистанционные и морфологические признаки из упомянутой статьи.

Эксперимент, как и предыдущий этап работы, проводился на основе корпуса RuCor, который для этого был разбит на два подкорпуса – обучающий и тестовый (70 и 30 % объема исходного корпуса соответственно). Оценка качества работы системы осуществлялась с помощью набора инструментов, использовавшегося в рамках конференции CONLL-2012 [Pradhan et al., 2012] и считающегося стандартом при анализе кореферентности. Для непосредственной оценки использовались две метрики: MUC [Vilain et al., 1995] и B³ [Bagga, Baldwin, 1998], являющаяся модифицированной версией первой. В соответствии с каждой метрикой измерялись точность, полнота и F-мера качества работы системы.

В ходе эксперимента были подсчитаны значения метрик качества для следующих версий системы: не использующей семантическую информацию (в табл. 3 *Soon*), опирающейся на семантические признаки только из одного источника и опирающейся на семантические признаки из обоих источников. В табл. 3 приведены значения метрик для всех четырех рассматривавшихся версий системы.

Из приведенных результатов видно, что в то время, как использование признаков на основе Википедии позволяет повысить полноту результатов, использование признаков на основе RuThes увеличивает точность распознавания связей. Таким образом, совмещение признаков, полученных из обоих источников, позволяет достигнуть максимального улучшения качества работы системы. Этим подтверждаются выводы, сделанные на основе второго этапа экспериментов, о полезности информации из Википедии, несмотря на более низкие параметры корреляции.

Таблица 3

Метрики качества для различных версий системы

Table 3

Quality metrics for different versions of the system

Система	MUC			В ³		
	Точность	Полнота	F-мера	Точность	Полнота	F-мера
Soon	72,76	59,49	65,46	71,01	44,50	54,71
Soon + Википедия	70,28	59,71	64,56	66,50	44,63	53,41
Soon + RuThes	72,72	59,43	65,41	71,15	44,44	54,71
Soon + RuThes + Википедия	73,57	60,01	66,10	71,77	44,93	55,26

Использование семантических признаков из обоих источников позволило улучшить распознавание кореферентности между именными группами, находящимися в нескольких классах отношений. С одной стороны, это таксономические связи, такие как гиперонимия или синонимия, в том числе между именованными сущностями (примеры (2)–(4)).

- (2) Сразу же с момента издания [альбом] стал настолько популярным, <...> что звукозаписывающая компания ЕМІ не успевала штамповать копии [пластинки] для всех желающих ее приобрести.
- (3) Выжившие после крушения [корабля] рассказали, что основная причина трагедии – то, что [теплоход] был очень старый.
- (4) Альбом первоначально был издан в 1975 году компаниями Harvest Records в [Великобритании] и Columbia Records в США. <...> в 1980 году альбом появился в [Соединённом Королевстве] с более высоким качеством...

С другой стороны, это отношения между именованной сущностью и именной группой, обозначающей один из аспектов этой сущности (пример (5)).

- (5) Виктор Вексельберг хотел бы ангажировать [Григория Перельмана] для работы в Кремниевой долине. Фортуна повернулась к [питерскому математику] лицом: сначала мировое признание, затем миллион долларов.

Продемонстрированный системой рост качества превышает результат, полученный в аналогичном эксперименте, описанном в статье [Toldova, Ionov, 2017]: максимальный прирост в 0,54 % MUC и 0,55 % В³ по сравнению с 0,26 и 0,19 % соответственно. Как можно заметить, наибольшее увеличение качества в обоих экспериментах дает использование информации

из RuThes. Поскольку в данном исследовании семантическая информация представлена в виде мер семантической близости в отличие от простого расстояния в онтологии, можно заключить, что более тщательная обработка семантической информации, в частности в форме подобных метрик, является более выигрышным способом представления семантической информации и позволяет в большей степени улучшить работу системы.

Поскольку описанная выше система использует для распознавания кореферентных связей тезаурусную и энциклопедическую информацию, за пределами ее возможностей на данный момент остаются случаи контекстной синонимии, опирающиеся на содержание текста, подобно примеру (6). Для таких случаев представляется необходимым использовать другие признаки, основанные на семантическом анализе текста или анализе совместной встречаемости слов. Подбор и оценка эффективности этих признаков, однако, находятся за рамками данной работы и должны стать предметом отдельного исследования.

- (6) Выходец из [Нигерии] решил остаться на ПМЖ в Израиле, поскольку на [родине] его якобы преследует опасный призрак.

Заключение

В данной работе описан ряд прикладных лингвистических экспериментов, целью которых было, во-первых, продемонстрировать важность семантической информации из энциклопедических источников для разрешения кореферентности и, во-вторых, создать работающую автоматическую систему для данной задачи с интегрированной тем или иным способом семантической информацией.

В ходе первых экспериментов было проведено сравнение различных мер семантической близости между референциальными выражениями, полученных на основе данных Википедии и тезауруса RuThes, с целью сравнения их пригодности для использования в системах распознавания кореферентности. Полученные результаты показали бóльшую релевантность данных RuThes в целом, в то время как данные Википедии оказалось возможным использовать для улучшения разрешения кореферентных связей именованных сущностей. Вторая серия экспериментов была нацелена на оценку качественного улучшения работы системы после имплементации разработанных ранее признаков. В результате разработана работающая система автоматического анализа кореферентности, включающая в себя семантическую информацию. Качество ее работы оказалось выше, чем у аналогичных систем, не использующих семантические признаки и использующих их более простые представления. Таким образом, подтверждается актуальность результатов предыдущего этапа экспериментов.

Важность результатов, полученных в результате проведенной работы, состоит в следующем:

- 1) экспериментально подтверждено повышение качества автоматического распознавания кореферентности в русском языке при использовании семантической информации из внешних источников;
- 2) проведены подсчет и сравнение мер семантической близости для русского сегмента Википедии и тезауруса RuThes, что позволило определить более репрезентативные меры такого рода и указало на ряд особенностей структуры Википедии, негативно влияющих на результаты их вычисления;
- 3) оценен вклад мер семантической близости в качество распознавания кореферентности в сравнении с другими признаками, основанными на семантической информации;
- 4) создана работающая система автоматического разрешения кореферентности с интегрированными семантическими признаками, доступная для изучения и улучшения.

Данная работа открывает несколько различных направлений дальнейших исследований. С одной стороны, существует большой объем задач по более тщательной оценке мер семантической близости для онтологий на русском языке. Существует также большое количество не рассмотренных в данной работе метрик, в том числе специализированных алгоритмов для оценки близости сущностей в Википедии (в частности, алгоритм, описанный в [Yeh et al., 2009]). Кроме того, описанные результаты могут открыть путь к дальнейшему изучению способов интеграции семантической информации в системы автоматического распознавания кореферентности как с точки зрения альтернативных источников данных, так и с точки зрения альтернативных способов их представления. Необходимо также учесть существование случаев контекстной синонимии, которые не могут быть распознаны за счет энциклопедической и тезаурусной информации и требуют иного подхода к анализу.

Наконец, интерес для теоретических исследований могут представлять затронутые в данной статье особенности структуры русской Википедии как онлайн-энциклопедии и ее возможные различия по сравнению как с традиционными энциклопедиями, так и с другими источниками структурированных данных: тезаурусами, такими как RuThes, или иноязычными сегментами Википедии.

Список литературы / References

- Крюков К. В., Панкова Л. А., Пронина В. А., Суховеров В. С., Шипилина Л. Б.** Меры семантической близости в онтологии // Проблемы управления. 2010. Вып. 5. С. 2–14.
Kryukov, K. V., Pankova, L. A., Pronina, V. A., Sukhoverov, V. S., Shipilina, L. B. Semantic similarity measures in ontology. Review and classification. *Control Sciences*, 2010, iss. 5, p. 2–14. (in Russ.)
- Лукашевич Н. В.** Тезаурусы в задачах информационного поиска. М.: Изд-во Моск. ун-та, 2011.
Loukachevitch, N. Thesauri in Information retrieval tasks. Moscow, Moscow State University Publ., 2011. (in Russ.)
- Толпегин П. В.** Автоматическое разрешение кореференции местоимений третьего лица русскоязычных текстов: Дис. ... канд. техн. наук. М., 2008.
Tolpegin, P. V. Automated resolution of third person pronouns coreference in Russian texts. PhD dissertation. Moscow, 2008. (in Russ.)
- Azerkovich, I.** Employing Wikipedia data for coreference resolution in Russian. *Artificial Intelligence and Natural Language. AINL 2017. Series: Communications in Computer and Information Science*, 2018, vol. 789, p. 107–112.
- Bagga, A., Baldwin, B.** Algorithms for scoring coreference chains. In: The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference, 1998, vol. 1, p. 563–566.
- Banerjee, S., Pedersen, T.** Extended gloss overlaps as a measure of semantic relatedness. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc., 2003, p. 805–810.
- Bogdanov, A. V., Dzhumaev, S. S., Skorinkin, D. A., Starostin, A. S.** Anaphora analysis based on ABBYY Compreno linguistic technologies. In: Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2014, vol. 13 (20), p. 89–102.
- Budanitsky, A., Hirst, G.** Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 2006, vol. 32 (1), p. 13–47.
- Haghighi, A., Klein, D.** Simple coreference resolution with rich syntactic and semantic features. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2009, vol. 3, p. 1152–1161.
- Harabagiu, S. M., Bunescu, R. C., Maiorano, S. J.** Text and knowledge mining for coreference resolution. In: Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies. Association for Computational Linguistics, 2001, p. 1–8.
- Kamenskaya, M. A., Khramoin, I. V., Smirnov, I. V.** Data driven methods for anaphora resolution of Russian texts. Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2014, vol. 13 (20), p. 241–250.
- Kameyama, M.** Recognizing referential links: an information extraction perspective. In: Proceedings of the ACL’97/EACL’97 workshop on Operational factors in practical, robust anaphora resolution. Madrid, Spain, 1997, p. 46–53.
- Kutuzov, A. B., Ionov M.** The impact of morphology processing quality on automated anaphora resolution for Russian. In: Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2014, vol. 13 (20), p. 232–240.
- Leacock, C., Chodorow, M.** Combining local context and WordNet similarity for word sense identification. In: WordNet. An Electronic Lexical Database. MIT Press, 1998, p. 265–283.

- Lesk, M.** Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In: Proceedings of the 5th Annual Conference on Systems Documentation. ACM, 1986, p. 24–26.
- Ng, V., Cardie, C.** Improving machine learning approaches to coreference resolution. In: Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002, p. 104–111.
- Ponzetto, S. P., Strube, M.** Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics, 2006, p. 192–199.
- Pradhan, S., Moschitti, A., Xue, N., Uryupina, O., Zhang, Y.** CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012, p. 1–40.
- Rada, R., Mili, H., Bicknell, E., Blettner, M.** Development and application of a metric to semantic nets. In: IEEE Transactions on Systems, Man and Cybernetics, 1989, iss. 19 (1), p. 17–30.
- Rahman, A., Ng, V.** Coreference resolution with world knowledge. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2011, vol. 1, p. 814–824.
- Resnik, P.** Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. IJCAI, 1995, vol. 1, p. 448–453.
- Seco, N., Veale, T., Hayes, J.** An intrinsic information content metric for semantic similarity in WordNet. In: Proceedings of the 16th European Conference on Artificial Intelligence. IOS Press, 2004, p. 1089–1090.
- Sharoff, S., Nivre, J.** The proper place of men and machines in language technology: Processing Russian without any linguistic knowledge. In: Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2011, vol. 10 (17), p. 591–605.
- Soon, W. M., Ng, H. T., Lim, D. C. Y.** A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 2001, vol. 27, no. 4, p. 521–544.
- Toldova, S. Ju. Roytberg, A., Ladygina, A. A., Vasilyeva, M. D., Azerkovich, I. L., Kurzukov, M., Sim, G., Gorshkov, D. V., Ivanova, A., Nedoluzhko, A., Grishina, Y.** Ru-Eval-2014: Evaluating anaphora and coreference resolution for Russian. In: Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2014, vol. 13 (20), p. 681–694.
- Toldova, S., Grishina, Yu., Ladygina, A., Vasilyeva, M., Sim, G., Azerkovich, I.** Russian coreference corpus. In: Almeida F. A., Barrera I.O., Toledo E. Q. (eds.). Input a Word, Analyze the World: Selected Approaches to Corpus Linguistics. Cambridge Scholars Publishing, 2016, p. 107–124.
- Toldova, S., Ionov, M.** Coreference Resolution for Russian: The Impact of Semantic Features. In: Computational linguistics and intellectual technologies: Proceedings of the international conference “Dialogue 2014”. Moscow, 2017, vol. 16 (23), p. 339–348.
- Vilain, M., Burger, J., Aberdeen, J., Connolly, D., Hirschman, L.** A model-theoretic coreference scoring scheme. In: Proceedings of the 6th Conference on Message Understanding, ser. MUC6 '95. Stroudsburg, PA, USA: Association for Computational Linguistics, 1995, p. 45–52.
- Wu, Z., Palmer, M.** Verb semantics and lexical selection. In: Proceedings of ACL-94, 1994, p. 133–138.

Yeh, E., Ramage, D., Manning, C. D., Agirre, E., Soroa, A. WikiWalk: random walks on Wikipedia for semantic relatedness. In: Proceedings of the 2009 workshop on graph-based methods for natural language processing. Association for Computational Linguistics, 2009, p. 41–49.

Материал поступил в редколлегию

Date of submission

08.09.2018

Сведения об авторе / Information about the Author

Азеркович Илья Леонидович, аспирант аспирантской школы по филологическим наукам факультета гуманитарных наук Национального исследовательского университета «Высшая школа экономики» (ул. Ст. Басманная, 21/4, Москва, 105066, Россия)

Ilya L. Azerkovich, National Research University Higher School of Economics (21/4 Staraya Basmannaya Str., Moscow, 105066, Russian Federation)

iazerkovich@gmail.com

SPIN-код 7211-6400

ORCID 0000-0002-6482-1137

Подходы к составлению лексических минимумов в России и за рубежом: проблемы и перспективы

Н. А. Муравьев, М. Ю. Ольшевская

*Национальный исследовательский университет «Высшая школа экономики»
Москва, Россия*

Аннотация

Представлен сравнительный обзор подходов, которые практиковались в российской традиции составления лексических минимумов русского языка, а также в зарубежных традициях, среди которых особое внимание уделено англо- и немецкоязычной традиции. Цель обзора – проследить направление развития науки в области составления лексических списков на стыке лексикографии и лингводидактики и определить критерии, на которые должны ориентироваться составители списков, чтобы отвечать запросам пользователей в рамках различных задач. Приводится определение понятия «лексический минимум», и задаются параметры сравнения лексических минимумов в последующем обзоре. Рассматриваются отечественные подходы к составлению списков с фокусом на лингводидактической проблематике, лежащей в основе всего русскоязычного направления. Обсуждаются зарубежные подходы, основное внимание уделяется спискам общего пользования и проблеме выделения лексического ядра. Сравниваются отечественная и зарубежные традиции. Приведенный в статье обзор дает основания утверждать, что вне зависимости от предназначения лексического минимума создание его требует сочетания как статистического, так и коммуникативно-ориентированного метода. Кроме того, для составления актуального и достоверного корпуса необходимо равномерное соотношение составных частей анализируемых данных: кроме корпуса текстов художественной литературы следует обратиться к данным устного корпуса и к таким различным по стилю и жанру источникам, как газетный, художественный и академический корпусы, а также корпус интернет-речи.

Ключевые слова

языковые минимумы, список слов, корпусная лингвистика, русский язык, критерии отбора лексического минимума, дисперсия слов, частотность, дисперсия, покрытие

Для цитирования

Муравьев Н. А., Ольшевская М. Ю. Подходы к составлению лексических минимумов в России и за рубежом: проблемы и перспективы // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 78–89. DOI 10.25205/1818-7935-2019-17-1-78-89

Approaches to the Composition of Lexical Minima in Russia and Abroad: Problems and Prospects

Nikita A. Muravyev, Maria Yu. Olshevskaya

*National Research University Higher School of Economics
Moscow, Russian Federation*

Abstract

This paper presents an overview of the existing approaches in Russia and abroad to the compilation of minimal vocabulary lists. Special attention is paid to the English, as well as to the German scientific tradition. The purpose of the overview is to track and compare the underlying lexicographical and lingvo-didactic trends from the beginning of the 20th century until now and to define the criteria for making a list that would match the expectations of the modern user. The first section of the article provides a definition of the notion *lexical minimum* and introduces the parameters of comparison for the wordlists under discussion. By lexical minimum we understand not only foreign learners' list but more broadly any wordlist which is compiled by minimization of the lexicon on the basis of statistical, pragmatic or mixed criteria. The wordlists are compared across four parameters: purpose of the list (general service list,

theoretical, lingvo-didactic), approach to compilation (statistical, pragmatic, mixed), sources of data (corpora, textbooks, questionnaires etc.) and text coverage in percent. The second section discusses the existing approaches to the lexical minima compilation in Russia with emphasis on pedagogical aspects which are prominent in the Russian tradition. The second section discusses the approaches used in German and English lexicographical traditions focusing on the problem of defining core vocabulary and compiling general service lists. The closing section of the article compares Russian and foreign traditions and summarizes the overview. The present overview suggests that the creation of minimal vocabulary lists requires a combination of both statistical and communicatively oriented methods. In addition to that, given the recent development of large corpora the new challenge arises as to provide a stylistically diverse and balanced corpus or a number of corpora that would serve as a proper vocabulary basis for a vocabulary list. Thus, in addition to the fiction texts, authors should include oral corpus data, as well as newspaper, art and academic sources, and internet speech.

Keywords

lexical minima, word list, corpus linguistics, Russian language, word list criteria, word dispersion, frequency, distribution, lexical overlap

For citation

Muravyev, Nikita A., Olshevskaya, Maria Yu. Approaches to the Composition of Lexical Minima in Russia and Abroad: Problems and Prospects. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 78–89. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-78-89

Введение

Концепция лексического минимума как набора наиболее значимых лексем языка, используемого в разнообразных практических целях, в том или ином виде существует уже не одно столетие. Подобные списки составлялись для разных теоретических и прикладных областей: образования, лексикографии, стенографии, философии языка, академического письма и многих других. По мере эволюции данной практики неоднократно менялось представление о том, какие лексемы языка являются значимыми и как правильно определять эту значимость в зависимости от поставленной задачи. Одни считают, что объективной мерой значимости является частотность слов в тексте, другие определяют значимость субъективно, но при этом руководствуются практическими соображениями вроде тематической репрезентативности, удобства представления, а третьи придерживаются промежуточных взглядов. Поэтому все многообразие подходов нуждается в осмыслении, сопоставлении и оценке.

Данная статья представляет собой обзор существующих подходов к составлению лексических минимумов в российской и в зарубежных традициях, среди которых особое внимание будет уделено наиболее влиятельным англо- и немецкоязычной традиции. Цель обзора – проследить направление развития науки в области лексических минимумов и определить наиболее значимые критерии для их составления.

Определение лексического минимума

Лексические минимумы имеют различную структуру и состав и создаются с самыми разными целями. Поэтому, для того чтобы иметь возможность сравнивать такие списки между собой, прежде всего, необходимо задать рамки того, что входит в это понятие, и определить параметры сравнения, релевантные для всех рассматриваемых списков. В своем изначальном, узком понимании лексический минимум представляет собой небольшого объема список слов, которые необходимо знать изучающим иностранный язык на определенном уровне владения языком. Чаще всего при этом слова для списка отбираются исходя из субъективных представлений составителей о коммуникативных потребностях изучающих иностранный язык. Такого рода списки составляют основу русскоязычной традиции, хотя в небольшом объеме встречаются и в зарубежных традициях (ср. [Bakonyi, 1934; Thorndike, Lorge, 1944; Carrol et al., 1971]). Параллельно с этим в зарубежной лексикографии возникает и получает развитие отдельная традиция составления списков «общего пользования», отражающих наиболее употребительный пласт лексики, или лексическое ядро языка (ср., к примеру, [West, 1953; Leech et al., 2001; Brezina, Gablasova, 2015]). Эти списки составляются в результате от-

бора, но отбор происходит не на коммуникативной, а на статистической основе с использованием коллекций текстов и полноценных корпусов. Такие списки в строгом смысле нельзя назвать лексическими минимумами, поскольку содержащаяся в них лексика не является минимально необходимой для каких-либо практических целей. Однако в целом, поскольку в эпоху больших данных составление списков без учета статистики фактически теряет смысл, для адекватной оценки текущей ситуации в области составления лексических минимумов исключительно необходимо учитывать и второй вид списков. По этой причине в данной статье предлагается расширенное понимание лексического минимума (далее ЛМ) как *любого списка лексики, полученного в процессе минимизации словарного состава языка ручным или автоматическим методом с учетом коммуникативных, статистических или смешанных критериев*.

Поскольку множество таких списков весьма разнородно, мы будем опираться на перечень параметров, которые будут служить ориентиром для обсуждения и сравнения рассматриваемых ЛМ. В число предлагаемых параметров входят следующие:

- предназначение (общего пользования, теоретические, учебные);
- подход к составлению (коммуникативный, статистический, смешанный);
- источники данных (корпус письменных текстов, материал опроса носителей, прочее);
- процент покрываемости текстов.

Первый параметр различает списки лексики, создаваемые для общего пользования в качестве теоретической модели и в помощь изучающим язык как иностранный. Второй параметр подхода к составлению различает опору на коммуникативные критерии (субъективные, тематические), на статистические критерии (объективные измерения частотности, дисперсии слов и др.) и комбинированные подходы. Следующий параметр характеризует источники данных, на основе которых строится список, а именно: текстовые корпуса, словари, учебная литература, опросники и прочее. Наконец, некоторые составители списков измеряют валидность полученного списка, тестируя степень покрытия их списком произвольного текста. Наличие таких данных также будет учитываться в настоящем обзоре.

Подходы к составлению лексических минимумов в России

Отечественными исследователями с 1950-х гг. предлагалось порядка 18 разных принципов отбора и формирования ЛМ, как, например, принцип стилистической нейтральности слова, сочетаемости, семантической и словообразовательной ценности, частотности и др. [Маркина, 2011]. В 1960–1980-е гг. И. В. Рахманов, И. Д. Салистра, В. Г. Костомаров, Э. А. Штейнфельдт, В. В. Морковкин, П. Н. Денисов разрабатывали и другие критерии отбора лексики. В результате основой для создания существующих ЛМ стали такие критерии, как частотность, сочетаемость, стилистическая нейтральность, семантическая и словообразовательная ценность, многозначность, и русская лексикографическая традиция опиралась на них.

В. В. Морковкин полагал, что основные проблемы при создании ЛМ сводятся к определению количества слов, источника их отбора и реализации процедуры по отбору [Морковкин, Дорогонова, 1976]. Вокруг критериев отбора релевантной лексики создания ЛМ развернулась полемика. Ю. А. Марков считал принцип частотности абсолютным при составлении ЛМ, а В. Г. Костомаров, В. В. Морковкин говорили о недостаточности критерия частотности как ведущего принципа создания ЛМ и необходимости учитывать практическую значимость, потребности пользователя. В. Г. Костомаров указывал, кроме того, что начиная со второй тысячи слов сложно провести границу по частотности, потому что все слова становятся примерно одинаково частотны [1963].

Еще одна из актуальных проблем при выделении ЛМ состоит в том, что слова, необходимые в повседневном общении, не всегда являются частотными. Поэтому один из первых ЛМ, созданный в 1930 г. для английского языка, не содержит таких слов, как «до свидания», «спасибо», «большой», «никогда», «хотеть» [Маркина, 2011]. В. В. Морковкин предлагал

решить эту проблему введением «субъективного» критерия отбора лексики путем проведения опроса среди школьников на предмет выделения самых необходимых слов в повседневном общении [Морковкин, Дорогонова, 1976].

Говоря об определении объема списка, также можно увидеть различие в подходах. В зарубежной лексикографии в связи с понятием коммуникативной компетенции в основу создания «Общеввропейской компетенции владения языком» легла концепция Я. А. Ван Эка [van Ek, 1987; Трушина, 2013]. Согласно данной концепции, объем словаря начального уровня зависит от целей общения на начальном этапе (для английского ЛМ это умение рассказать о себе, семье, ближайшем окружении, понимать вывески, объявления, открытки, заполнять анкеты) и составляет 120 слов (survival list), а для повседневного общения потребуется порядка 2 000 слов.

Сходным образом представлен и круг тем в ЛМ русского языка для элементарного владения [Владимирова и др., 2001]: «Рассказ о себе», «Работа», «Учеба», «Рабочий день», «Свободное время», «Отдых», «Семья». Слова, вошедшие в минимум, дают также возможность общаться на такие темы, необходимые для уровня выживания в иноязычной среде, как «Питание», «Здоровье», «Погода», «Транспорт», «Покупки», «Город, ориентация в городе». Он включает также ограниченный объем лексики, описывающей атрибуты учебного процесса (преподаватель, учебник и др.), которая менее значима для повседневного общения, но является неотъемлемой частью обучения иностранному языку. Кроме того, в минимуме представлены различные формулы речевого этикета, необходимые изучающим русский язык как иностранный при решении актуальных коммуникативных задач.

Кроме того, существуют разногласия по поводу выбора источников текстов для создания ЛМ. Основное противоречие состоит в том, что для создания ЛМ русского языка ранее брались словари частотности (например, [Штейнфельдт, 1963; Засорина, 1977; Лёнгрен, 1993]), толковые и переводные словари, списки лексики к учебникам (см., например, словарь «2 380 слов, наиболее употребительных в русской разговорной речи», составленный коллективом авторов Университета дружбы народов им. П. Лумумбы), а не собственно тексты из художественной литературы, СМИ и т. п. [Морковкин, Дорогонова, 1976].

По мнению Е. И. Маркиной [2011], практика использования НКРЯ до недавнего времени в значительной мере недооценивалась составителями ЛМ. Вместе с тем была произведена попытка создать ЛМ с помощью автоматизированных систем для выделения терминологического словника по методологии А. А. Поликарпова. В результате с помощью морфоанализатора DicTUM, разработанного в Лаборатории компьютерной лексикологии и лексикографии филологического факультета МГУ им. М. В. Ломоносова (О. В. Кукушкина, А. А. Поликарпов), было отобрано по 100 слов для двух специальностей [Маркина, 2011].

Из современных подходов можно отметить совместную работу филологического факультета Барселонского университета и филологического факультета МГУ им. М. В. Ломоносова. Коллективом предпринята попытка создания иллюстративного интерактивного ЛМ. Этот проект разрабатывается на базе существующих лексических минимумов ТРКИ для испанских студентов, изучающих русский язык в рамках специализации «Славянская филология» [Дунаева, Соррилья, 2010]. Ресурс предполагает работу с лексикой на различных уровнях: семантическом, синтаксическом, грамматическом. Для каждого слова в этом ЛМ указывается ударение, произношение, лексическая, толковая, графическая иллюстрации, а также перевод. Ресурс представлен серией заданий, направленных на усвоение обучающимся лексики на основных уровнях, включающих словосочетание и текст. Данный ресурс, несомненно, представляет большой интерес для учебных целей, однако не совершенствует существующий ЛМ, поскольку список слов ресурса базируется на существующих ЛМ русского языка как иностранного.

Говоря о новом типе ЛМ русского языка, В. В. Морковкин предлагал считать «единицей учета при освоении лексики... не слово как таковое, а слово в определенном значении» [Морковкин, Дорогонова, 1976]. Вместе с тем в российских ЛМ в качестве единицы описа-

ния по-прежнему выступает слово в совокупности его значений. В некоторых списках, составленных европейскими авторами (например, в [West, 1953]), единицей описания является не отдельное слово, а так называемая «словарная семья» (= word family – базовое слово во всех его значениях, в том числе в разных частеречных значениях) и его дериваты, которые учащиеся могут понять, опираясь на знание правил словообразования. Тем не менее в силу слишком разной употребительности дериватов одной лексемы и по некоторым другим причинам эта концепция так и не получила широкого распространения.

В. В. Морковкин предлагал также идею создания универсального ЛМ, состоящего из двух частей: 1) списка слов, отобранных на основе частотных словарей, а также на основе «интуиции» составленного тематического «распространителя» алфавитного списка с нумерацией тем и пометой к каждому слову; 2) материалов для поаспектной работы над русской лексикой, «трудностей» русского языка, включающих синонимы, антонимы, паронимы, омонимы, несвободные сочетания и т. д. Слова из первого списка автор предлагал считать обязательными для усвоения, а из второго – факультативными. Такое распределение позволило бы избежать таких пробелов, как наличие в основном корпусе слова «малина» при отсутствии слова «смородина», хотя оба они принадлежат к одной тематической группе [Морковкин, Дорогонова, 1976]. Однако в существующем ЛМ элементарного уровня русского языка [Андрюшина, Козлова, 2006] такое разделение не очевидно: в описании ЛМ нет информации, все ли слова из списка являются одинаково важными и учитываемыми при тестировании элементарного уровня (ТЭУ). Тем не менее, проанализировав данный ЛМ, можно увидеть, что в общих чертах он соответствует критериям В. В. Морковкина. ЛМ состоит из 3 частей, содержит алфавитный список частотных слов, тематические группы, последняя часть издания – приложения, о которых писал В. В. Морковкин.

Таким образом, можно сказать, что в отечественной лексикографии проблема выделения ЛМ до сих пор является актуальной. Среди основных спорных вопросов остаются, на наш взгляд, следующие:

- проблема отбора слов для ЛМ (количество, способ отбора, источники отбора);
- круг предметно-тематических областей («тем»), которые соотносятся со списком важнейших слов для владения русским языком на элементарном (начальном) уровне;
- необходимость отнесения частотности к ведущим критериям отбора слов в ЛМ.

Практика составления ЛМ в России оказывается при этом почти полностью ориентированной на лингводидактическую задачу обучения русскому языку как иностранному. В качестве основных их источников используются преимущественно ранее составленные ЛМ и частотные списки, которые модернизируются и дополняются данными анкетных опросов носителей и другими вспомогательными материалами.

Учитывая, что появление первых теоретических работ по вопросу создания ЛМ в отечественной лексикографии относится к 1950-м гг., а в зарубежной науке – к 1930-м гг., обратимся далее к истории вопроса за рубежом.

Подходы к составлению лексических минимумов за рубежом

Попытки создания ЛМ были предприняты в рассматриваемых европейских странах (Англия, Франция, Германия) уже в 20–30-х гг. XX в.

В 1930 г. Ч. К. Огден создал словарь базовой английской лексики Dictionary of Basic English, состоящий из 850 слов [Ogden, 1930]. Концепция данного словаря основана не на принципе частотности, а на идее поиска минимального лексикона, достаточного для выражения главных идей о мире. Данный ЛМ позиционировался автором как мини-модель языка, способная заменить естественный язык в качестве средства общения. Другими словами, это был искусственный язык, созданный автором из лексических единиц, отобранных для пояснения основной информации о мире.

Впоследствии в словаре было отмечено два существенных недостатка: 1) не разрешается проблема многозначности отобранных слов (было подсчитано, что 850 единиц, вошедших

в минимум Огдена, имеют более 12 000 значений; 2) отсутствует целый пласт слов, необходимых для повседневного общения, но не выражающих каких-либо «идей» о мире (например, «до свидания», «спасибо», «большой», «никогда», «хотеть» и др.).

В связи с критикой изданного словаря в начале 1930-х гг. М. Уэст начал разработку другой концепции лексического минимума, связанной с преподаванием английского языка как иностранного, – Reading Movement. Находясь на службе в Индии, М. Уэст интуитивно отбирал слова для обучающихся, необходимые индийцам для чтения литературы на английском языке. Вручную ему удалось отобрать около 1 800 слов. В 1934–1935 гг. он доложил итоги своей работы на международной конференции лингвистов, где выступили также другие ведущие специалисты в области преподавания английского языка: Г. Палмер, Э. Торндайк, Л. Фосетт. Ими был составлен доклад под названием *Interim Report on Vocabulary Selection* (1936), в котором эти ученые предложили избрать в качестве ведущего принципа отбора минимума лексики для освоения ее иностранцами критерий частотности. Кроме того, были выделены следующие критерии составления минимизированного словника: 1) структурная ценность (включались строевые слова языка); 2) тематическая широта (исключались термины); 3) стилистическая нейтральность (исключались разговорные слова и сленг); 4) наличие слов-определителей (включались слова, необходимые для формулирования дефиниций в словарях); 5) словообразовательная способность [Маркина, 2011].

Впоследствии на основе этих критериев в 1953 г. М. Уэст подготовил новый список [West, 1953], насчитывавший 2 000 слов с дериватами. Процесс выбора слов опирался на опыт членов комитета и включал как объективные, так и субъективные критерии. Объективный отбор был основан главным образом на списке 1 500 самых частотных слов английского языка [Fauccett, 1936], в то время как субъективные критерии были основаны на списке слов Палмера [Palmer, 1931], с одной стороны, и словаре Уэста [West, 1927], с другой, были включены слова со служебными функциями, а именно предлоги, местоимения, вспомогательные глаголы, союзы и определители. Также были включены слова, использование которых не ограничивалось определенным временем или местом (универсальность), а слова, используемые в ограниченных областях (религия, моральные понятия, собственные имена), были исключены. Слова, которые были сочтены полезными для личного и профессионального развития учащихся в возрасте от 12 до 18 лет, относились к широким предметным областям и потому вошли в список. К каждому слову были сделаны такие пометы, как отнесенность к частям речи, LCB и иллюстративный пример, а также указывалось процентное соотношение употребимости LCB в конкретном значении, что и стало отличительной особенностью *General Service List* (GSL).

Одним из базовых критериев при создании данного списка был параметр простоты лексики, отбираемой для коммуникации на начальном уровне владения языком, в результате чего отобран минимальный набор простой лексики. Однако этот критерий оказался ненадежным, поскольку слова, выделенные авторами, не всегда являлись частотными, и был необходим дополнительный объективный критерий, которым стал критерий покрываемости текстов предлагаемым списком слов. Теперь считается, что список может быть признан педагогически эффективным и надежным только при большом проценте покрытия текстов (порядка 95%). Методы, положенные в основу данного списка, до сих пор активно используются в европейской учебной лексикографии.

Дидактического направления придерживались и другие составители частотных лексических списков того времени (см. [Thorndike, Lorge 1944; Lorge 1949; Carroll et al., 1971] для английского, а также [Morgan, 1928; Bakonyi, 1934] для немецкого языка). В первый период списки составлялись на материале образовательных текстов для школьников, а с течением времени стали использоваться газетные статьи и другие печатные издания.

Основной проблемой этих и других ЛМ того времени был, прежде всего, малый объем корпусов и поверхностное применение статистики, причиной чего был, в частности, недостаток качественных вычислительных инструментов. В расчет бралась почти исключительно

абсолютная частотность слов, которая напрямую зависит от словарно-тематического состава конкретного корпуса. С чисто лингвистической точки зрения проблемой этих списков был чрезмерный крен в сторону существительных по сравнению с другими частями речи (аналогичная картина наблюдается и в более поздних статистических подходах), а также игнорирование целостности семантических полей. К примеру, в список могла попасть лишь часть наименований дней недели или месяцев. Кроме того, такие списки, как видно из приведенных выше рекомендаций, были плохо приспособлены для образовательных целей и не могли использоваться в качестве ЛМ для изучающих иностранный язык.

В то же время, как отмечается в [Маркина, 2011], все больше исследователей осознают необходимость введения содержательных критериев, которые заменили бы или, по меньшей мере, дополнили критерий частотности. Так, в 1950–1960-е гг. французскими лингвистами Ж. Гугенеймом и Р. Мишеа был предложен принцип «присутствия в сознании» (фр. *disponibilité*, англ. *availability*). Данный принцип предполагал отбор такой лексики, которая была бы полезной в повседневной жизни, поскольку она неизбежно возникает в ситуациях конкретного общения: «Мы используем слова *autobus* (автобус), *bouton* (пуговица, кнопка), *fourchette* (вилка) только тогда, когда они вызываются темой разговора» [Гугенейм, 1961]. Данный список был сформирован с помощью опроса школьников: в каждой из 16 основных тем (части тела, одежда, дом, продукты питания и т. п.) было необходимо отобрать по 20 необходимых лексических единиц. В результате обработки статистических данных был создан *Dictionnaire Fondamental de la Langue Française* [Gougenheim, 1958].

В немецкой лексикографии появляются списки, имеющие статистическую основу, но при этом также ориентированные на коммуникативные потребности пользователей. Достигалось это введением дополнительных мер, таких как включение экспертами менее частотных слов для заполнения лакун в семантических полях или устранение «лишних» частотных слов. Одним из ярких примеров применения такого подхода является список из 1 084 лемм (595 000 словоформ) Й. Пфеффера [Pfeffer, 1970]. Список был составлен автором на основе собранного им корпуса объемом 650 000 словоупотреблений, содержащего записи устной речи на повседневные темы от носителей из ФРГ, ГДР, Швейцарии и Австрии, и дополнен списком наиболее важных существительных, глаголов и прилагательных на материале опроса школьников. В списке также присутствуют собственные дополнения слов, которые, по мнению автора, должны завершить минимально необходимый лексический ряд, такие как *луна* в ряду слов *солнце* и *звезды* или слова *мотор* и *руль* в дополнение к слову *мотоцикл*.

С начала 2000-х гг. с ростом объема корпусов и формированием репрезентативных подкорпусов устной и письменной речи и частных жанрово-стилевых разновидностей языка исследователи все больше стали обращать внимание на функциональное разнообразие представленной в корпусе лексики, а развитие технологий позволило расширить инструментарий статистических метрик для выявления релевантных слов. В качестве примера можно привести список [Leech et al., 2001], составленный на основе 100-миллионного корпуса BNC. Список упорядочен по алфавиту с разбиением на лексемы и словоформы каждой лексемы и по частотности (рангам). В нем также представлена статистика использования единиц по подкорпусам (письменному и устному). Как и в работах предшественников, при составлении этого списка используется мера абсолютной частотности, измеряемая в количестве словоупотреблений на миллион слов, с нижней границей отсечения в 10 вхождений на миллион, а также разбиение на ранги. Однако с появлением возможности анализировать и сравнивать отдельные подкорпусы по регистрам, жанрам и стилям возникает необходимость соизмерять объемы подкорпусов и вырабатывать новые статистические метрики их сбалансированности. Для этого вводятся две дополнительные меры: мера характерности и мера дисперсии. Мера характерности представляет собой меру вероятности появления слова в подкорпусе, тогда как мера дисперсии определяет охват (т. е. количество секторов из 100 равных секторов в подкорпусе с данным словом) и распределение (т. е. равномерность распределения данного слова по секторам объемом в миллион слов).

Учет регистровых, жанровых и стилевых различий в сочетании с более детальной проработкой статистических метрик позволяет достичь большей точности в оценке частотности слов в языке, а также дает возможность приспособить список слов под более конкретные практические нужды. «Новая волна» статистических подходов к составлению списков наблюдается и в немецкой лингвистике. В 2005 г. Р. Джонс и Э. Чирнер собрали сбалансированный корпус объемом 4,2 млн словоупотреблений с автоматической лемматизацией и частотной разметкой [Johns, Tschirner, 2006]. На основе этого корпуса они извлекли 4 000 наиболее частотных лемм и упорядочили их по частотности и по алфавиту. В первой части списка все слова были даны вместе, с указанием части речи и с английскими переводами, примерами и важными коллокациями, релевантность которых была рассчитана по минимальной частоте на миллион слов. Во второй части списка те же слова были объединены в отдельные таблицы по тематическому принципу. Авторы списка утверждали, что их корпус покрывает от 80 до 90 % произвольного неспециализированного текста, что было позднее подтверждено посредством квантитативного анализа в работе [Winnerlöv, 2014].

В 2013 г. на основе 4 корпусов текстов: LOB (Lancaster-Oslo-Bergen Corpus), BNC (British National Corpus), BEO6 (The BEO6 Corpus of British English (BEO6), EnTENTen12, создан список New General Service List [Brezina, Gablasova 2015] объемом в 3 000 слов. Необходимость создания нового списка общего пользования взамен по-прежнему широко используемого списка Уэста [West, 1953] обосновывается авторами, во-первых, тем, что часть лексики в его списке уже устарела (к примеру, имеется слово «телеграф», но отсутствует «компьютер»). Кроме того, используемый при его составлении критерий обязательной стилистической немаркированности отбираемой лексики игнорирует некоторые стилистически окрашенные частотные лексические единицы, необходимые даже на базовых уровнях владения языком. Итоговый объем проанализированной авторами лексики составил 12 миллиардов словоформ. В новом получившемся списке оказалось 2 494 леммы, покрывающие 80,1–81,7 % наугад выбранного неспециализированного текста.

В последнее время проблематика ЛМ активно исследуется в рамках японского проекта «Basic German Vocabulary for Foreign Language Learners: A data-driven Approach» по составлению учебных минимумов немецкого языка. В работе [Okamura et al., 2012] на основе нового объемного корпуса Basic German Corpus объемом в 845 млн словоупотреблений с двумя примерно равными подкорпусами немецкоязычных онлайн-форумов и газетных текстов производится анализ лексики с целью выделения лексического ядра с акцентом на большем объеме и большей дифференцированности статистических измерений. Вывод, к которому приходят исследователи на основе сопоставительного анализа двух подкорпусов, состоит в том, что недифференцированный подход к составлению списка лексики приводит к нежелательному смещению важных жанровых и стилевых различий текстов, что говорит о необходимости более подробного статистического анализа.

Появление корпусов большого объема и надежной статистической методологии дало возможность задаться еще одним существенным вопросом о том, насколько статистически выделимо лексическое ядро языка. Попытка поиска ответа на этот вопрос в [Scharloth et al., 2016] на материале того же Basic-German Corpus дала отрицательный результат: частотность слов монотонно убывает, и график не имеет явных изгибов, позволяющих провести деление на ядерную и остальную часть лексики. На этом основании авторы заключают, что граница между ядерной и неядерной лексикой должна проводиться исходя из конкретных практических целей составителей списка.

Таким образом, зарубежный опыт составления ЛМ показывает, что значительным методологическим шагом вперед стало активное использование текстовых корпусов, а с ростом их объема и жанрового разнообразия – учет жанрово-стилевых особенностей лексики. В немалой степени прогресс в этой области также предопределило развитие статистических методов, при помощи которых стало возможно сопоставлять между собой лексемы с точки

зрения не только частотности, но также, среди прочего, стабильности во времени и стабильности по типам текстов.

Основные различия в российских и зарубежных подходах

Как показывает настоящий обзор, преобладающей тенденцией в зарубежных подходах к составлению ЛМ является использование в качестве основы корпуса или корпусов письменных, а в последнее время также и устных текстов. В свою очередь, российские подходы в большей степени консервативны и чаще всего опираются на существующие словари и учебные пособия, хотя иногда также используют частотные списки слов. Во многом это объясняется тем, что российские ЛМ составляются с ориентацией на коммуникативные потребности иностранцев, изучающих русский язык, тогда как большинство составителей зарубежных ЛМ последнего времени ставит более общую задачу выявления лексического ядра языка, что, однако, не исключает появления конкурентных коммуникативно ориентированных и смешанных списков. Тем не менее, и в России, и за рубежом отбор слов в списки с коммуникативной ориентацией, как правило, производится по субъективным критериям, которые в недостаточной степени отражают статистические характеристики реального языкового узуса. Это видно из сравнения процента покрываемости ЛМ произвольно выбранных неспециализированных текстов на языке списка, который оказывается выше у ЛМ, составленных на статистической основе, что доказывает их большую валидность по сравнению с коммуникативно-ориентированными и смешанными ЛМ.

Результаты обзора суммированы в таблице ниже.

Сравнение российского и зарубежных подходов к созданию лексического минимума
Comparing of Russian and foreign lexical minima approaches

Показатель	Лексикография		
	русская	английская	немецкая
Предназначение	Учебные	Теоретические, общего пользования	Теоретические, общего пользования, учебные
Подход к составлению	Преимущественно коммуникативный	Преимущественно статистический	
Источники данных	Учебники, словари частотные списки ранее созданные частотные списки и ЛМ	Корпуса, словари, частотные списки, опрос носителей	
Какой процент текстов должен покрывать ЛМ	70	не менее 80	

Заключение

Приведенный в статье обзор обнаруживает целый ряд моментов, которые необходимо учитывать составителям ЛМ. Прежде всего, составление ЛМ в настоящее время практически непредставимо без корпусной основы. Корпус или корпуса, которые используются в качестве основы, должны быть достаточного объема и в достаточной мере отражать разнообразие языка с точки зрения регистров, жанров и стилей. В свою очередь, сами списки должны демонстрировать высокий процент покрытия исходного корпуса и содержать лексемы, максимально охватывающие все представленные типы текстов. К количественным характеристикам, помимо абсолютной частотности, относятся меры стабильности и продуктивности.

К качественным характеристикам относятся тематическая организация, грамматическая информация и типичные контексты употребления слова. Объем списка, как показывает практика, не играет принципиальной роли и должен определяться исходя из конкретных целей, преследуемых составителями.

Список литературы / References

- Андрюшина Н. П., Козлова Т. В.** Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение. М.; СПб.: ЦМО МГУ, Златоуст, 2006.
Andryushina, N. P., Kozlova, T. V. A minimal Russian wordlist for foreign learners. Elementary level. General knowledge. Moscow, St. Petersburg, Zlatoust, 2006. (in Russ.)
- Владимирова Т. Е., Нахабина М. М., Соболева Н. И., Андрюшина Н. П.** Государственный стандарт по русскому языку как иностранному. Элементарный уровень. 2-е изд., испр. и доп. М.; СПб.: Златоуст, 2001. 28 с.
Vladimirova, T. E., Nakhabina, M. M., Soboleva, N. I., Andryushina, N. P. State Standard for the Russian Language as a Foreign Language. Elementary level. 2nd ed., corr. and add. Moscow, St. Petersburg, Zlatoust, 2001, 28 p. (in Russ.)
- Гугенейм Ж.** Некоторые выводы статистики словаря // Методика преподавания иностранных языков за рубежом. М.: Прогресс, 1967. Вып. 1. С. 302.
Gughenheim, J. Some conclusions of the statistics of the dictionary. In: Methods of teaching foreign languages abroad. Moscow, Progress, 1967, vol. 1, p. 302. (in Russ.)
- Дунаева Л. А., Соррилъя К. М.** Стандартный лексический минимум в электронном практикуме по лексике русского языка // Русский язык за рубежом. 2010. № 6. С. 22–28.
Dunaeva, L. A., Sorrilja, K. M. A standard minimal Russian wordlist for foreign learners in the online platform for learning Russian vocabulary. *Russian Language Abroad*, 2010, no. 6, p. 22–28. (in Russ.)
- Засорина Л. Н.** Частотный словарь русского языка: Около 40 000 слов / Под ред. Л. Н. Засориной. М.: Рус. яз., 1977. 936 с.
Zasorina, L. N. Frequency dictionary of the Russian language: About 40 000 words. Ed. by L. N. Zasorina. Moscow, Rus. lang., 1977, 936 p. (in Russ.)
- Костомаров В. Г.** Принципы отбора лексического минимума // Русский язык в национальной школе. 1963. № 1. С. 31.
Kostomarov, V. G. Principles of creating a minimal wordlist. *Russian Language in the National School*, 1963, no. 1, p. 31. (in Russ.)
- Лённгрен Л.** Частотный словарь современного русского языка. Uppsala, 1993.
Lyonngrén, L. Frequency Dictionary of Modern Russian Language. Uppsala, 1993.
- Маркина Е. И.** Лингводидактические основы разработки лексических минимумов по русскому языку как иностранному (для разных уровней и профилей обучения): Автореф. дис. ... канд. пед. наук. М., 2011. 24 с.
Markina, E. I. Linguodidactic basis for creating minimal wordlists for Russian foreign learners. Abstract. Moscow, 2011. 24 p. (in Russ.)
- Маркина Е. И., Соррилъя К. М.** Основные подходы к минимизации лексики в российской и европейской учебной лексикографии // Вестник РУДН. Серия: Вопросы образования: языки и специальность. 2011. № 3. С. 77–84.
Markina, E. I., Sorrilja, K. M. Main approaches to vocabulary minimization in the Russian and the European lexicography. *Vestnik RUDN. Series: Issues in education: languages and profession*, 2011, no. 3, p. 77–84. (in Russ.)
- Морковкин В. В., Дорогонова И. А.** О новом типе лексического минимума современного русского языка // Русский язык за рубежом. 1976. № 2. С. 59–62.
Morkovkin, V. V., Dorogonova, I. A. On the new type of a minimal modern Russian wordlist. *Russian Language Abroad*, 1976, no. 2, p. 59–62. (in Russ.)
- Трушина Л. Б.** Русский язык как иностранный в специальных целях: прошлое, настоящее, будущее. Опыт исторического обзора / Под ред. В. Г. Костомарова. М., 2013. С. 8–9.
Trushina, L. B. Russian as a foreign language for special purposes: past, present and future. A historical overview. Ed. by V. G. Kostomarov. M., 2013. P. 8–9. (in Russ.)

- Штейнфельдт Э. А.** Частотный словарь современного русского литературного языка. Таллин, 1963.
Steinfeld, E. A. A frequency dictionary of modern Russian literary language. Tallinn, 1963. (in Russ.)
- Bakonyi, H.** Die gebräuchlichsten Wörter der deutschen Sprache für den Fremdsprachenunterricht stufenmäßig zusammengestellt. München, 1934.
- Brezina, V., Gablasova, D.** Is there a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 2015, vol. 36/1, p. 1–22.
- Carroll, J. B., Davies, P., Richman, B.** The American Heritage word frequency book. Boston, Houghton Mifflin, 1971.
- Ek, J. A. van.** Scope. Levels. Strasbourg, Council of Europe, 1987, vol. 2.
- Faucet, L., West, M., Palmer, H., Thorndike, E. L.** The Interim Report on Selection of the English Language as a Foreign Language. London, 1936.
- Gougenheim, G.** Dictionnaire fondamental de la langue française. Paris, Didier, 1958.
- Johns, R., Tschirner, E.** A frequency dictionary of German: Core vocabulary for learners. Routledge, 2006.
- Leech, G., Rayson, P., Wilson, A.** Word Frequencies in Written and Spoken English: Based on the British National Corpus. London, Longman, 2001.
- Lorge, I.** Semantic count of the 570 commonest English Words. New York, Columbia University Press, 1949.
- Morgan, B. Q.** German Frequency Word Book. New York, Macmillan, 1928.
- Ogden, C. K.** Basic English, London. Kegan Paul, Trench Trubner & Co., Ltd. 1930.
- Okamura, S., Lange, W., Scharloth, J.** Methoden der Bestimmung des Kernwortschatzes Deutsch. Grundwortschatz Deutsch: Lexikografische und fremdsprachendidaktische Perspektiven. Tokyo, Studienreihe der Japanischen Gesellschaft für Germanistik, 2012, P. 29–44.
- Palmer, H.** Second interim report on vocabulary selection submitted to the Eighth Annual Conference of English Teachers under the auspices of the Institute for Research in English Teaching. Tokyo, IRET, 1931.
- Pfeffer, J. A.** Grunddeutsch. Basic (Spoken) German Dictionary. Englewood Cliffs, 1970.
- Scharloth, J., Okamura, S., Lange, W.** Gibt es einen Kernwortschatz? Datengeleitete Perspektiven auf die Erstellung von Grundwortschatzen für Deutsch als Fremdsprache. Brunetti, S. u.a. (Hrsg.). Versprachlichung von Welt – Il mondo in parole. Festschrift zum 60. Geburtstag von Maria Lieber. Stauffenburg, 2016.
- Thorndike, D. L., Lorge, I.** The Teacher's Word Book of 30,000 Words. New York, NY, Bureau of Publications, Teachers College, Columbia University, 1944.
- West, M.** A general service list of English words. London, Longman, 1953.
- West, M.** The new method readers (new series). Bombay and Calcutta, Longmans, Green, 1927.
- Winnerlöv, J.** Wie viele Wörter muss man kennen, um einen deutschen Text zu verstehen? Eine quantitative und qualitative Untersuchung zur lexikalischen Textdeckung. Stockholms Universitet, 2012.

*Материал поступил в редколлегию
Date of submission
24.10.2018*

Сведения об авторах / Information about the Authors

Муравьев Никита Алексеевич, преподаватель Школы лингвистики Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

Nikita A. Muravyev, Lecturer of the Faculty of Humanities, School of Linguistics, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)

nmuravev@hse.ru

ORCID 0000-0001-9586-5928

SPIN 8474-3643

Ольшевская Мария Юрьевна, старший преподаватель Школы лингвистики Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

Maria Yu. Olshevskaya, Senior Lecturer of the Faculty of Humanities, School of Linguistics, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)

molshevskaya@hse.ru

ORCID 0000-0002-1050-0784

SPIN 2146-2398

Исследование потенциала компьютерных программ для стилистического и переводческого анализа текста и его практическое применение

В. Г. Сибирцева, Н. Х. Фролова

*Национальный исследовательский университет «Высшая школа экономики»
Нижний Новгород, Россия*

Аннотация

Предлагается критический обзор существующих компьютерных ресурсов с точки зрения их использования в филологическом анализе. В гуманитарных науках, таких как социология, психология и менеджмент, существуют различные типы программного обеспечения для многомерного анализа текста (проверка авторства, определение совместимости слов, выявление эмоциональной окраски текста и т. д.), которые могут применяться и в филологических исследованиях для интерпретации текста. Однако для филологии они остаются *terra incognita*: традиционно изучение и анализ текста в литературоведении связаны с его индивидуальной интерпретацией исследователем, поэтому до сих пор преобладает консервативный подход, не подкрепленный данными контент-анализа или статистики.

Различные программы были успешно протестированы на материале русской и иностранной художественной литературы, а в дальнейшем интегрированы в обучение студентов бакалавриата и магистратуры факультета гуманитарных наук (Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород). Накопленный пятилетний опыт позволил получить неединичные достоверные данные и выполнить сравнительный анализ компьютерных программ, таких как AntConc concordancer, многофункциональный контент-анализатор LEKTA, LF Alighner, TextAnalyst SDK. На фактическом материале доказывалась необходимость перехода от традиционного сбора художественного материала и анализа различных дискурсов к широкому применению веб-инструментов. Практическое применение компьютерных программ заполняет разрыв между современными требованиями цифровых гуманитарных наук (Digital Humanities) и консервативным подходом многих исследователей.

Авторами статьи приводятся краткие аннотации курсовых и дипломных работ, посвященных различным видам анализа текста с опорой на программные инструменты. Отдельные аспекты, посвященные перспективным, но малоиспользуемым компьютерным ресурсам, описываются подробно. При этом подчеркивается необходимость избирательного подхода к общедоступным лингвистическим корпусам, узкоспециализированным программам-конкордансам и частотным анализаторам. Применение компьютерных инструментов в анализе текста обязательно требует дальнейшей традиционной интерпретации материала. Достижение необходимого баланса в использовании программных ресурсов может означать оптимизацию анализа художественного оригинального или переводного текста наряду с определением новых критериев оценки репрезентативности и валидности результатов подобных литературоведческих и лингвистических исследований.

Ключевые слова

многомерный анализ текста, компаративный анализ, компьютерные инструменты в филологии, перевод, Digital Humanities

Для цитирования

Сибирцева В. Г., Фролова Н. Х. Исследование потенциала компьютерных программ для стилистического и переводческого анализа текста и его практическое применение // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 90–101. DOI 10.25205/1818-7935-2019-17-1-90-101

Testing Computer Programs' Potentialities in Stylistic and Translational Comparative Text Analysis

Vera G. Sibirtseva, Natalia H. Frolova

*National Research University Higher School of Economics
Nizhny Novgorod, Russian Federation*

Abstract

Traditionally, text and literature abstracts analysis is associated with the individual perception and creative interpretation of the researcher. Though many software applications are being created for multidimensional text analysis from the point of view of verification of authorship, identifying the compatibility of words and emotional coloring of the text, etc., the conservative approach prevails. No matter how many different types of software for the Humanities (such as sociology, psychology and management) there exist, they are still rarely used in philological research while interpreting a text.

The current article provides a critical review of technology enhanced philological analysis. Various computer programs were successfully tested on the material of Russian and foreign literature, and later on used to train future specialists from the Humanities department (National Research University "Higher School of Economics", Nizhny Novgorod). The accumulated experience allowed us to obtain a reliable data and perform a comparative analysis of computer programs, such as AntConc concordancer, LEKTA, LF Aligner, TextAnalyst SDK, a multifunctional content analyzer.

The article describes a survey conducted at NRU HSE to address the need to move from traditional collection of authentic materials and analysis of various discourses to the widespread use of web-based tools designed to accelerate this process and follow the contemporary requirements of the Digital Humanities. Brief prefaces of the term papers and the theses devoted to various methods of text analysis based on software tools are considered and described in detail. Specific examples emphasize the need for a selective approach to publicly available linguistic corpora and highly specialized concordance programs, as well as to tools for aligning translated texts and frequency analyzers. However, the use of computer tools in text analysis should be combined and completed with traditional methods of linguistic and stylistic interpretation. The purpose is to achieve the necessary balance in the use of software resources which can mean optimizing the comparative analysis of original or translated texts, along with the definition of new criteria for assessing the representativeness and validity of the results of literary or linguistic studies.

Though the proposed gap filling between the traditional and innovative techniques of text interpretation is rather novice, computer assisted text analysis is promising and tends to be in the mainstream of comparative linguistic and stylistic research.

Keywords

multimodal text analysis, computer tools in philological education, comparative analysis, text translation, Digital Humanities

For citation

Sibirtseva, Vera G., Frolova, Natalia H. Testing Computer Programs' Potentialities in Stylistic and Translational Comparative Text Analysis. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 90–101. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-90-101

Введение

Традиционно изучение и анализ художественных произведений связаны с индивидуальным восприятием и творческой интерпретацией исследователя. В настоящее время создается много программных приложений для многомерного анализа текста. Предлагается проверка авторства, выявление сочетаемости слов и эмоциональной окрашенности текста и т. д. Тем не менее, консервативный подход в исследованиях по литературе сохраняется до настоящего времени. Для гуманитарных наук (таких как социология, психология и менеджмент) существуют различные виды программного обеспечения, однако они достаточно редко применяются в филологических исследованиях при интерпретации текста.

Идея поверить «алгеброй гармонию» не должна восприниматься буквально, компьютер не может заменить профессионального филолога. Но в процессе текстового анализа всегда появляются задачи, связанные с рассмотрением большого объема материала или с рутинным поиском однородной информации. И тогда компьютерные инструменты могут облегчить обработку материала и сэкономить время. При таком подходе к области лингвистического

и филологического (в частности, компаративного) анализа открываются новые возможности и перспективы.

Основные причины пренебрежения существующими компьютерными программами в области филологических исследований следующие: отсутствует удобный пользовательский интерфейс или пошаговые инструкции для лингвистов и филологов; средства автоматической обработки текста не структурированы с точки зрения филолога; число примеров применения программного обеспечения при анализе литературных отрывков очень малочисленно, и поэтому алгоритмы, учитывающие частотность и валидность автоматической обработки, отсутствуют.

В настоящей статье описан пятилетний опыт апробации веб-инструментов в качестве средства стилистического и лингвистического анализа целого ряда коллекций и банков различных по жанру и регистру текстов силами студентов факультета гуманитарных наук (Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород). Исследовательские проекты представляют собой курсовые и выпускные квалификационные работы, посвященные анализу содержания текстового материала с применением компьютерных программ, анализаторов и веб-инструментов. Все они были загружены в систему LMS вуза, и, следовательно, большинство из них находится в режиме ограниченного доступа. Наиболее интересные исследовательские работы студентов бакалавриата и магистратуры представлены в табл. 1.

Программные ресурсы, использованные для анализа и интерпретации художественных текстов, представляют собой две большие группы: программное обеспечение, встроенное в лингвистические исследовательские корпуса, и дополнительные компьютерные программы. Состоявшиеся защиты курсовых и выпускных квалификационных работ свидетельствуют о том, что данные компьютерные ресурсы успешно применялись студентами для определения стилистической окраски текста или его перевода, для выявления ключевых мотивов произведения, а также в исследовании компьютерных средств для изучения языка на аутентичных текстах. Таким образом, описанные в настоящей статье способы компьютерного текстового анализа могут использоваться как профессиональными лингвистами, так и новичками.

Цифровые коллекции текстов и лингвистические корпуса

Развитие цифровых технологий и интерес исследователей к стилистическим вариантам русского языка стал причиной появления цифровых лингвистически аннотированных корпусов. Работа с корпусами является на сегодняшний день одним из ведущих методов лингвистического исследования, при котором можно решить множество различных проблем. Крупным проектом, базирующимся на Национальном корпусе русского языка (НКРЯ)¹, является постоянно пополняемый раздел учебного портала НКРЯ «Проблемы русской стилистики» под руководством А. И. Левинзон [2007] и Ю. М. Кувшинской [2008]. Авторы проекта исследуют закономерности выбора лексических единиц и соответствующую частоту их использования в речи XX–XXI вв. Полное лингвистическое аннотирование текстов и удобные опции поиска облегчают сбор исследовательских данных.

В преподавании иностранных языков и литературы также могут использоваться национальные языковые корпуса как средства многомерного анализа стиля и языка автора. Однако частота использования лексических единиц не является релевантным показателем при исследовании стиля, поскольку характерный для лингвистических корпусов широкий охват всевозможных текстов (в том числе современных нехудожественных) препятствует глубокому изучению текстов одного автора.

¹ URL: <https://ruscorpora.ru/>.

Исследовательские работы студентов

Students research works

ФИО студента	Тема / Topic	Аннотация и ссылка
Одинцова Анна Николаевна	Контент-анализ переводов цикла сказок Д. Р. Кипплинга «Just So Stories» в программе ЛЕКТА / Content analysis of the translations of the number of fairy tales by D. R. Kipling “Just So Stories” via the program LEKTA	Описывается оригинальный метод анализа переводов сказок Д. Р. Кипплинга с применением программы Lekta (лексико-семантический текстовый анализатор). Результатом обработки данных является выделение факторов, определяющих неочевидные взаимосвязи в текстах. Метод позволяет оценить адекватность передачи авторских смыслов и может быть использован в практических нуждах переводоведения. https://www.hse.ru/edu/vkr/126185253
Самсонова Галина Игоревна	Компаративный анализ полисемии цветообозначений (на материале коллекций примеров русского, английского и французского языков) / Comparative analysis of the color terms polysemy (based on collections of Russian, English and French examples)	Рассматривается проблематика многозначности колоративной лексики, проводится компаративный анализ цветообозначений русского, английского и французского языков для установления особенностей, сходств и различий значений. Проводится работа с данными из корпусов текстов. Составляются триграммы, то есть наиболее вероятно встречаемые слова в окружении того или иного цветообозначения. https://www.hse.ru/edu/vkr/125370574
Внюленова Любовь Андреевна	Компьютерные средства анализа характеристик героев в произведениях В. П. Крапивина / Computer tools for analyzing the characteristics of characters from the works by V. P. Kravivin	Проводится анализ характеристик главных героев произведений В. П. Крапивина с использованием лексико-семантического анализатора ЛЕКТА. Выявленные ключевые характеристики сопоставляются с данными, полученными исследователями традиционными способами. https://www.hse.ru/edu/vkr/86916268
Молокова Светлана Вячеславовна	Составление корпуса русско-английских эквивалентов интернет-коммуникации и методика работы с ним / Building a corpus of Russian-English equivalents of Internet communication and methods of working with it	Описывается методика построения корпуса русско-английских эквивалентов интернет-коммуникации, проводится исследование некоторых понятий с применением инструментов ресурса Sketch Engine: Concordance, WordSketches, Thesaurus и Sketch-Diff. Производится оценка результатов работы корпуса, предлагаются возможные пути его дальнейшего усовершенствования и способы применения в прикладных целях. https://www.hse.ru/edu/vkr/153009761
Аджигитова Юлия Руслановна	Создание приложения «Словарь ключевых слов английских романов-антиутопий» (на материале романов Дж. Оруэлла «1984», О. Хаксли «Brave New World») / Creation of the application “Dictionary of keywords of English dystopian novels” (based on the novels by G. Orwell “1984”, O. Huxley “Brave New World”)	Выявлены основные различия и сходства рассматриваемых антиутопий. Обнаружено, что с помощью ключевых слов можно дополнить наблюдения исследователей. По итогу работы было создано веб-приложение «Словарь ключевых слов британских антиутопий», ориентированное на пользователей, знающих английский язык и интересующихся антиутопиями, филологов и лингвистов. https://www.hse.ru/edu/vkr/182595148

ФИО студента	Тема / Topic	Аннотация и ссылка
Овчинникова Наталья Александровна	Особенности контекстного анализа посредством веб-технологий / Peculiarity of Context Analysis by Means of Web-tools	Выявляются и анализируются наиболее частотные программные продукты для контекстного анализа. В связи с тем, что возможности человека-переводчика ограничены, прежде всего, по времени, предлагается возможное решение данной проблемы путем создания и развития веб-инструментов для машинного перевода. https://lms.hse.ru/?ap_work&h_id=64CCF754-7750-4847-BC97-A4942AA02B69
Овчинникова Наталья Александровна	Роль коннотации лексических единиц в контекстном анализе при работе с корпусами / Corpus Use in Context Analysis of the Connotation of the Word	Проводится уточнение значения коннотации, а также связь изучения данного явления с компьютерными технологиями. Целесообразность применения веб-инструментов показана на примере использования корпусов текстов. Целью исследования является определение роли коннотации выявления ее функций в контекстном анализе лексических единиц. https://lms.hse.ru/?ap_work&h_id=24AF64B6-1AB2-48BA-92AA-462084F80EFB
Ураев Дмитрий Юрьевич	Сравнительный анализ эффективности применения веб 2.0 инструментов в процессе изучения английского языка / Web 2.0 based tools in English learning: comparative approach	Обучение иностранным языкам все теснее связано с использованием аутентичных современных текстов и во многом обусловлено развитием компьютерных технологий и непрерывным увеличением их быстроедействия. Рассматриваются и сопоставляются различные инструменты веб 2.0. https://lms.hse.ru/?ap_work&h_id=12F971D0-FDEF-44DA-AC39-4EE6565B34D4
Владимирский Арсений Ильич	Мотивы жизни, смерти и бессмертия в трагедии У. Шекспира «Гамлет» и их реализация в русских переводах / Motives of life, death and immortality in the tragedy by William Shakespeare's «Hamlet» and their implementation in Russian translations)	Проводится частотно-семантическое исследование, выделяется семантическое ядро и центр слов форм «жизнь», «смерть» и «бессмертие» и впоследствии создаются лексико-семантические микрополя мотивов; анализируются переводы «Гамлета» Уильяма Шекспира на русский язык с использованием конкорданса AntConc; выявляются сходства и различия на уровне мотивов между оригиналом и переводами. https://www.hse.ru/edu/vkr/220137419
Кислянищина Елена Александровна, Сажина Алёна Сергеевна	Изучение лексики английского языка с применением информационных технологий / Vocabulary learning by means of ICT	Информационно-коммуникационные технологии призваны ускорить процесс ввода в активный оборот и запоминания новых лексических единиц из актуальных текстов, а также пополнить список современных синонимов и фразеологизмов. В печати
Манаенкова Валерия Сергеевна, Абрамова Мария Александровна	Морфологический и синтаксический состав свободных ритмов XVIII века и свободного стиха 2-й половины XX века. Традиции и новаторство / Morphology and Syntax of Free Verse in German Poetry of the 18 th Century and of the Second Half of the 20th Century: Tradition and Innovation	Проводится анализ изменения синтаксиса и морфологического состава предложений в нерифмованном немецком стихе. Ведется работа над созданием представительной коллекции текстов, выявлением характеристик стихов при помощи программных средств (конкордансер, выравниватель), проводится сопоставление стихов-оригиналов и их переводов на русский язык. В печати

Для изучения наследия отдельного автора необходимы специальные лингвистически аннотированные цифровые собрания текстов. Образцовым современным проектом, учитывающим достижения Digital Humanities и не утратившим удобство в использовании, является проект «Толстой-диджитал» [Bonch-Osmolovskaya, Kolbasov, 2015]. Сохраняя все достоинства национальных корпусов (объем материала, полное морфологическое и синтаксическое описание лексики, метаданные, интеллектуальный поиск), он обогатился мультимедийной поддержкой (аудио, фото, карты и т. д.) и доступом к образовательным мероприятиям (музеи и новости в режиме реального времени). Корпус в настоящий момент состоит из многочисленных томов известного русского писателя Льва Толстого, находящихся в процессе оцифровки и обработки [Tolstaya, 2016], ставя конечной целью создание онлайн-энциклопедии, которая позволит пользователям получать доступ ко всем текстам Толстого, по ключевым фразам находить конкретные цитаты, тексты или отрывки эпистолярного наследия писателя. Проект поддерживается ведущими мировыми университетами, включая Московский государственный университет и Гарвардский университет.

Не менее известен частотный словарь-конкорданс публицистики другого русского писателя Ф. М. Достоевского [Kopotev, 2003], созданный несколько ранее. Данный корпус текстов можно назвать удачной коллекцией готовых произведений, опирающейся на тридцатитомное академическое собрание сочинений Ф. М. Достоевского. Словарь позволяет получать контекстные и частотные характеристики любого слова или словоформы, а также находить их позицию в тексте в собрании сочинений, т. е. проводить сквозной поиск по всем текстам Достоевского. Оцифрованные и аннотированные материалы словаря-конкорданса вместе со словарем литературных текстов писателя, созданным Институтом русского языка РАН, дают полное представление о частотных характеристиках основного корпуса текстов Достоевского.

Наряду с национальными корпусами и корпусами одного автора существуют коллекции, содержащие разные варианты перевода одного текста, систематизированные и аннотированные в соответствии с конкретными целями. Так, русско-французский параллельный поэтический корпус первой трети XIX в. помогает исследовать специфику переводов нескольких авторов и анализировать оригинальные и переводные идиомы.

Параллельный сборник переводов «Слова о полку Игореве» – еще один ресурс для сравнения переводов. Возможности этого корпуса одного конкретного текста довольно широки, но прежде всего пользователь может анализировать языковые конструкции в параллельных версиях и тем самым получать полную картину восприятия «Слова о полку Игореве» в современном мире [Orekhov, Rybina, 2017].

Поскольку рассмотренные ресурсы посвящены художественным текстам, было бы логично предположить, что исследователи обращаются к данным корпусам и коллекциям постоянно. В частности, открывается перспектива изучения индивидуального стиля автора: исследование рифмы, стилистических клише и т. д.

Вышеупомянутые программы имеют удобный интерфейс, что очень важно для исследователя-филолога, однако их использование в лингвистических и литературоведческих исследованиях значительно ограничено: все они являются негибкими литературными сегментами. С помощью корпуса Ф. М. Достоевского невозможно интерпретировать произведения любого другого автора, например, Н. В. Гоголя или М. А. Булгакова. Но перспектива и актуальность таких коллекций очевидна – филологи получают хороший образец оцифрованных текстов для изучения художественных произведений с полным набором инструментальных средств: системой поиска и цитирования, лингвистической разметкой, метаданными и тезаурусом.

Программные инструменты филологического анализа

Инструмент обработки естественного языка – это специальный продукт, который может применяться при анализе художественных текстов. В некоторых исследовательских публи-

кациях содержится сводное описание инструментов, позволяющих им выполнять лингвистическую обработку текста. Однако филологи в основном избегают использования компьютерных программ в своих исследованиях [Cordell, 2015], игнорируя цифровизацию гуманитарных дисциплин (Digital Humanities). Следует отметить, что в контексте настоящей статьи рассматриваются только программы, помогающие литературоведческому анализу и интерпретации текста. За рамками статьи остаются такие программные продукты, как инструменты для фонетического и фонологического анализа текста, автоматизированные создатели словарей, инструменты автоматической индексации и извлечения терминологии.

Все существующие инструменты и программы можно разделить на несколько групп:

- 1) программы для анализа морфологических и синтаксических особенностей текстов;
- 2) программы статистического анализа текста.

Эти инструменты могут быть интересны филологам, поскольку каждый из них имеет свои преимущества. В 2001 г. появился TextAnalyst², интерфейс которого далек от современных разработок, однако он по-прежнему используется в образовательных целях, так как содержит полезные компоненты реализации комплекса функций автоматического анализа для русских и английских текстов. Программа позволяет выделить основное содержание текста, сформировать гипертекстовые базы данных, выполнить поиск информации в текстах и т. д. Интересными представляются следующие описанные функции ресурса:

- контент-анализ с гиперссылками – идентификация семантической структуры текста в виде иерархии тем и подтем;
- семантический поиск связей между запрошенными словами (с учетом скрытых семантических связей между запрошенными словами и контекстом);
- автоматически создаваемая «выжимка» текста, представляющая собой семантический портрет текста с набором ключевых слов и наиболее содержательных фраз.

Работа над проектами, посвященными филологическому анализу с использованием компьютерных инструментов, будет более эффективной при создании междисциплинарных проектных групп.

Опыт совместных исследовательских проектов (Digital Humanities projects) показывает, что применение программ для анализа литературных текстов затруднено следующими факторами: отсутствие информации о работоспособности того или иного ресурса на сегодняшний день; дороговизна и отсутствие открытого доступа к продукту; отсутствие четкой понятной инструкции для непрофессионального пользователя. Тем не менее, практическое использование некоторых веб-инструментов в настоящем исследовании продемонстрировало различные возможности для анализа, а затем и интерпретации литературных текстов с помощью современных программ.

По инициативе студентов и преподавателей ВШЭ были экспериментально проверены следующие программы: конкордансер AntConc³, многофункциональный текстовый анализатор LEKTA⁴, программа LF Aligner⁵ для выравнивания текста. Например, в процессе изучения разных переводов одного и того же текста можно отслеживать стилистические изменения в дискурсе и обнаруживать особенности стратегии перевода. Поскольку художественные тексты, как правило, содержат большое количество лакун и случаев асимметричного перевода, анализ и интерпретация произведений при помощи программных средств становится увлекательной исследовательской задачей. Несмотря на широкие возможности, указанные программы достаточно редко используются литературоведами.

Бесплатный ресурс LF Aligner можно применять для изучения несоответствий в различных переводах одного текста. Наглядным примером использования программных продуктов

² URL: <https://analyst.ru/index.php?lang=eng&dir=content/products/>.

³ URL: <https://www.laurenceanthony.net/software.html/>.

⁴ URL: <http://content-analysis.ru/>.

⁵ URL: <https://sourceforge.net/projects/aligner/>.

в литературоведении на начальных этапах может являться проект по анализу нескольких переводов на русский язык романа И. Хмелевской «Все красное» [Сибирцева, 2013], осуществленный в рамках совместной работы студентов и преподавателей. Для анализа данного произведения уже существующий лингвистический польско-русский параллельный корпус не подходит: как и рассмотренные выше национальные корпуса и коллекции текстов, он не позволяет добавлять тексты из области интересов конкретных исследователей. По сравнению с лингвистическими корпусами, программа LF Aligner для выравнивания параллельных текстов может стать дополнительным инструментом при анализе перевода по нескольким причинам. Прежде всего, визуализация нескольких вариантов перевода удобна для анализа. Во-вторых, объем программы не ограничивается двумя текстами; при необходимости могут быть добавлены четыре или более текстов на разных языках. И, наконец, удобный интерфейс программы для филолога, а также простые команды не требуют специальных навыков программирования, и весь процесс обработки текстов занимает всего несколько минут.

Разумеется, возможности программы LF Aligner намного шире, чем визуализация сравниваемых фрагментов, но наглядность в процессе сравнения может значительно облегчить восприятие нескольких текстов и сэкономить время на поиск необходимых фрагментов в каждом тексте, взятом по отдельности.

Программа для контент-анализа ЛЕКТА, разработанная в Нижегородском государственном университете и имеющая ограниченные бесплатные возможности⁶, также использовалась в анализе литературных текстов с применением компьютерных средств. Поскольку компьютерные программы дополняют друг друга и фактически усиливают любые исследования, для получения достоверных данных необходимо использовать как качественные, так и количественные методы. В лингвистике и литературоведческих исследованиях допускается применение контент-анализа. Практичность этого подхода заключается в том, что контент-анализ собирает чисто лингвистическую информацию о текстовых характеристиках и на ее основе пытается идентифицировать семантические особенности текста. Суть контент-анализа с применением компьютера – породить достоверную гипотезу о содержании текста на основе внешних (количественных) характеристик текста на уровне слов и словосочетаний и, как следствие, сделать выводы об особенностях авторского слова: намерении, реляции, системе ценностей и т. д. Использование контент-анализа в интерпретации текста накладывает определенные ограничения, о которых переводчику необходимо помнить: текстовый анализ в значительной степени формализуется, производится простая количественная регистрация языковых единиц, что позволяет прогнозировать содержание текста по семантике включенных в него единиц. Практика доказывает эффективность контент-анализа в документальной прозе, фольклоре, а также в авторских сказках. Однако тексты с большим количеством метафор, языковой игрой и сложным сюжетом не могут быть достоверно интерпретированы при помощи контент-анализа.

Контент-анализ представляет собой метод обнаружения и оценки специфических характеристик текстов и других носителей информации (например, видеороликов, телевизионных и радиопрограмм, интервью, ответов на открытые вопросы и т. д.). Представленная возможность компьютерной «кросс-обработки» текстов с дальнейшей интерпретацией результатов широко используется в социологии, позволяет идентифицировать скрытое информационное содержание и его форму. Не следует забывать, что такой анализ значительно затруднен для неподготовленного исследователя-филолога. Студент или преподаватель должен научиться работать в программе ЛЕКТА самостоятельно или с помощью специалиста.

Исследования, упомянутые в данной статье, были сделаны студентами-магистрантами после прослушивания лекций по работе с программой контент-анализа. Для изучения литературного текста с помощью программы ЛЕКТА были выбраны многочисленные переводы

⁶ См.: Рюмин А. Блог о контент-анализе. Софт. URL: <http://content-analysis.ru/index.php/luchshij-soft-dlya-kontent-analiza/>.

сказки Д. Р. Кипплинга «Как леопард стал пятнистым» [Kipling, 2013]. Авторские сказки Д. Р. Кипплинга отличаются простотой стиля и языка, план выражения непосредственно соответствует плану содержания, а лингвистические характеристики текста показывают его сходство с интервью.

Программа LEKTA способствовала полуавтоматизированному поиску ключевых слов и фраз, которые образуют точки пересечения текстовых значений. Факторный анализ (сравнение совпадающих слов) помог сделать более обоснованные выводы о семантическом аспекте текста, как поверхностном, так и скрытом. В ходе исследования переводов одной главы из сказки «Как леопард стал пятнистым» выяснилось, что многие ключевые слова связаны с цветом: например, в описании особенностей сказочной топографии (High Veld) и шкуры леопарда преобладает желтый и коричневый цвет.

В ходе исследования были не только проанализированы преимущества и недостатки одиннадцати известных переводов сказки, но и произведена компиляция гипотетически наиболее полного перевода, включающего в себя различные фрагменты авторских вариантов. Контент-анализ позволил выявить скрытый смысл текста и его переводов, а также провести полный семантический анализ сказки. Таким образом, программа для количественного социологического анализа доказала свою состоятельность в компаративных литературоведческих исследованиях. Подобный анализ представляется целесообразным для циклов коротких текстов, многозначность семантики которым не свойственна. В этом случае компьютерный поиск семантических связей может принести очень интересные результаты.

Конкордансер AntConc прост в эксплуатации и требует минимальной подготовки для текстового анализа, в связи с чем широко используется переводчиками. При этом программа вполне применима и для филологических исследований. AntConc – бесплатная программа, которая помогает применять самые простые приемы лексического анализа пользовательских текстовых коллекций, например выявлять типичную сочетаемость слов и наиболее часто встречающиеся словоформы. Неоспоримым преимуществом этой программы является способность обрабатывать общемировые языки. AntConc не может объединить различные морфологические словоформы в одну лексему (как в LEKTA), но этой проблемы можно избежать, если знать правила комбинации различных лексем. Набор инструментов AntConc довольно разнообразен, но только некоторые из них можно применять в процессе работы с литературным текстом: программа позволяет выводить согласование ключевых слов в контексте (KWIC); выбирать ключевые слова в тексте; получать частотные, алфавитно-частотные и инвертированные текстовые словари; составлять списки совпадений с искомым словом.

Данная программа применялась в интерпретации повестей В. Крапивина [2001]. В частности, для анализа стиля В. Крапивина использовались четыре текста: «Журавленок и молнии», «Лето кончится не скоро», «Бабушкин внук и его братья» и «Самолет по имени Сережка». Объектом исследования был наиболее часто встречающийся авторский типаж, так называемый «Крапивинский мальчик» [Разумихин, 1982]. Молодые исследователи попытались выявить особенности языкового поведения и характера героев Крапивина на основе коллокаций слов. Прежде всего, были проанализированы слова «мальчик», «мальчишка» и имена главных героев. С помощью частотного анализа и списков коллокаций было установлено, что в текстах этого автора часто встречаются словосочетания «весело сказал мальчик» и «мальчик улыбнулся». Книги Крапивина в основном ориентированы на подростков, и неудивительно, что имя главного героя часто сочетается с такими словами, как «обрадовался», «усмехнулся», «засмеялся», «весело», «улыбнулся»; независимо от того, какие проблемы и неприятные ситуации случаются с главным героем, автор идеализирует мир, изображая мальчика как радостного и счастливого. Тексты также содержат много описаний внешнего вида персонажа: «кудрявый мальчик», «симпатичный мальчик, рыжий», «рыжеватый мальчик», «длинноволосый мальчишка», «белобрысый мальчишка», «веснушчатый мальчишка», «мальчишка в синей рубашке», «мальчишка в широкой кепке», «желтоглазый мальчишка».

В результате анализа художественного текста, выполненного программой AntConc, была установлена корреляция между основными чертами «Крапивинского мальчика», намеченными в качестве рабочей гипотезы, и их количественными показателями. Наблюдения исследователей произведений В. Крапивина, работающих в традиционном ключе, были подтверждены статистически.

Выводы

В результате экспериментального использования программных инструментов была доказана их эффективность с точки зрения выбора стратегии перевода, анализа коллокаций или контекстуального анализа. Идея «Литературный текст vs Компьютерный анализ» представляется несостоятельной, потому что это заблуждение основано на двух изначально ложных тезисах: «компьютер способен оценить каждый оттенок смысла» и «литературные исследования не требуют компьютерного анализа вообще».

В настоящем обзоре намеренно были описаны бесплатные и частично бесплатные программы. Необходимость привлечения финансирования всегда препятствует экспериментальной исследовательской работе и является негативным фактором для исследователей (особенно студентов), подталкивающим к отказу от использования веб-инструментов. Как сказано выше, обсуждаемые программы не создавались специально для литературоведческого анализа и не стремятся полностью заменить научного исследователя. Специалисты по переводу, профессионалы в области лингвистики и литературоведы должны учитывать новые возможности цифровизации филологии. Вместе с тем данные компьютерные инструменты необходимо тестировать на новом материале, чтобы выявить возможные ограничения в их использовании. Каждая программа требует определенных навыков при работе с ней. Выводы можно найти в табл. 2.

Таблица 2

Характеристики исследуемых программ в сопоставлении

Table 2

Comparative characteristics of the programs under study

Программа Program	Большой объем текстов Great text volume	Бесплатность Free access	Удобство использования User-friendliness	Многоязычность Multilingual	Мультифункциональ- ность Multifunctioning
TextAnalyst SDK	+	+	–	–	–
LF Aligner	+	+	+	+	–
LEKTA	+	–	+	–	+
AntConc	+	+	–	+	+

В ходе экспериментального использования программ на основе тестирования были определены наиболее ценные для большинства опрошенных студентов и преподавателей характеристики представленных в табл. 2 инструментов. Выделенные преимущества:

- 1) способность обрабатывать большие массивы данных;
- 2) бесплатный доступ как для отдельных лиц, так и для организаций;
- 3) понятный интерфейс, не требующий специальных навыков;

- 4) сокращение времени работы;
- 5) возможность загрузить готовый текст без необходимости вводить его вручную.

Общим преимуществом всех подобных программ является способность обрабатывать большие объемы текстов за значительно меньшее время, чем при традиционном обращении к бумажным носителям текста.

Дигитализация всех сфер повседневной и научной жизни требует от филологов освоения новых направлений (Digital Humanities) и умения применять на практике инновационные способы интерпретации текста и анализа его содержания. Постепенно эти навыки формируются и с успехом используются, несмотря на нехватку IT-специалистов для оказания технической и просветительской поддержки в университетах. Более того, само создание языкового корпуса или учебной коллекции текстов занимает много времени, требует специализированной подготовки и скоординированной работы филологов и программистов. Но как только это происходит, у исследователей появляется много новых возможностей интерпретации слова или стиля определенного автора или его переводчика. Поэтому настоятельно рекомендуется чаще обращаться к программам, упомянутым в данной статье.

Необходимо, однако, уметь выбирать компьютерные программы и целесообразный способ контент-анализа, четко соотносить требования исследования и желание исследователя. Представленные в статье исследовательские работы являются примером данного осознанного выбора.

Область исследований на стыке компьютерных и гуманитарных наук неуклонно расширяется. Но компьютерный анализ не может заменить лингвистическую интерпретацию, проведенную филологом лично. Результаты, полученные в процессе компьютерного анализа, доказывают свою достоверность и эффективность и являются многообещающими в области литературоведческого анализа, а потому заслуживают дальнейшего внимания научной общности.

Список литературы / References

- Крапивин В. П.** Бабушкин внук и его братья // Крапивин В. П. Собр. соч.: В 30 т. М.: Центрполиграф, 2001. Т. 18. 437 с.
- Krapivin, V. P.** Babushkin vnuk i ego brat'ya. In: Krapivin, V. P. Sobranie sochinenij: In 30 vols. Moscow, Tsentrpoligraf, 2001, vol. 18, 437 p. (in Russ.)
- Кувшинская Ю. М.** Проблемы русской стилистики. 2008. URL: <https://studiorum-ruscorpora.ru/stylistics/>.
- Kuvshinskaya, Yu. M.** Problemy russkoj stilistiki [Russian Stylistics Problems]. 2008. (in Russ.) URL: <https://studiorum-ruscorpora.ru/stylistics/>.
- Левинзон А. И.** Использование НКРЯ в преподавании дисциплины «Риторика» в вузе // Национальный корпус русского языка и проблемы гуманитарного образования. М.: AST, 2007. С. 160–173.
- Levinzon, A. I.** RNC use in discipline Rithorics teaching in higher institutions. In: RNC and problems of humanitarian education. Moscow, AST Publ., 2007, p. 160–173. (in Russ.)
- Разумихин А.** Правило без исключений, или Прозрачная злость и интеллигентные мальчики Владислава Крапивина // Урал. 1982. № 8. С. 149–152.
- Razumikhin, A.** The unexceptional rule or crystal anger and decent boys by Vladislav Krapivin. *Ural*, 1982, no. 8, p. 149–152. (in Russ.)
- Сибирцева В. Г.** Технология использования параллельного подкорпуса Национального корпуса русского языка и коллекций текстов в обучении иностранным языкам // Rocznik Instytutu Polsko-Rosyjskiego. 2013. № 2 (5). URL: <https://ip-r.org/wp-content/uploads/2014/01/Sibircewa-Rocznik-IPR-2-5-2013.pdf/>.
- Sibirtseva, V. G.** Technologies of using parallel subcorpora of RNC and text collections in teaching of foreign languages. *Rocznik Instytutu Polsko-Rosyjskiego*, 2013, no. 2 (5). (in Russ.) URL: <https://ip-r.org/wp-content/uploads/2014/01/Sibircewa-Rocznik-IPR-2-5-2013.pdf/>.

- Bonch-Osmolovskaya, A., Kolbasov, M.** Tolstoy digital: Mining biographical data in literary heritage editions. In: CEUR Workshop Proceedings. 2015. URL: <https://ceur-ws.org/Vol-1399/>.
- Cordell, R.** How Not to Teach Digital Humanities. 2015. URL: <http://ryancordell.org/teaching/how-not-to-teach-digital-humanities/>.
- Kipling, J. R.** How the Leopard Got His Spots. 2013. URL: <https://etc.usf.edu/lit2go/79/just-so-stories/1304/how-the-leopard-got-his-spots/>.
- Orekhov, B., Rybina, M.** Le bestiaire médiéval dans un contexte surréaliste : à propos d'une traduction de Slovo o polku Igoreve par Philippe Soupault. In : Verba sonandi : Représentation linguistique des cris d'animaux. Presses Universitaires de Provence, 2017, p. 321–331.
- Tolstaya, F.** Tolstoy goes digital. *East-West digital news*, 2016. URL: <https://www.ewdn.com/2016/09/27/tolstoy-goes-digital/>.

Материал поступил в редколлегию
Date of submission
01.10.2018

Сведения об авторах / Information about the Authors

Сибирцева Вера Григорьевна, кандидат филологических наук, доцент департамента социальных наук Национального исследовательского университета «Высшая школа экономики» (ул. Б. Печерская, 25, Нижний Новгород, 603155, Россия)

Vera G. Sibirtseva, Associate Professor: HSE Campus in Nizhny Novgorod, Faculty of Humanities (Nizhny Novgorod), School of Social Sciences, National Research University Higher School of Economics (25 B. Pecherskaya Str., Nizhny Novgorod, 603155, Russian Federation)
vsibirtseva@hse.ru

Фролова Наталья Хайдаровна, кандидат педагогических наук, факультет гуманитарных наук, доцент департамента лингвистики и иностранных языков Национального исследовательского университета «Высшая школа экономики» (ул. Б. Печерская, 25, Нижний Новгород, 603155, Россия)

Natalia H. Frolova, Associate Professor: HSE Campus in Nizhny Novgorod, Faculty of Humanities (Nizhny Novgorod), School of Applied Linguistics and Foreign Languages, National Research University Higher School of Economics (25 B. Pecherskaya Str., Nizhny Novgorod, 603155, Russian Federation)
nfrolova@hse.ru

УДК 811.161.1'373

DOI 10.25205/1818-7935-2019-17-1-102-114

О любопытстве и смежных понятиях в русском языковом сознании

А. Е. Бочкарев

*Национальный исследовательский университет «Высшая школа экономики»
Нижний Новгород, Россия*

Аннотация

Статья посвящена изучению любопытства в плане концептуальной его конфигурации в русском языковом сознании. Обращение к Национальному корпусу русского языка представляется в этой связи тем более оправданным и необходимым, что содержание анализируемого концепта, как и экспликация связанных с ним ситуаций и установок мнения, не представлено в «готовом виде» ни в одном из отдельно взятых случаев употребления, а реконструируется на совокупности всех возможных употреблений.

По данным Национального корпуса русского языка, рус. *любопытство* входит в отношения дополнительности с *интересом*, *удивлением*, *волнением*, *нетерпением*, *надеждой*, *желанием*, *скукой*, *праздностью*, *пороком*, *похотью*, *грехом*, позволяющими реконструировать основные ситуации любопытства в отношении к сопутствующим переживаниям, поступкам и действиям, а также базисным установкам мнения, задающим оценку анализируемого чувства сообразно избираемой по случаю «концептуальной схеме» как *любопытства-познавательного интереса*, *любопытства-скуки*, *любопытства-праздности* или *любопытства-греха*.

Судя по засвидетельствованным употреблениям, приложимые к любопытству предикаты варьируются в зависимости от позиции в синтаксической структуре высказывания. В функции семантического объекта к любопытству приложимы предикаты *испытывать*, *внушать*, *удовлетворять*, *выказывать*, *сдерживать*, *скрывать*; в функции субъекта – *появляться*, *просыпаться*, *охватывать*, *снестать*, *распирать*, *терзать*, *побеждать*, *говорить*, *заставлять*, *одолевать*, *побуждать*, *губить* и проч.

В метафорическом отображении рус. *любопытство* определяется в терминах пропозициональной модели других предметных областей, о чем свидетельствуют избираемые по случаю дескрипции вида *говорить*, *советовать*, *руководить* ('внутренний голос'), *просыпаться*, *пробуждаться* ('живое существо'), *мучить*, *грызть*, *глодать* ('злой зверек'), *влечь*, *побуждать*, *одолевать* ('неподвластная сила'), *разгораться*, *разжигаться*, *жечь* ('огонь') в функции семантического предиката. Причем аргументом в пользу такого употребления, а заодно и интерпретантом метафорической транспозиции, служит не подлежащее сомнению мнение, что для обозначения особо примечательных свойств любопытства нет более подходящих вспомогательных объектов, кроме разве только внутреннего голоса, силы, огня и живого существа.

Ключевые слова

чувства и эмоции, языковая картина мира, лингвоспецифичные слова, установка мнения, пропозициональная модель, концептуальная метафора, аксиологические нормы

Для цитирования

Бочкарев А. Е. О любопытстве и смежных понятиях в русском языковом сознании // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 102–114. DOI 10.25205/1818-7935-2019-17-1-102-114

Curiosity and Related Concepts in the Russian Linguistic Consciousness

Andrey E. Bochkarev

*National Research University "Higher School of Economics"
Nizhny Novgorod, Russian Federation*

Abstract

The article is devoted to *lyubopytstvo* 'curiosity' as a language-specific word in order to show its specific conceptual configuration in the Russian language consciousness. In this regard, the National Russian Corpus is more appropriate,

© А. Е. Бочкарев, 2019

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Том 17, № 1
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2019, vol. 17, no. 1

because a conceptual configuration of an analyzed concept is not present in a “finished” form in any single utterance, but may be reconstructed only on the totality of all possible utterances. It can be manifested in many different ways: distribution, ability to accumulate some Russian “key ideas”, predisposition to be associated with some emotional attitudes, concepts, propositional and metaphorical models.

According to the National Russian Corpus, curiosity is usually felt for everything which may be of interest, and defies the imagination: another man’s life, news and policy, death, abroad and foreigners, origins and workings of the universe, friends’ husband’s salary, danger and suspense, someone’s life stories, scientific discoveries, etc. In different contexts, curiosity is defined in relation to interest, surprise, excitement, hope, desire, idleness, sin, etc., that allow us to reconstruct some conventional situations of curiosity, as well as related feelings, acts, opinions and axiological norms, in conformity with different “conceptual schemas” of curiosity as a cognitive interest, boredom, idleness, or sin.

The propositional model provides information that predicates applied to *lyubopytstvo* ‘curiosity’ vary with the position in the syntactic structure of the proposition. As a semantic object curiosity is felt, constrained, excited, masked, and satisfied; as a semantic subject it appears, covers, grows, and encourages.

In the metaphoric mapping, *lyubopytstvo* ‘curiosity’ is redefined over categorical boundaries in terms of a propositional model appropriated for an inner voice, a human being, a living creature, an inevitable force, or a flammable mixture. By analogy with an inner voice it calls, tells and counsels; by analogy with a living creature it is waking up, brings out; by analogy with the beast it gnaws and bites; by analogy with an inevitable force it covers, overcomes, leads and wins; by analogy with a peculiarly flammable mixture it inflames and burns. Such use becomes so common that native speakers don’t pay more attention to metaphorical expressions like curiosity killed someone or to burn and consume oneself in curiosity but take them almost for the authentic characteristic of curiosity.

Keywords

feelings and emotions, language picture of the world, language-specific words, opinion, propositional attitude, conceptual metaphor, axiological norms

For citation

Bochkarev, Andrey E. Curiosity and Related Concepts in the Russian Linguistic Consciousness. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 102–114. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-102-114

Вводные замечания

О любопытстве сказано немало – и философами, и психологами, и культурологами, и антропологами, определяющими искомое понятие в связи с врожденной склонностью искать новое в окружении, потребностью к преодолению энтропии, стремлением к познанию, поведенческими нормами, предписаниями совести и морали, внешней и внутренней необходимостью поступка, интересом, потребностью и необходимостью [Berlyne, 1966; Loewenstein, 1994; Изард, 1999; Ильин, 2001; Рябцева, 2003; Litman et al., 2005; Елшанский и др., 2016; Аванесян, Башмакова, 2017].

Малоизученным остается вместе с тем определение любопытства как «состояния души», как, впрочем, и обследование лингвоспецифических особенностей обозначающих его выражений на значительном корпусе употреблений. В данном случае недостаточно сказать, что *любопытство* противопоставляется *любопытности* в рамках характерной для русского языка аксиологической поляризации ‘достойно’ – ‘недостойно’, ‘хорошо’ – ‘плохо’ [Апресян, 1995. С. 351; Шмелев, 2005. С. 398]; требуется обследовать, кроме того, и некоторые другие параметры, в том числе дистрибутивные возможности, предметно-понятийные соответствия, способность аккумулировать «ключевые идеи» русской языковой картины мира, предрасположенность ассоциироваться с определенными переживаниями, представлениями и ситуациями [Зализняк, 2015]. Критерием лингвоспецифичности как особой концептуальной конфигурации, характерной именно для данного языка, становятся в таком случае специфические условия употребления, гарантом достоверности полученных результатов – представительный корпус исследовательского материала.

Предметная область

Обратившись к Национальному корпусу русского языка¹, можно в первом приближении установить, что наряду с соответствующим именным выражением означенное эмоциональ-

¹ URL: <http://ruscorpora.ru/>.

ное состояние выражается предикатами состояния, в том числе глагольными выражениями *испытывать любопытство, быть вне себя от любопытства*, прилагательным *любопытный*, наречием *любопытно*, и что квазисинонимами *любопытства* могут быть *любопытность, интерес, внимание*. При этом, как показывают примеры общим числом свыше тринадцати тысяч вхождений, причин для любопытства в русском языковом сознании много, как много и ситуаций, в которых испытывают чувство любопытства.

Судя по засвидетельствованным употреблением, сферой действия глагола или любого другого пропозиционального оператора в значении любопытства становится, по сути, всё, что представляет интерес, поражает воображение, будоражит фантазию, представляется новым, незнакомым, запретным: положение дел, неизвестность, взрослая жизнь, противоположный пол, театральная премьера, чужая жизнь, мучения, смерть, заграница, посторонний мужчина, запертая на ключ кладовка, научное открытие, новости политики, содержимое котомки, коллекция ключей, незнакомец, возраст собеседника, устройство вселенной, обнаженное тело, незнакомая книга, невиданные кушанья, житейская история, зарплата мужа подруги, тайна, дизайн автомобиля и проч.² Из любопытства (*просто из любопытства, любопытства ради*) совершают самые разные, порой необдуманные поступки: отправляются в неизвестном направлении, идут в театр посмотреть на заезжую знаменитость, заглядывают в словарь иностранных слов, прислушиваются к разговору случайных встречаемых, всматриваются в лица прохожих, разглядывают части своего или чужого тела, изучают философию, занимаются политикой, интересуются подробностями дела, расспрашивают, подсматривают, отправляются на край света, лишаются девственности, впутываются в истории, делают глупости, знакомятся, изменяют, приобщаются к наркотикам и просто живут.

Ситуаций любопытства действительно много, как много в релевантном для человека контексте жизни причин и поводов для любопытства. Но общей для всех ситуаций любопытства остается, несмотря на различия, прототипическая модель: *X* вызывает любопытство *Y*-а или *Y* испытывает любопытство к *X*-у (где *X* – предмет любопытства, *Y* – субъект любопытства). Отчасти сходной оказывается, кроме того, и аффективная реакция, некоторое внутреннее состояние *R*, судить о котором можно по эксплицированным в контексте процессам и действиям *P*, в которых это состояние проявляется.

Симптоматика любопытства

С симптоматикой исследуемого эмоционального состояния соотносятся, в частности, засвидетельствованные в корпусе указания на выражение лица: *На рассеяннo-улыбчивых лицах, оживленных как бы беспредметно, появилось любопытство* (Д. Гранин), голос: *Тайный жар любопытства придавал его голосу воркующие нотки* (Ф. Искандер), взгляд: *В ее взгляде было жгучее кошачье любопытство к моим потрохам и уверенность в хозяине* (Ф. Искандер), глаза: *В глазах его даже появилось что-то вроде любопытства ко мне* (В. Войнович), губы: *Губы ее поджаты от любопытства* (А. Е. Рекемчук), рот: *Витяка с разинутым от любопытства ртом* (А. Е. Рекемчук), вытянутую шею: *Что? – с любопытством спросила Лэсси, вытягивая шею* (Е. Хаецкая), покраснение: *Пунцовая от любопытства Клодин нетерпеливо переминалась с ноги на ногу на крыльце дома* (В. Михальский), дрожь: *И такое любопытство меня охватило, что дрожь по телу пошла* (А. С. Новиков-Прибой).

На физическую и психологическую реакцию тела в состоянии любопытства указывают, кроме того, употребленные по случаю глагольные выражения с причинным предлогом *от*. Как показывают примеры, от любопытства округляются глаза: *Глаза девицы округляют-*

² В интересах сопоставительного анализа ср. предметную область любопытства в изложении Плутарха. Здесь и страсть узнавать об укромном и тайном, и желание услаждать слух рассказами про растление девицы, распутство женщины, готовящуюся тяжбу, распрю братьев, и желание разузнать, что творится разом всюду: в домах богачей и в лачугах бедняков, в царских палатах и в спальнях новобрачных, и стремление познать, отчего одни растения цветут и зеленеют, а другие, разом расточивши свои запасы, остаются голы и нищи [Плутарх, 1983].

ся – в них испуг и любопытство (Г. Алексеев), прищуривают глаза: *...и прищурил глаза от любопытства* (А. Белянин), загораются глаза: *...загорелись от любопытства глаза ученого* (В. Постников), бледнеют: *...побледнев от любопытства* (С. Липкин), горят уши: *У Люси с утра до вечера нестерпимо горели от любопытства тонкие уши, будто кто-то долго и с наслаждением их драл* (К. Г. Паустовский), краснеют: *Волк раскраснелся от любопытства* (М. Петросян), загораются: *...прямо загорелась от любопытства* (Л. Дворецкий), горят: *...и весь горел от любопытства* (Н. В. Кожевникова), сгорают: *Просто сгораю от любопытства!* (М. А. Алданов), замирают: *Замирая от любопытства и интереса, я заглянул в клеенчатую тетрадь Людмилы Сильвестровны* (М. А. Булгаков), трепещут: *Ну и что? – спросил я, трепеща от любопытства* (Ф. Искандер), тают: *...прямо таяла от любопытства* (А. Приставкин), изнывают: *Симочка изнывала от любопытства узнать, о чем он пишет и для кого* (А. Солженицын), изнемогают: *Она изнемогала от любопытства* (Г. Алексеев), заболевают: *...а я могу заболеть от любопытства* (Т. Окуневская) и чуть ли даже не умирают: *Она явно умирала от любопытства* (М. А. Кучерская).

Параметризация

При сходном механизме любопытства как стремления *Y*-а узнать нечто, представляющее для него интерес (*X*), содержание любопытства варьируется вместе с тем от ситуации к ситуации, так что раскрыть его можно только с учетом таких переменных, как основание, область действия, интенсивность, физические ощущения, сопутствующие эмоционально-психологические переживания, предметно-понятийные соответствия и, что немаловажно, оценка с противоположным аксиологическим знаменателем в фокусе базисных суждений вида *Из любопытства выросла фундаментальная наука* (Г. Горелик), *...это движущая сила как истории, так и современности* (В. А. Каверин), *...любопытство <...> сгубило род человеческий* (М. Ю. Лермонтов), *Любопытной Варваре на базаре нос оторвали* и проч., на которых базируются культура и социальные институты вроде семьи, церкви и государства³. В одном случае любопытство воспринимается как естественная потребность человека к обретению новых знаний, как движущая сила познания: ср. *живое исследовательское любопытство* (С. С. Аверинцев); в другом переосмысливается как нечто предосудительное, порицаемое и заведомо вредное для человека и окружающих: ср. *Любопытство – худший из человеческих пороков* (Н. Александров).

По основанию любопытство может быть природным: *...и теперь его одолело природное любопытство* (Т. Сахарова), врожденным: *Но, думаю, мое врожденное любопытство как-нибудь переживет подобный удар* (Т. Сахарова), естественным: *...любопытство мое слишком естественно* (Ф. М. Достоевский), невольным: *Нет, не ожидание, а какое-то невольное любопытство* (В. Брагин), законным: *Он умирал от законного любопытства* (В. Ф. Панова), особым: *Сие чудовище ни в ком из нас сначала не возбудило особого любопытства* (Ф. М. Достоевский), обыкновенным: *У меня пропало обыкновенное человеческое любопытство* (И. Меттер), чистым: *Позвонила только из чистого любопытства* (Т. Богомолова), простым: *...я со своего столичной физиономией не возбудил в ней даже простого любопытства* (А. П. Чехов), отвлеченным: *...а тот смотрел с каким-то отвлеченным любопытством* (В. В. Набоков), пустым: *...из пустого любопытства* (А. Солженицын), сексуальным: *...пробуждение сексуального любопытства* (В. А. Гиляровский), эротическим: *...страдали болезненным эротическим любопытством* (М. А. Кузмин), половым: *В этих вопросах не содержится еще никакого особенного полового любопытства* (К. И. Чуковский);

³ С излагаемыми в базисных суждениях аксиологическими нормами-формулировками коррелируют отчасти словарные толкования любопытства: 1. Мелочный интерес ко всяким, даже несущественным подробностям. ♦ *Спрашивать из пустого любопытства. Праздное любопытство.* 2. Стремление узнать, увидеть что-н. новое, проявление интереса к чему-н. ♦ *Возбудить, удовлетворить чье-н. любопытство* (С. И. Ожегов).

по области действия – исследовательским: *...исследовательское любопытство* (В. Войнович), научным: *Ну, зачем ради научного любопытства мучить животных?* (А. Волков), профессиональным: *Из профессионального любопытства, не более* (С. Видулов), житейским: *Удовлетворять ваше чисто житейское любопытство я не стану* (А. Маринина), хозяйственным: *Женщины глядели на меня с хозяйственным любопытством* (Ф. Искандер), деревенским: *Румяная молодая с откровенным деревенским любопытством тут же возникла у калитки* (Г. Щербакова), крестьянским: *Я понял, что крестьянское любопытство непобедимо* (Ф. Искандер), всеобщим: *...всеобщее любопытство* (Н. Троицкий), человеческим: *...удовлетворить человеческое любопытство* (В. Крапивин, Д. Струев), людским: *Надо куда-нибудь подальше от людского любопытства, догадок и подглядыв* (С. Д. Кржижановский);

по времени действия – минутным: *...с минутным любопытством смотрели на экипаж* (Л. Н. Толстой), мгновенным: *Вызывает в массах мгновенное любопытство, на это и расчет* (А. Слаповский);

по интенсивности и силе проявления – легким: *...легкое любопытство* (Д. Корецкий), возрастающим: *...с возрастающим любопытством* (А. Вампилов), удвоенным: *Я с удвоенным любопытством посмотрел на него* (И. С. Тургенев), немалым: *Заряд любопытства, которое кошку сгубило, во мне еще был немалый* (В. Войнович), большим: *...я прочитала их с большим любопытством* (А. Мильчин), великим: *Упорство его родителей вселило в него великое любопытство* (М. Д. Чулков), огромным: *Испанцы с огромным любопытством и радостью знакомятся с православным храмом* (А. А. Данилова), чрезвычайным: *Конструктор с чрезвычайным любопытством рылся в чемодане, отвечал на вопросы нехотя* (В. Шукшин), крайним: *...и сгорали крайним любопытством, желая узнать истинную причину* (А. Т. Болотов), чрезмерным: *Одна старуха от чрезмерного любопытства вывалилась из окна, упала и разбилась* (И. Сухих), неумным: *Неумное любопытство* (И. А. Ефремов), неукротимым: *...неукротимое любопытство к жизни* (К. И. Чуковский), бешеным: *...и глаза ее зажглись бешеным любопытством, как всегда, когда речь заходила о чем-нибудь, касающемся неведомого ей в любви* (А. Кабаков);

по внешним проявлениям – неприкрытым: *...с неприкрытым любопытством* (Н. Горланова), нескрываемым: *И что было дальше? – с нескрываемым любопытством спросил Колосенцев* (А. Маринина), неподдельным: *И кого вы ищете на этот раз? – с неподдельным любопытством осведомился главврач* (Т. Сахарова), откровенным: *Парень с откровенным любопытством, как на иностранца, смотрит на старика* (В. Распутин), видимым: *...потом с видимым любопытством спросил* (Ю. Казаков), явным: *...и поглядывал на нас с явным любопытством* (Н. Мандельштам), простодушным: *...с простодушным любопытством наблюдал развитие инцидента* (А. Стругацкий, Б. Стругацкий), искренним: *...изобразив на лице искреннее любопытство* (А. Маринина), наигранным: *...с наигранным любопытством поинтересовалась Алла* (А. Беляков), притворным: *Сердась на притворное любопытство Зелинского, кричал, не сдерживаясь* (К. Фукс), ленивым: *...с ленивым любопытством к самому себе* (В. Аксенов), торопливым: *Я жил в торопливом и голодном любопытстве к завтрашнему дню* (И. Меттер), лихорадочным: *Я с лихорадочным любопытством следил за своей новой жизнью* (А. Т. Аверченко), сдержанным: *Она согласилась и вошла к Бурдалакову в комнату со сдержанным любопытством* (В. Войнович), затаенным: *С затаенным любопытством смотрел он на пейзажи Байкала* (А. И. Алдан-Семенов), молчаливым: *Они разглядывают меня с молчаливым любопытством* (Ю. Трифонов), тайным: *...и я издавна с тайным любопытством и страхом посматриваю на нее* (В. Распутин), живым: *А еще за что? – с живым любопытством спросила она* (А. Маринина), веселым: *А что такое? – с веселым любопытством спросил Бомбардов* (М. А. Булгаков), радостным: *...я их рассматривал с каким-то радостным и ненасытным любопытством* (И. С. Тургенев), восторженным: *...и большие серые глаза его снова загораются восторженным любопытством* (Н. Ф. Гильярди), нежным: *...они*

на меня посмотрели с нежным любопытством (М. Ю. Лермонтов), страстным: Мы здесь со страстным любопытством следим за эволюцией советского человека сквозь его условную, заказную литературу (Г. П. Федотов), неистовым: ...сидели в ложе и с неистовым любопытством ждали выхода на сцену того, чье выступление только что возвестил председатель (В. П. Катаев), озорным: ...было озорное любопытство, перемешанное со стыдом и горечью напрасных сожалений (Б. Окуджава), насмешливым: Рита с насмешливым любопытством посмотрела на сестру (Ю. Вяземский), насмешливо-недоуменным: ...и тот увидел это насмешливо-недоуменное любопытство (В. Гроссман), серьезным: Он, не отрываясь, глядел с недоумением и жадным, серьезным любопытством в лицо этого странного человека (А. И. Куприн), вопрошающим: ...посмотрела с каким-то вопрошающим любопытством (А. И. Куприн), пытливым: ...с пытливым любопытством (В. П. Катаев), снисходительным: ...с выражением снисходительного любопытства (А. Волос), уважительным: ...оглядывала меня с уважительным любопытством (Ф. Искандер), почтительным: Папа с почтительным любопытством смотрел на бабушку (Л. Н. Толстой), благодушным: ...полный благодушного любопытства (А. Н. Толстой), добродушным: Бабы продолжали с добродушным любопытством рассматривать наши синие мундирчики и даже забыли о белье (А. П. Чехов), доброжелательным: ...спросил Харламтий Диогенович с доброжелательным любопытством (Ф. Искандер), застенчивым: ...она взглядывала на нас с выражением застенчивого любопытства (В. Г. Короленко), растерянным: ...и с растерянным любопытством спросила (Р. Погодин), настороженным: ...с настороженным любопытством (В. Войнович), подозрительным: ...и опять мое любопытство казалось подозрительным (Ю. Казаков), равнодушным: Я прочитал в себе равнодушное любопытство палача (В. Личутин), безучастным: ...с безучастным любопытством смотрел на себя со стороны (А. Стругацкий, Б. Стругацкий), боязливым: С боязливым любопытством я приблизился к козлу (В. Астафьев), робким: Он вполз внутрь и с робким любопытством стал озираться (М. Гиголашвили), опасливым: Вытянув шею, он с опасливым любопытством глядел на Юрковского (А. Стругацкий, Б. Стругацкий), испуганным: Сверху, с печки, с испуганным любопытством глядели головы Авдотьи-работницы и хозяйкиной кошки-фаворитки (Ф. М. Достоевский), угрюмым: ...некоторые глядели с угрюмым любопытством (Ю. В. Трифонов), отчужденным: ...с отчужденным любопытством (В. Токарева), странным: ...и с каким-то странным любопытством смотрел на брата (В. Шукшин);

по физическим ощущениям – холодноватым: ...с холодноватым любопытством потребителя всякой экзотической дури размышлял о том, куда это его угораздило проскочить (Л. Улицкая), горячим: ...и сразу же этим возбудил горячее любопытство (Ф. Искандер), жгучим: ...и его охватило жгучее любопытство узнать, правда ли это (Л. А. Чарская), острым: ...Теперь каждый въезд в наш лагерь вызывал острое любопытство (Д. Самойлов), обостренным: Манефа была здесь впервые и с обостренным любопытством прикинула к окну (Е. Хаецкая), щекочущим: ...щекочущее любопытство одержало верх (Д. Симонова), притупленным: ...вылупив глаза, с каким-то притупленным любопытством, и только глухой ропот пробежал иногда с одного конца избы до другого (Д. В. Григорович), тупым: ...и с тупым любопытством смотрит (Л. Н. Толстой), нестерпимым: ...любопытство иногда просто нестерпимое (А. Славовский), желчным: ...подслушивала их разговор с желчным любопытством (В. А. Соллогуб), сладким: И тотчас вновь началось сладкое любопытство друг к другу (А. Г. Колмогоров), сладостным: ...и похабные фотографические карточки не возбуждали сладостного любопытства (А. И. Куприн), сладострастным: ...и я опять увидел в ее глазах сладострастное любопытство (М. П. Арцыбашев), вялым: ...с вялым любопытством наблюдал за перипетиями интриги (А. Маринина), болезненным: Нет ли в этом болезненного любопытства? (Вс. В. Иванов), мучительным: ...уже нет у меня мучительного любопытства к людям иных культур, стран, наций и рас (А. Ким), ненасытным: ...от природы одаренный ненасытным любопытством

вом (М. Е. Салтыков-Щедрин), жадным: ...жадное мальчишеское любопытство (Ю. О. Домбровский), алчным: *Публичные женщины с алчным любопытством смотрели на необычных посетителей* (Г. И. Чулков);

по сопутствующим эмоционально-психологическим переживаниям – печальным: ...с живым и печальным любопытством посмотрел в глаза Пинегину (В. Гроссман), тревожным: ...с тревожным любопытством глядел на прокуратора (М. А. Булгаков), брезгливым: ...и посмотрел на него с брезгливым любопытством (В. Пелевин), злорадным: ...вымолвил он равнодушно, но не без злорадного любопытства (В. Голованов), оскорбительным: *Все они смотрели на него с каким-то оскорбительным любопытством* (Ф. М. Достоевский), мстительным: *И новое, мстительное любопытство, с которым Федор Иванович не мог совладать, толкнуло его на третий заход* (В. Дудинцев), страшным: *Сквозь собственное уныние, со страшным любопытством (как дети сквозь плач) я прислушиваюсь: ошибётся он или нет?* (Ф. Искандер), жутким: ...увлекаемый жутким любопытством и, кажется, еще другим чувством (А. П. Хейдок), жестоким: ...жестокое любопытство публики (А. Ф. Кони), враждебным: ...старалась отвлечь это враждебное любопытство, которое я ощущал у себя на затылке (Ф. Искандер), неудовлетворенным: ...детское неудовлетворенное любопытство давно было забыто (Ю. Бондарев);

по морально-этической оценке – мелочным: *Я, однако, не мог победить в себе мелочного любопытства* (Г. А. Газданов), мелким: *Да и злорадное мелкое любопытство мучило ее* (З. Н. Гиппиус), никчемным: *Меня все еще удивляло это никчемное любопытство* (Р. В. Иванов-Разумник), бесплодным: ...чтобы не будить в себе бесплодного любопытства к чужим жизням и смертям (В. Конецкий), неразумным: *Я с пристальным вниманием и неразумным любопытством провинциала рассматривал в гостиной картины, фотографии, книги, благоговей перед святой стариной* (В. Астафьев), бессмысленным: ...упрекая себя за бессмысленное любопытство (С. Носов), глупым: ...из глупого любопытства (Г. С. Эфрон), невинным: *А что вам нравится в женщинах? – спросила Ася, закинув голову с невинным любопытством* (И. С. Тургенев), легкомысленным: ...легкомысленное любопытство или желание кому-либо отомстить (Г. Львов), нескромным: ...все очень нетерпеливо желали видеть и потому рассматривали ее с несколько нескромным любопытством (Н. С. Лесков), неуместным: *Он посмеялся над скучным анекдотом, сетуя в душе на свое неуместное любопытство* (В. Валеева), излишним: *Следовало бы посадить вас недельки на две за излишнее любопытство* (К. Г. Паустовский), назойливым: ...и делалось неприятно от этого назойливого любопытства (А. Белый), бесцеремонным: *Сельчане с бесцеремонным любопытством рассматривали «служилого»* (Г. М. Марков), бестактным: *Неприятно было также видеть, с каким бестактным любопытством наши офицеры рассматривали иностранцев* (А. А. Игнатъев), досадным: ...избежать досадного любопытства глупой черни (В. Т. Нарезный), неприличным: *А это неприличное любопытство! Что ела, что пила?* (Д. Донцова), постыдным: ...и вовсе не из постыдного любопытства (И. Андреев), бесстыдным: ...бесстыдное любопытство (Б. Хазанов), нездоровым: ...и нездоровое любопытство зажигалось в ее глазах (Л. А. Чарская), патологическим: ...и нет в ней этого нервного патологического любопытства к греху (А. В. Амфитеатров), нехорошим: ...с плохо скрываемым, нехорошим любопытством (Е. Евтушенко), дурным: ...спрашиваю я не без дурного любопытства (Г. Д. Катанян), гадливым: *И ты тоже? – спросила она с гадливым любопытством* (Л. М. Леонов), подленьким: ...бабы на тебя смотрят с обидной жалостью да подленьким любопытством (Ф. Абрамов), подоночным: *Обстановка паскудная, нездоровая ажитация, подоночное любопытство* (И. Э. Кио), свинячьим: *Чтобы вы удовлетворили свое свинячье чувство любопытства?* (Э. Володарский), обывательским: ...любопытство было скорее обывательским, чем естествоиспытательским (Д. Симонова), заурядным: *Зачем ей нужно про меня все знать? Заурядное любопытство?* (О. Новикова), вульгарным: *Честно говоря, Киру раздирало на части вульгарное любопытство* (Д. Донцова), пошлым: ...пошлое любопытство (Б. Л. Пастернак),

бабьим: ...*просто бабье любопытство – что же ты за зверь такой?* (Ю. О. Домбровский), суетным: *На лице Ксении выразилось суетное любопытство, о причинах которого он догадывался* (К. Симонов), праздным: *Тут не праздное любопытство, а реальное дело* (Ю. Трифонов), греховным: ...*тут не одно лишь легкомыслие, но и греховное любопытство* (М. Попов), лукавым: ...*лукавое любопытство* (С. А. Дангулов), преступным: ...*замирая от смешанного с омерзением преступного любопытства* (Л. Юзефович), наказуемым: *Всякий раз любопытство оказывается наказанным* (Ю. К. Олеша);

по участию субъекта – женским ⁴: ...*но их подводило извечное женское любопытство* (Ф. Искандер), бабьим: ...*бабье любопытство, известное дело* (М. А. Алданов), детским: ...*с детским любопытством первооткрытия* (А. Слаповский), ребяческим: ...*просто с одним ребяческим любопытством: попробую!* (М. Е. Салтыков-Щедрин), мальчишеским: ...*мальчишеское любопытство непобедимо* (В. Астафьев), юношеским: *Было только естественное юношеское любопытство ко всему новому* (Л. Гурченко), мужским: *А с Ольгой взяло просто мужское любопытство* (О. Ворошилова), стариковским: *С этим и пришли стариковским любопытством: поглядеть, много ли придет провожать* (Ю. Трифонов).

Предметно-понятийные соответствия

Реконструировать возможные ситуации любопытства позволяют также установленные в корпусе корреляции. Как свидетельствуют примеры из Национального корпуса русского языка, *любопытство* сочетается по принципу дополнительности с любознательностью: ...*на ее лице и восторг, и любознательность, и любопытство, чуть-чуть детское* (С. А. Дангулов), интересом: ...*его охватывает вдруг интерес, род любопытства* (В. Маканин), вниманием: ...*я слушаю еще раз со вниманием и новым любопытством* (А. И. Куприн), жаждой знаний: *Одной из особенностей человека и человечества является любопытство и жажда знаний* (Д. Зимин), чем-то новым: ...*из любопытства ожидаешь чего-то нового* (М. Ю. Лермонтов), удивлением: ...*она посмотрела на меня с удивлением и любопытством* (Г. А. Газданов), изумлением: ...*с любопытством и изумлением наблюдали за светилом мировой физики* (А. Стругацкий, Б. Стругацкий), восхищением: ...*с любопытством и восхищением разглядывает Васю* (А. А. Образцов), недоумением: *Император по-прежнему внимательно, с жутким любопытством и недоумением, читал полицейские доклады о террористах* (М. А. Алданов), робостью: ...*начинали смотреть с любопытством и некоторой робостью* (М. Чулаки), замешательством: ...*не то с любопытством, не то с замешательством* (А. Битов), ожиданием: *Глаза круглые, неглупые, несколько усталые, но тотчас засветились любопытством и неким немym ожиданием* (В. Шукшин), нетерпением: *Любопытство и нетерпение овладели мной* (С. Д. Кржижановский), радостью: *Испанцы с огромным любопытством и радостью знакомятся с православным храмом* (А. А. Данилова), любованием: ...*со смешанным чувством любопытства <...> и некоторого даже любования* (К. Симонов), симпатией: *На него с особенным любопытством и симпатией смотрит публика* (А. С. Макаренко), нежностью: ...*чувство какой-то особенной нежности, смешанной с любопытством* (Н. Э. Гейнце), волнением: ...*охваченный внезапным волнением и любопытством* (Б. Хазанов), фантазией: *Теперь-то я понимаю, что примитивное любопытство дало пищу фантазии* (М. Шишкин), желанием: ...*с любопытством и затаенным желанием* (А. И. Куприн), надеждой: ...*с тайным любопытством, а может быть, и с надеждой* (В. Соловьев), возбуждением: ...*замечала на их лицах какое-то острое любопытство, прикрываемое возбуждением* (М. Палей), азартом: ...*азарт тревожного любопытства: а вдруг?* (Ф. Искандер), удовольствием: *Он запомнил разговорившее от любопытства и удовольствия лицо соседки* (В. Ф. Кормер), истомой: *Но он*

⁴ По частотным характеристикам любопытством отличаются в значительно большей степени женщины, чем мужчины.

чувствовал в себе только ночную истому да любопытство (Ф. К. Сологуб), любовью: *Какая это была тайная обжираловка любви и любопытства!* (Ф. Искандер), наивностью: *...проглянут наивность и любопытство* (В. А. Солоухин), неосторожностью: *Руська и два его товарища имели неосторожность (ну, и любопытство тоже) раза два удить с американцами рыбу* (А. Солженицын), почтением: *Своей таинственной и замкнутой жизнью он вызывал любопытство и почтение* (Д. Самойлов), уважением: *К преподавателям я относился с любопытством, но без должного уважения* (С. Довлатов), беспокойством: *Потом, не в силах преодолеть то ли беспокойство, то ли любопытство, спросил, безразлично глядя на Вику* (М. Зосимкина), осторожностью: *...любопытство в нем одолело осторожность* (А. П. Платонов), опасением: *...любопытство пересиливало в них возможные опасения* (Б. Окуджава), грустью: *Грусть и любопытство больших глаз делали ее взор еще тяжелее* (К. А. Федин), боязнью: *Кудыка покряхтел, раздираемый надвое любопытством и боязнью* (Е. Лукин), испугом: *Скоро в углу около дверей образовалась темная куча людей, лица которых с любопытством и испугом тянулись из-за плечей стоявших впереди* (В. Г. Короленко), содроганием: *...с содроганием и любопытством* (В. В. Набоков), страхом: *...смесь любопытства и страха* (А. Куприн), паникой: *Сочетание любопытства с паникой, в которую он ее повергал, образовывало странную волнующую смесь* (Т. Устинова), ужасом: *...вглядывается с ужасом и любопытством в невозмутимого преступника* (В. В. Набоков), стыдом: *Усмехаясь от гадливого, пополам со стыдом, любопытства, он осваивал свою долю у себя на квартире* (Л. М. Леонов), омерзением: *И впечатления были подобные: смесь омерзения и любопытства* (А. Иванов), безразличностью: *Я смотрела на его вихры со смешанным чувством безразличности и любопытства* (К. Сурикова), искушением: *...боролась с этим искушением и победить беса любопытства не могла* (А. П. Чапыгин), чувственностью: *...женщина отдается не из-за чувственности, а из любопытства* (Вс. В. Иванов), супружеской изменой: *Даже не измена, я жену люблю. Это будет только из любопытства* (Е. Белкина), пороком: *Любопытство – худший из человеческих пороков* (Н. Александрова), грехом: *...и кто без греха любопытства* (Т. Соломатина), первородным грехом: *Но на дне сердца затаилась еще любовь вместе с первородным грехом, любопытством* (И. И. Лажечников), смертным грехом: *Но мы потакаем смертному греху: любопытству* (А. Лазарчук), тщеславием: *...из собственного любопытства и тщеславия* (С. Осипов), жадностью: *И вновь неистребимые жадность и любопытство охватывают меня* (В. В. Вишневский), завистью: *Любопытство и зависть Кости разгораются еще больше* (Н. Дубов), похотью: *...странное сочетание страха, похоти и вялого любопытства* (С. Осипов), бездельем: *А любопытство от безделья усугубляется* (А. Ларина), скукой: *...от скуки и любопытства* (А. Н. Толстой), праздностью: *...длинноногая праздность таит любопытство* (А. Белый), злом: *...источником зла является любопытство, лживость, похотливость* (В. В. Вересаев).

Таким образом, с любопытством ассоциируются в языке самые разнообразные понятия – и интерес, и восхищение, и жажда знаний, и грусть, и страх, и тщеславие, и стыд, и безделье, и праздность, и измена, и грех. Основанием тому служит, как видно, и многообразие ситуаций: ср. *...остался у нее в гостинице из простого мужского любопытства* (И. Муравьева); и потребность обозначить причину или предмет любопытства: ср. *...с радостным любопытством к жизни* (И. А. Бунин); и стремление эксплицировать сопутствующие переживания: ср. *...был и страх, и жгучее любопытство* (В. Шукшин); и колебания в выборе номинации, именовать ли данное эмоциональное состояние как *любопытство*, *любопытливость*, *тщеславие* или *интерес*: ср. *...приходят из любопытства или тщеславия* (А. Найман); и разнообразие исторически подвижных аксиологических норм, задающих оценку анализируемого чувства как производную от скуки, праздности или познавательного интереса как *любопытство-познавательный интерес* (эпистемологическое любопытство), *любопытство-скука*, *любопытство-праздность* или *любопытство-грех*.

Пропозициональная модель

Пропозициональная модель, построенная на Национальном корпусе русского языка, предоставляет, кроме того, информацию о том, что приложимые к *любопытству* действия-предикаты варьируются в зависимости от позиции в синтаксической структуре высказывания.

В функции семантического объекта к *любопытству* приложимы предикаты вызывать: *...у нас в галерее мужчин заметный недокомплект, а потому каждый новый представитель сильного пола вызывает легкий ажиотаж и нездоровое любопытство* (И. Павская), испытывать: *К новым блюдам я всегда испытываю любопытство* (М. Спиричева), доставлять: *...монолог доставляет удовольствие или хотя бы вызывает любопытство* (Б. Рогинский), дразнить: *...мне хотелось подразнить любопытство Маши* (В. В. Вересаев), возбуждать: *Она возбуждала любопытство* (П. Д. Боборыкин), будить: *...будить ее любопытство* (А. Н. Бенуа), подстрекать: *Ужели он только хотел подстрекнуть женское любопытство?* (М. Ю. Лермонтов), разжигать: *Это страшно разжигает женское любопытство* (Д. И. Хармс), удовлетворять: *...он не может удовлетворить их любопытство к своей работе* (В. Г. Распутин), внушать: *...внушает отвращение, брезгливое любопытство и страх* (А. Азольский), проявлять: *...очень рано проявила любопытство ко всему взрослому* (А. Слаповский), выказывать: *Любопытства ко мне он не выказал* (В. Конецкий), выставлять: *...неловко так выставить свое любопытство* (А. Иличевский), изображать: *Думаю, сначала она изображала любопытство к нему, чтобы вызвать во мне ревность* (В. Пелевин), сдерживать: *Не в силах сдержать любопытства, я приоткрыла дверь и стала в щелку наблюдать за мужем* (Д. Донцова), утолять: *...утолить любопытство* (С. Есин), насыщать: *...я насытил свое любопытство, сколько было мне нужно* (Ф. И. Буслаев), скрывать: *...нам плохо удавалось скрыть любопытство, с которым мы разглядывали ее* (Н. И. Ильина).

В функции субъекта *любопытство* появляется: *На рассеянно-улыбчивых лицах, оживленных как бы беспредметно, появилось любопытство* (Д. Гранин), зарождается: *Зародилось какое-то жуткое чувство любопытства* (С. Н. Сергеев-Ценский), мелькает: *...и в том и в другом глазу у него мелькнуло любопытство* (С. Залыгин), охватывает: *...ее охватило любопытство* (О. Онойко), растет: *У него все росло смешанное чувство любопытства к Тихону и страха перед этим скуластым, непонятным мужиком* (М. Горький), беспокоит: *...любопытство начинало меня беспокоить* (А. С. Пушкин), волнует: *...любопытство ее тяжело волнует* (М. Горький), снедает: *Любопытство снедало: а что же там, в гастрономе, случилось* (А. Азольский), разбирает: *Вас разбирает любопытство?* (Ю. Трифонов), пронимает: *Но меня проняло невольное любопытство* (С. М. Голицын), раздирает: *Любопытство раздирало меня на кусочки* (И. Павская), заедает: *Бедняка еще больше любопытство заедает* (А. С. Новиков-Прибой), разъедает: *...меня прямо-таки разъедало любопытство к жизни; и я из любопытства впутывался в самые невероятные истории* (Е. Евтушенко), терзает: *Меня просто терзает любопытство. Таких удивительных вещей я еще никогда не видал* (М. А. Булгаков), бурлит: *Город забурился от любопытства* (С. М. Голицын), пересиливает: *...любопытство пересиливало страх* (А. Григоренко), побеждает: *Однако рано или поздно любопытство обязательно побеждало* (И. Нечаева), заносит: *...куда меня занесло любопытство* (С. Б. Рассадин), губит: *Просто меня губит любопытство* (Л. М. Леонов), исчезает: *Их было столько, что исчезало даже любопытство к этим промокшим женщинам с бумажными розами на шляпах или просто на животе* (А. Н. Толстой) и т. д.

Смысловые преобразования

В метафорическом отображении *любопытство* переосмысливается в зависимости от избираемой по случаю установки мнения и в качестве внутреннего голоса, и в качестве живого существа, и неподвластной человеку силы, и злобного зверька, и огненной стихии, и исчис-

ляемой по объему субстанции. Отбору соответственно подлежат предикатные дескрипции, подходящие главным образом для внутреннего голоса, живого существа, силы, зверька, силы или огня.

В качестве **внутреннего голоса** любопытство говорит: *Но любопытство заговорило в ней* (Н. Садур), советует: *Чувство любопытства с таким убеждением советовало мне узнать, какие были эти бумаги* (Л. Н. Толстой); в качестве **живого существа** – зарождается: *Зародилось какое-то жуткое чувство любопытства* (С. Н. Сергеев-Ценский), просыпается: *...у меня проснулось любопытство исследователя* (Ю. Сенкевич), пробуждается: *...в коллегах пробудилось любопытство* (Ю. Гутова, Е. Романенко), шевелится: *У меня начинает шевелиться любопытство: какие мысли копошатся в графской голове?* (А. С. Новиков-Прибой); в качестве **злобного зверька** – мучает: *Ее мучит любопытство* (М. М. Пришвин), толкает: *...она поняла, что ее толкает любопытство* (С. Н. Сергеев-Ценский), гложет: *Но гложет нездоровое любопытство* (В. Дубовский), грызет: *...любопытство начинает уже грызть* (митрополит Антоний), насыщается: *Его любопытство уже насытилось* (В. Дудинцев); в качестве **силы** – руководит: *...и сыщиками руководит любопытство* (А. Вайнер, Г. Вайнер), охватывает: *Тогда его охватило любопытство* (К. Симонов), распирает: *Но меня просто распирало от любопытства* (А. Моторов), влечет: *...жгучее любопытство, которое властно влекло его к женщине* (М. П. Арцыбашев), заносит: *...куда меня занесло любопытство* (С. Б. Рассадин), побуждает: *...любопытство могло побудить Михаила Юрьевича распечатать чужое письмо* (П. А. Висковатов), заставляет: *Любопытство заставляет ее заглянуть туда и сделать страшное открытие* (В. Коршунов), пересиливает: *...любопытство пересилило* (Е. Топильская), берет: *И вдруг меня такое любопытство взяло* (Э. Радзинский), перебарывает: *Постепенно мужское любопытство перебороло эмоции* (А. Троицкий), одолевает: *В дороге меня одолевало любопытство* (Н. Н. Пунин), берет верх: *Но любопытство взяло верх, и комендант постучался* (М. Сосенкова), владеет: *Филиппом сейчас более всего владело любопытство* (Г. Полонский), губит: *Любопытство сгубило кошку. Любопытство погубит и Лизавету* (Е. Козырева); в качестве **огненной стихии** – разгорается: *Узнал он, что я бывал за границей, и любопытство его разгорелось* (И. С. Тургенев), разжигается: *Любопытство мое разожглось* (Г. В. Иванов), жжет: *Любопытство так и жгло Марата* (Ф. Искандер), в качестве **исчисляемой по объему субстанции** – иссекает: *Если любопытство не иссякло, можно осмотреть еще несколько квадратов и перейти к прямоугольникам* (С. Юрский).

Разумеется, в условно-истинностном определении метафорические дескрипции вида *любопытство гложет* или *любопытство жжет* представляются неприемлемыми и даже, можно сказать, ложными по сравнению с дескрипциями вида *испытывать любопытство*. Но если о соответствии присоединенного предиката судить по речевому обиходу, о сообразности пропозиции следует тогда судить только и только по тому, насколько обиходным является такое словопотребление. Больше того, такое понимание любопытства укореняется в обыденном сознании настолько, что носители языка не обращают более внимания на метафоричность выражений вида *тайный жар любопытства* (Ф. Искандер), а принимают их чуть ли не за объективную характеристику того, чем является «на самом деле» любопытство. Аргументом в пользу такого употребления, а заодно и интерпретантом метафорической транспозиции, служит не подлежащее сомнению мнение, что в наиболее характерных своих проявлениях любопытство может действительно овладевать по аналогии с силой, разгораться и жечь по аналогии с огнем, глотать и грызть по аналогии со зверьком, иссякать по аналогии с исчисляемым ресурсом.

Так приходим к следующим выводам.

- Рус. *любопытство*, пусть и имеет переводные эквиваленты в других языках, лингвоспецифично.
- О лингвоспецифичности *любопытства* можно судить и по предрасположенности к определенному употреблению, и по построенной на множестве употреблений пропозициональ-

ной модели, и по способности к специфическим переосмыслениям, и по засвидетельствованным предметно-понятийным соответствиям.

- В наиболее частотных своих употреблении *любопытство* входит в отношения дополнительности с *интересом*, *удивлением*, *волнением*, *нетерпением*, *желанием*, *надеждой*, *желанием*, *скукой*, *праздностью*, *пороком*, *похотью*, *грехом* и проч., позволяющими отобразить возможные ситуации любопытства в отношении к сопутствующим переживаниям и поступкам, а также базисным, пусть и не всегда эксплицированным, установкам мнения.

- Приложимые к *любопытству* предикаты варьируются в зависимости от позиции в синтаксической структуре высказывания. В функции семантического объекта к любопытству приложимы предикаты *испытывать*, *внушать*, *удовлетворять*, *выказывать*, *сдерживать*, *скрывать*; в функции субъекта – *появляться*, *просыпаться*, *охватывать*, *снедать*, *распирать*, *терзать*, *побеждать*, *говорить*, *заставлять*, *одолевать*, *побуждать*, *губить* и проч.

- В метафорическом отображении рус. *любопытство* определяется в терминах пропозициональной модели других предметных областей, о чем свидетельствуют избираемые по случаю дескрипции типа *говорить* ('внутренний голос'), *просыпаться* ('живое существо'), *мучить* ('злостный зверек'), *руководить* ('сила'), *разгораться* ('огонь'), *иссекать* ('ресурс') в функции семантического предиката. Причем аргументом в пользу определения любопытства по аналогии с внутренним голосом, силой, огнем или живым существом служит такое существенное обстоятельство, что для обозначения особо примечательных свойств любопытства в русском языковом сознании нет более подходящих вспомогательных объектов, кроме, быть может, только внутреннего голоса, силы, огня или живого существа.

Список литературы / References

- Аванесян М. О., Башмакова Я. П. Исследование эксплицитных и имплицитных представлений о любопытстве // Вестн. Санкт-Петербургского ун-та. Психология и педагогика. 2017. Т. 7, № 3. С. 235–248.
- Avanesyan, M. O., Bashmakova, I. P. The Study of Explicit and Implicit Representations of Curiosity. *Bulletin of the St. Petersburg University. Psychology and Pedagogy*, 2017, vol. 7, no. 3, p. 235–248. (in Russ.)
- Апресян Ю. Д. Образ человека по данным языка // Апресян Ю. Д. Избр. тр. М.: Языки русской культуры, 1995. Т. 2: Интегральное описание языка и системная лексикография. С. 348–388.
- Apresyan, Yu. D. Human Image According to the Language. In: Апресян, Yu. D. Selected works. Moscow, Yazyki russkoj kul'tury, 1995, vol. 2: Integral Language Description and Systematic Lexicography, p. 348–388. (in Russ.)
- Зализняк Анна А. Лингвоспецифичные единицы русского языка в свете контрастивного корпусного анализа // Компьютерная лингвистика и интеллектуальные технологии: В 2 т. По материалам ежегодной Международной конференции «Диалог». М., 2015. Вып. 14 (21), т. 1. С. 683–695.
- Zalizniak, Anna A. Russian Language-specific Words as an Object of Contrastive Corpus Analysis. In: Computational Linguistics and Intelligent Technologies, Materials of the Annual International Conference "Dialogue", 2015, issue 14 (21), vol. 1, p. 683–695. (in Russ.)
- Елшанский С. П., Ануфриев А. Ф., Камалетдинова З. Ф., Сапарин О. Е., Семенов Д. В. Психометрические показатели русскоязычной версии теста "Склонность к любопытству и исследованию" (CEI-II) // Вестн. Костромского гос. ун-та. Серия: Педагогика. Психология. Социокинетика. 2016. Т. 22, № 2. С. 113–117.
- Yelshansky, S. P., Anufriev, A. F., Efimova, O. S., Semyonov, D. V. The Russian-language version of the "Curiosity and Exploration Inventory" (CEI-II). *Bulletin of the Kostroma State University. Series: Pedagogics. Psychology. Social kinetics*, 2016, vol. 22, no. 2, p. 113–117. (in Russ.)
- Изард К. Э. Психология эмоций. СПб.: Питер, 1999. 464 с.
- Izard, C. The Psychology of Emotions. St. Petersburg, Piter Publ., 2002. (in Russ.)

- Ильин Е. П.** Эмоции и чувства. СПб.: Питер, 2001. 752 с.
Ильин, Е. Emotions and Feelings. St. Petersburg, Piter Publ., 2001, 752 p. (in Russ.)
- Плутарх.** О любопытстве / Пер. Н. В. Брагинской // Плутарх. Соч. М.: Худож. лит., 1983. С. 479–491.
Plutarch. On Being a Busybody. In: Plutarch. Selected Writings. Moscow, Khudozhestvennaya literatura, 1983, p. 479–491. (in Russ.)
- Рябцева Н. К.** Любопытство как организующая и как дезорганизующая сила // Логический анализ языка. Космос и хаос: концептуальные поля порядка и беспорядка / Сост. и отв. ред. Н. Д. Арутюнова. М.: Индрик, 2003. С. 158–190.
Ryabtseva, N. K. Curiosity as an organizing and a disorganizing force. In: Arutyunova N. D. (ed.). Logical analysis of the language. Space and chaos: conceptual fields of order and disorder, Moscow, Indrik, 2003, p. 158–190. (in Russ.)
- Шмелев А. Д.** Плюрализм этических систем в свете языковых данных эмоции // Зализняк Анна А., Левонтина И. Б., Шмелев А. Д. Ключевые идеи русской языковой картины мира: Сб. ст. М.: Языки славянской культуры, 2005. С. 398–409.
Shmelev, A. D. Pluralism of Ethical Systems in the Light of Linguistic Data of Emotion. In: Zalizniak, Anna A., Levontina, I. B., Shmelev, A. D. Key Ideas of the Russian Language Image of the World. Moscow, Yazyki russkoj kul'tury, 2005, p. 398–409. (in Russ.)
- Berlyne, D. E.** Curiosity and exploration. *Science*, 1966, no. 153, p. 25–33.
- Litman, Jordan A., Hutchins, Tiffany L., Russon, Ryan K.** Epistemic curiosity, feeling-of-knowing, and exploratory of behavior. *Cognition and Emotion*, 2005, vol. 19 (4), p. 559–582.
- Loewenstein, G.** The Psychology of curiosity: a Review and Reinterpretation. *Psychological Bulletin*, 1994, vol. 116, p. 75–98.

Материал поступил в редколлегию
 Date of submission
 08.09.2018

Сведения об авторе / Information about the Author

Бочкарев Андрей Евгеньевич, доктор филологических наук, профессор департамента литературы и межкультурной коммуникации, факультет гуманитарных наук Национального исследовательского университета «Высшая школа экономики» (ул. Б. Печерская, 25/12, Нижний Новгород, 603155, Россия)

Andrey E. Bochkarev, Dr. Hab. (Philology), Professor, Department of Literature and Intercultural Communication, Faculty of Human Sciences, National Research University "Higher School of Economics" (25/12 B. Pecherskaya Str., 603155, Nizhniy Novgorod, Russian Federation)

abotchkarev@hse.ru
 bochkarev.andrey@mail.ru
 ORCID 0000-0002-9650-8604
 SPIN-код 3213-6724

Ментальные аспекты лингвистической реконструкции (на материале «Гетики» Иордана)

Ю. М. Трофимова

*Национальный исследовательский
Мордовский государственный университет им. Н. П. Огарева
Саранск, Россия*

Аннотация

Статья имеет целью установление некоторых черт остготского языка VI в., которые доступны для наблюдения в латинском тексте «Гетики». Исследовательская посылка основана на том факте, что текст написан готом на неродном языке и что влияние родного языка должно проявить себя в соответствии с законами интерференции. Это влияние выявляется на довербальном уровне с опорой на понятия текстового менталитета и текстовой ментальности в аспекте прогнозируемого расширения правил лингвистической реконструкции. Обоснование ментальных основ лингвистической реконструкции осуществляется в статье на базе текстовой образности, особенно чувствительной к влиянию национального менталитета. Анализ выявил достаточно ограниченный в своем диапазоне характер образности, соотносящейся с немногочисленными сферами внешней действительности: человек, растительный, животный мир, неодушевленные предметы, природа и очень редкие абстрактные понятия. Поскольку латинская вербализация практически всех перечисленных сфер обнаруживает готские лексические соответствия, зафиксированные в Готской Библии, рассмотрение текстовой образности «Гетики» завершается построением языковой картины мира в том ее реконструируемом виде, который мог быть свойствен текстовому уровню остготского языка VI в. Новизна статьи заключается в развитии практики лингвистической реконструкции в соответствии с современными подходами в языкознании, наиболее продуктивными и популярными в сфере когнитивной лингвистики, активно оперирующей понятиями менталитета и ментальности. Новым является и предложенный способ реконструкции текстовых явлений. Актуальность статьи определяется обращением к историческому текстовому материалу, позволяющему установить истоки современных явлений, характерных для процесса текстопостроения, активно изучаемого на базе разных языков и типов текста.

Ключевые слова

остготский язык, латинский язык, текстовый менталитет, текстовая ментальность, языковая личность, лингвистическая реконструкция, текстовая образность, языковая номинация, довербальный уровень текста

Для цитирования

Трофимова Ю. М. Ментальные аспекты лингвистической реконструкции (на материале «Гетики» Иордана) // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 115–124. DOI 10.25205/1818-7935-2019-17-1-115-124

Linguistic Reconstruction and Its Mental Foundations (Based on Jordanes' *Getica*)

Yulia M. Trofimova

*Ogarev Mordovia State University
Saransk, Russian Federation*

Abstract

The article is aimed at revealing some peculiarities of the Ostrogothic language of the 6th century which are preserved in the Latin text of the *Getica* on the mental level of text formation. This text being written by a Goth who was not a native Latin speaker is sure to contain some Ostrogothic language features accessible for observation due to the phenomenon of linguistic interference. Mother tongue inevitably influences the process of text formation when it is car-

ried out in a foreign language and the manifestations of its realization can be traced in accordance with the principles of linguistic reconstruction.

Purpose. The idea of the article is to show that the procedure of linguistic reconstruction can be enlarged according to the achievements of contemporary linguistics comprising the theory of text and its mental foundations such as textual mentality. The article deals with one of the reconstructed manifestations of the Ostrogothic language of the 6th century, i.e. Jordanes' tropology. Tropology seems to be most sensitive to mental interpretation resulting in its specific realization in the text. It is stated that the tropology of the Latin text of the *Getica* is conditioned by the native Gothic mentality and could have been equally verbalized with the help of the corresponding Gothic lexical equivalents in case Jordanes wrote the *Getica* in Gothic. This type of mentality according to the article is located on the preverbal level of text formation showing that the author when describing a non-verbal situation selects some important from his/her viewpoint elements of the situation and verbalizes them by means of lexical resources of the language he/she uses. Jordanes' tropology as is evident from the article can be evaluated as his choice of the elements of the imaginative description of a text situation.

Results. The analysis shows that Jordanes' tropology is rather restricted in its scope and comes to an inconsiderable in number outer world spheres, i.e. human body, vegetative resources, non-animated objects, facts relating to animals, landscape, water, few natural elements such as lightning, storm and very few abstract phenomena. The point is that nearly all the Latin lexemes used by Jordanes as the foundation of his tropology have Gothic lexical equivalents found in the Gothic Bible which up to the present moment remains the only source of the Gothic language material. The Gothic lexemes equivalent to the Latin text variants are represented in the article in the form of a certain structure relating to the so-called language world picture which could underlie the imaginative manner of interpreting the outer world the by Gothic native speakers and Jordanes in particular.

Conclusion. It is emphasized in the article that the described level of Jordanes' tropology may primarily be indicative of his lingual personality, but on the other hand his lingual personality cannot be isolated from the speech practice characteristic of the Ostrogothic language of that period. The facts stated in the article are to be accepted as one of the spheres of the Gothic text formation inherent in its natural mode of functioning since the Gothic Bible being a word-for-word translation from the Ancient Greek is not illustrative in this regard.

Keywords

The Ostrogothic language, the Latin language, textual mentality, lingual personality, linguistic reconstruction, tropology, lingual nomination, preverbal text level

For citation

Trofimova, Yulia M. Linguistic Reconstruction and Its Mental Foundations (Based on Jordanes' *Getica*). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 115–124. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-115-124

Введение Лингвистическая реконструкция в контексте понятий менталитета и ментальности

Ограниченный объем сохранившихся текстов на готском языке предрасполагает к поиску каких-либо новых, еще неопробованных источников изучения готской языковой специфики, доступной для установления в силу разработанности положений современной лингвистики. Интерес в этой связи вызывает труд Иордана «Гетика» [Иордан, 1960], который написан на латинском языке, но написан готом, жившим в среде, где готский язык еще бытовал в активном употреблении.

Лингвистический анализ текста предполагает непосредственное апеллирование к личности автора, однако в этом случае все факты такого рода оказываются весьма приблизительными. На основе различных умозаключений, сопоставлений и выводов историками и лингвистами были предположительно установлены некоторые факты его биографии и литературной деятельности. С другой стороны, одно немаловажное обстоятельство, которое непосредственно его касается и о котором Иордан сам сообщает в своем труде, уже более века является предметом активной дискуссии. Эта фраза помещена в самом конце «Гетики», где Иордан заверяет читателя, что не прибавил ничего лишнего в пользу племени готов, из которого происходит (*{316} nec me quis in favorem gentis praedictae, quasi ex ipsa trahenti originem, aliqua addidisse credat*).

Данный факт некоторыми исследователями признается достаточным, чтобы считать Иордана готом, однако еще в XIX в. возникло предположение, что он был аланом. Вопрос о его этнической принадлежности дискутируется по сей день, и различные доводы все еще приводятся. Иногда они представляют собой некий третий вариант интерпретации упомянутого факта, как, например, у Д. С. Конькова, который выразился так: «Иордан в этом смысле может утверждать свое происхождение от готов, но себя при этом к готам не причисляет, его высказывание свидетельствует только о наличии готов среди его предков» [Коньков, 2012. С. 71].

В нашем случае доля аланской крови в венах Иордана, если таковая и была, не имеет значения. Важно лишь то, какой язык был для него родным: готский или аланский. Чем же важен этот момент в данном случае? Не чем иным, как языковой способностью Иордана.

Х. Ф. Массманн, активно пользовавшийся готским языком и даже писавший на нем стихи, не избежал критики со стороны лингвистов, отметивших, что одно из сочиненных им готских предложений «продумано по-немецки» («*Sein gotischer Satz ist deutsch gedacht...*» – Его готское предложение продумано по-немецки...) [Peeters, 1980. S. 204]. К. Петерс тонко подметил влияние родного языка на производимые опыты с чужим (особенно мертвым) языком. Сейчас это явление именуется интерференцией, которая чаще всего выявляется на уровне текста.

От вывода Петерса остается сделать только один шаг к популярным в современной лингвистике понятиям менталитета и ментальности, поскольку фраза *deutsch gedacht* явно предполагает ментальные основы текста. Однако в связи со сказанным закономерен вопрос: а если «Гетика» была *gotisch gedacht*, т. е. «продумана по-готски»? Выделимы ли в принципе ментальные основы латинского текста Иордана, унаследованные от готской языковой ментальности?

Термины «ментальность» и «менталитет» иногда употребляются как синонимы, но чаще они дифференцируются, получая у разных авторов не вполне идентичные интерпретации.

Из всех имеющихся определений ментальности наиболее оптимальной нам представляется ее трактовка у В. В. Колесова, по мнению которого ментальность заключает в себе «национальный способ выражения и восприятия мира, общества и человека в формах и категориях родного языка, а также способность истолковывать явления как их сущности» [2006. С. 13]. Способность истолковывать – это именно то, на что следует нацеливаться при комментировании латинского текста Иордана, конкретизируя при этом ментальность именно как текстовую.

Менталитет как понятие по преимуществу рассматривается вне лингвистики. Иногда лингвисты даже заявляют о непричастности лингвистики к данному феномену. Было, например, высказано мнение, что менталитет, в принципе, сфера не лингвистики, не психолингвистики, не когнитивной лингвистики, а социальной и национальной психологии [Попова, Стернин, 2015. С. 17]. Тем не менее, в этой же работе содержится и авторское, уже собственно лингвистическое, понимание менталитета, который представляется как специфический способ восприятия и понимания действительности, определяемый совокупностью когнитивных стереотипов сознания, характерных для определенной личности, социальной или этнической группы людей [Там же. С. 13].

Следует подчеркнуть, что в изучении любых проявлений менталитета, зафиксированного текстами на древних языках, роль лингвистики, тем не менее, весьма значительна, так как все факты, касающиеся менталитета этноса, могут устанавливаться лишь при помощи дошедших до нас текстов, созданных на этих языках. Нельзя не согласиться с мнением Е. М. Масленниковой, отметившей, что главным источником, обеспечивающим «выход» на национальные образы мира, был и остается текст [2008. С. 105]. Однако в целом понятие менталитета действительно следует отнести к более отвлеченным концептуальным сферам, реализующим специфический способ интерпретации действительности, свойственный отдельной личности, социальной или этнической группе людей. В статье акцент ставится на этнической группе,

а именно готской прослойке Остготского королевства VI в. (гот. *Ostrogotans þiudangardi*), и на характерном для нее менталитете, в той или иной форме отразившемся в тексте «Гетики». В этой связи и в этом ракурсе менталитет можно квалифицировать и как текстовый.

С этих позиций далее и рассматривается текст Иордана, с переносом рассуждений о менталитете и ментальности в сферу национально детерминированных моделей и приемов конструирования текста, которые определяют процесс его создания. Именно этот момент составит главный интерес при комментировании труда Иордана с лингвистической точки зрения, а также с позиции того, насколько возможно выделить текстовый менталитет и текстовую ментальность Иордана на базе его латыни и определить, в какой мере они оказываются *gotisch gedacht*.

В этом плане допустимо заявить об особой сфере лингвистической реконструкции, которая традиционно понимается лишь как комплекс приемов и процедур воссоздания незасвидетельствованных языковых состояний и которая всегда ориентировалась лишь на восстановление праязыковых явлений. Однако для современной лингвистической реконструкции характерен также переход от реконструкции отдельных форм и элементов к уровневой реконструкции, что позволяет вплотную подойти и к реконструкции текста (в индоевропеистике – архаичный поэтический язык, язык права, мифологии и т. п.) [Нерознак, 1990. С. 409].

Для готского языка данное обстоятельство имеет особое значение. Дело в том, что единственным источником сведений о готском языке остается Готская Библия, т. е. переведенный в IV в. вестготским епископом Вульфиллой текст Писания с древнегреческого языка. Поскольку перевод Готской Библии представляет собой пословное следование оригиналу, то можно заключить, что о текстовом уровне готского языка, характерном для его естественного функционирования, мы ничего не знаем. Ментальные основы лингвистической реконструкции, обусловленные интерференцией родного языка, простирающейся сквозь латинскую вербализацию, и свидетельствующие, таким образом, о готской языковой ментальности, в состоянии указать на некоторые черты свойственного готскому языку текстопостроения, которые отразились в латинском тексте «Гетики».

Текстопостроительные характеристики «Гетики» как основа ее образности

«Гетика» в аспекте текстопостроения

Еще в конце XIX в. Ф. Вреде установил, что остготский диалект VI в. (как он его поименовал) в фонетическом отношении демонстрирует более молодое состояние, нежели язык Готской Библии («*keinen bedeutend jüngeren und entwickelteren Lautstand als der bibelgotische*») [Wrede, 1891. S. 200]. По всей вероятности, такое заключение будет справедливым и для других уровней остготского языка VI в., в том числе и для текстового. Для целей настоящего исследования целесообразно исходить из признания текста самостоятельным уровнем языка. То, что текст является языковым уровнем, отмечалось в лингвистике неоднократно. У этого положения, однако, больше противников, нежели сторонников. Текст в подавляющем числе случаев продолжают локализовать на уровне речи. Не вдаваясь в подробности дискуссии, отметим только, что реально существующая возможность типизации текстовых структур позволяет предположить, что текст может рассматриваться и как самостоятельный уровень языка.

Исходя из признания текста самостоятельным уровнем языка, можно рассматривать текстовые явления «Гетики», делая акцент на особенностях несохранившегося древнего (как родного) языка, отразившихся в языке, принятом в соответствующем социуме, в данном случае в латинском.

Латынь Иордана изучена исключительно подробно как с точки зрения его стиля, так и с позиции конкретных языковых аспектов. Общим местом всех исследований является самая негативная оценка латинского языка «Гетики», который всем, привыкшим к классической латыни, всегда казался грубым и почти варварским («*uncouth and almost barbarous*»)

[Mierow, 1915. P. 16]. Но иногда ситуацию оценивают более сдержанно, ссылаясь на то, что это неродной язык и что здесь мы имеем перегруженный стиль («un estilo recargado»), содержащий элементы разговорного, народного языка («elementos de la lengua cotidiana y popular») и риторические фразы, которые присущи возвышенному литературному стилю («un estilo mas elevado y literario») [Introducción, 2001. P. 23].

Однако подобные работы не касаются текстового уровня. К тому же Иордан и его труд редко рассматриваются в тесной связи с его литературными современниками («Jordanes and his work are rarely discussed in close relation to his literary contemporaries») [Swain, 2010. P. 247–249], что сильно снижает возможность противопоставления «Гетики» латинской текстовой ментальности того периода. Однако готская специфика «Гетики» у некоторых лингвистов не вызывает сомнения, и, в частности, А. Сёби обоснованно заявляет о наличии в «Гетике» готских элементов, как бы незначительны они ни были («the existence of Gothic elements in the text, however insignificant they may be») и уверенно говорит о наличии у Иордана «собственного голоса» («a voice of his own») [Søby, 2002. P. 484]. Собственный голос Иордана, который обнаруживается некоторыми исследователями в самых разных местах «Гетики», также может свидетельствовать об отражении в ней готской языковой ментальности.

Для многих аспектов комментирования «Гетики» с целью выявления готской части ее ментальности существенно учесть, где она была написана. На этот счет имеются разные точки зрения, но наиболее убедительной представляется позиция Е. Ч. Скржинской, доказавшей, что местом ее создания была Равенна, ставшая с 540 г. византийской [Скржинская, 1960а. С. 60–61]. Но Остготское королевство, как известно, окончательно пало в 555 г., и поэтому готский язык еще звучал там достаточно продолжительное время. Более того, король остготов Теодорих в свое длительное правление (493–526 гг.) готовил почву для того, чтобы сделать готский язык первым официальным языком Остготского королевства («prepara il terreno... all'avvento della lingua gotica come prima lingua ufficiale del suo regno») [Scardigli, 1964. P. 164]. Поэтому закономерно предположить, что Иордан находился в сфере влияния живого готского языка, и все выводы, сделанные в статье на основе изложенных выше теоретических положений, могут непосредственно касаться текстовых явлений остготского языка VI в.

Текстопостроение «Гетики» следует рассмотреть, прежде всего, с довербального уровня, а именно с того момента, когда автор в соответствии со своей текстовой ментальностью начинает отбор признаков номинируемой ситуации для ее последующей вербализации. Поскольку кроме Иордана неизвестен какой-либо другой гот VI в., писавший по-латыни, то все рассматриваемые здесь текстовые явления могут быть оценены только сквозь призму его языковой личности. С другой стороны, отмеченные выше факты довербальной специфики готского языка, характерные лишь для труда Иордана и его языковой личности, подводят к вопросу о том, насколько они были свойственны в целом готской языковой системе того периода и сколько вообще было грамотных готов, следовавших таким же нормам.

Довербальный уровень можно идентифицировать как внутреннюю форму текста и соотносить ее с проблемой языковой номинации, а в данном случае, если ее сузить, – текстовой номинации. Сущность понятия текстовой номинации заключается в том, что уровень текста вербализует внеязыковую ситуацию [Тураева, 1986. С. 22]. Продолжая высказанную З. Я. Тураевой мысль, заметим, что на уровне текста внеязыковая ситуация отображается избирательно на базе какого-то ограниченного и субъективного набора признаков, поскольку отобразить «фотографически» всю внеязыковую ситуацию, ставшую объектом авторского описания, семиотическими средствами языка невозможно. Это под силу другим семиотическим системам – живописи, кино, фотографии, но языковое описание ситуации не может быть таким подробным, иначе текстовое построение потеряет не только всякие границы, но и статус текста. Внутренняя форма текста должна пониматься лишь как понятие, соотносимое с суммой некоторых признаков внеязыковой ситуации, получивших вербализацию. Например, во фрагменте о смерти Атиллы упоминается такая деталь, как девушка с опущенным лицом,

плачущая под покрывалом. Трудно сказать, упоминалась ли соответствующая подробность в хрониках, которыми пользовался Иордан, но, думается, что *плачущую девушку с опущенным лицом под покрывалом* он добавил от себя, продуцируя тем самым ситуацию с признаками почти художественного текста: {254; 255} *puellamque demisso vultu sub velamine lacrimantem*.

Деталь с плачущей под покрывалом девушкой и есть признак ситуации, отображенной Иорданом в соответствии с его собственным видением излагаемого материала, т. е. текстовой ментальностью. Если же эта деталь не почерпнута из хроник, то ее можно признать чертой языковой личности Иордана, тяготеющей к мелким подробностям при вербализации внеязыковой ситуации.

Вспомним в этой связи, что говорил Иордан, создавший свою работу на базе сочинения Кассиодора, о собственных добавлениях в текст по сравнению с его трудом: {3} *addedi... mea dictione permiscens* «добавил... своими высказываниями перемешивая». Что конкретно добавил он при изложении фактов? Возможно, также и частности типа «плачущей под покрывалом девушки».

Другими словами, довербальный уровень текстопостроения в аспекте внутренней формы текста – это то, какое содержание автор имеет в виду вербализовать и какие опорные точки этого содержания (признаки внеязыковой ситуации) он имеет в виду отобразить и облечь в вербальную форму. В данном случае их латинская вербализация не заслоняет довербальной основы текста, и можно полагать, что эта структурная часть текста продиктована готской языковой ментальностью.

Образность «Гетики» как проявление готской языковой ментальности

В этом же ключе можно рассмотреть и специфическую образность Иордана, которая раскрывает особенности его менталитета с точки зрения видения им мира и его комментирования. Образность Иордана наиболее очевидна в сравнениях и метафорах. В целом образы Иордана, которые приводятся ниже, распадаются на несколько тематических групп: соматические метафоры и сравнения, фитообразы, зоообразы, неодушевленные объекты как основа сравнений, некоторые явления природы, абстрактные образы. Если оценивать каждый образ как своего рода признак внеязыковой ситуации, получивший в тексте «Гетики» латинскую вербализацию, и принимать во внимание довербальный уровень, который был «продуман по-готски» (*gotisch gedacht*), то можно полагать, что эта образность была вполне характерна для готского языка VI в., и Иордан как его носитель вполне естественно воплощал ее в латинском тексте.

Некоторые образы имеют тенденцию к повторению (чрево, утроба, рука, тело), что указывает на определенную черту языковой личности Иордана, однако используемая им при этом лексика разнообразна. Так, например, для образа чрева находим:

{46} *Danaper autem ortus grande palude, quasi ex **matre** profunditur* «Днепр великим болотом как бы из материнской утробы рождается»; {17} *unde Vagi fluvius velut quodam **ventrae** generatus in Oceanum undosus evolvitur* «откуда река Ваги, словно чревом порождаемая, в океан волнуясь катится»; {25} *Ex hac igitur Scandza insula quasi officina gentium aut certe velut **vagina** nationum... Gothi... egressi* «из этого острова Скандза как бы из мастерской племен или точнее утробы народов... вышли... готы».

Образ с рукой (*brachium* и *manus* !) также повторяется:

{18} *ex quo quasi quodam **brachio** exiente, sinu distento, Germanicum mare efficitur* «из которого (из океана) как будто какая-то выступающая рука возникает Германское море»; {172} *Africa... a rei publicae Romanae corpus gentilis **manus** abstulerat* «Африка... от тела Римского государства варварской рукой отторгнута».

Образ тела также отмечен дважды:

{261} *dividuntur regna cum populis, fiuntque ex uno **corpore** membra diversa* «отделяются королевства с народами, отпадают от единого тела разные члены»; См. также приведенный выше пример {172}.

Склонность к соматическим образам видна и в следующих примерах:

{53} *Eufratem Tigrimque navigeros ad opinionem maximam perennium fontium cupiosis fundit uberibus* «Евфрат и Тигр судоходный из обильных сосцов вечных источников изливает»; {156} *Adriaeque pelagus velut lingua porrecta a Tyrreno* «от Адриатического моря, словно язык, вытянутая до Тирренского»; {104} *morbis... quod nos ante hos novem annos experti sumus, faciem totius orbis foedavit* «мор, который мы испытали девять лет тому назад, лицо всего круга (т. е. земного) обезобразил»; {75} *in modum spinae, quem costas ut cratem intexunt* «в форме хребтовой кости, ребра которой соединены как бы в плетенке»; {106} *qui vitam solent aliorum dente maledico lacerare* «чужую жизнь злоречивым зубом разрывать».

Фитообразы у Иордана немногочисленны, они включают лист лимона, гриб, цветы:

{16} *insula magna, nomine Scandza, in modum folii cetri* «большой остров в виде листа лимона»; {30} *quae in extremis Asiae finibus ab Oceano eoroboro in modum fungi primum tenuis* «которое (появляется) на крайних границах Азии от северо-восточного океана в виде гриба»; {252} *ad Vandalarum sobulem, quae trino flore pululabat, redeundum est (вернуться)* «к потомству Вандалария, которое распустилось тройным цветком»; {316} *ex eorum latissima prata paucos flores legisse, unde inquirenti pro captu ingenii mei coronam contexta* «с их (т. е. готов) обширнейших лугов немногочисленные цветы собрал, из которых для талантливых умов венков сплел».

Неодушевленные объекты как основа сравнения ограничиваются короной, при этом образ с короной отмечен устойчивостью, он упомянут трижды, и вербализуется с помощью одной и той же лексемы:

{74} *quae patria in conspectu Moesiae sita trans Danubium corona montium cingitur* «эта страна перед Мезией через Данубий опоясывается короной гор»; {6} *pelagi ripa, quam diximus totius mundi circum, in modum coronae ambiens fines suos* «берег, который называем кругом всего мира, в виде короны, огибающий его пределы»; {34} *Dacia... ad coronae speciem arduis Alpibus emunita, iuxta quorum sinistrum latus* «Дакция... с виду короной крутых Альп огражденная».

Зоообразы антропоцентричны: лев (об Аттиле), саранча (об Атаульфе):

{212} *velut leo venabulis praessus speluncae aditus* «словно лев, рогатинами прижатый к входу в пещеру»; {159} *quid primum remanserat, more locustarum erasit* «истребил там подобно саранче, что еще оставалось».

Крайне немногочисленны образы ландшафта и природных стихий (остров, обтекаемый водами (1), волна (2), молния (3), ураган / буря (4)):

(1) {148} *haec... in modum insulae influentium aquarum* «эта [Равенна]... в виде острова, обтекаемого водами»;

(2) {143} *populosque diversarum gentium quasi fonte in uno e diversis partibus scaturiente unda* «и народы различных племен, как бы с различных сторон переливающиеся волны»;

(3) {102} *ibique dum equos exercitumque lassum refoveret, ilico Cniva cum Gothis in modum fulminis ruit* «когда он расположил на отдых лошадей и войско, налетел на него, подобно молнии, Книва с готами»;

(4) {126} *nam mox ingentem illam paludem transierunt, ilico Alpidzuros, Alcildzuros, Itimaros, Tuncarsos et Boiscos, qui ripae istius Scythiae insedebant, quasi quaedam turbo gentium rapuerunt*. «Как вскоре они {126} перешли болото, то алпидзуров, алцилдзуров, итимаров, тункарсов и боисков, которые сидели на побережье этой самой Скифии, как ураган, народов захватили».

Один из абстрактных образов касается вопросов веры, и здесь Иордан достаточно эмоционален. Его темперамент выдает не столько лексема *perfidia* (имеющая ОНЗ, т. е. основное номинативное значение, «вероломство» и уже развившая к тому времени ЛСВ «отпадение от христианской веры»), сколько несогласованное определение к ней (*virus* «яд»), а также характеристика тех, кто поддавался арианскому вероучению: {132, 133} *qui venientes rudibus*

et ignaris ilico perfidiae suae virus infundunt «которые придя, (в сердца или души?) грубых и невежественных тотчас яд своей ереси вливают».

В другом абстрактном образе Иордан, напротив, поэтичен: {250} *Thorismud filius eius flore iuventutis ornatus* «сын Торисмунд, украшенный цветом юности».

Оставляя в стороне заимствованные образы (*малое суденышко, мелкая рыбешка, стоячие воды предков*), взятые Иорданом у Руфина вместе с целым фрагментом его работы [Скржинская, 1960а. С. 31], отметим только зоообраз *пчелиный рой*, с которым сравниваются готы, вышедшие из Скандинавии: {9} *gens... ab huius insulae gremio velut examen apium erumpens* «племня... вырвавшись, словно пчелиный рой, из лона этого острова».

Этот образ, по замечанию Е. Ч. Скржинской [1960б. С. 188], был повторен Иорданом за другими древними историками, и он всегда вспоминается при обсуждении вопроса об истонной родине готов, которой с подачи Иордана традиционно считается Скандинавия. На этот счет существует достаточно много мнений, и не все сходятся в том, что готы пришли на континент из Скандинавии, а точнее с территории современной Швеции. Однако для Иордана, как отмечают исследователи, неразрывно связаны Скандинавия как *origo* (т. е. происхождение) готов, их выход оттуда и образ пчелиного роя («Scandza, its role as the *origo* of the Goths, their permanent departure from there, and the analogy of Goths to bees are all introduced simultaneously») [Gillett, 2006. P. 154–155].

Исходя из довербального уровня текста, который детерминируется как текстовой ментальностью, так и национальным менталитетом в целом, можно в определенной мере оценить практику текстопостроения того периода, сосредоточившись на следующем вопросе. Если бы «Гетика» писалась по-готски, то может ли это означать, что ее образность была бы той же самой, и что вся разница сводилась бы лишь к использованию готских лексем вместо соответствующих латинских? Если ответить на этот вопрос утвердительно, то дальнейшие рассуждения могут принять следующий вид. В целом текстовая образность Иордана имеет достаточно сниженный характер, его сравнения и метафоры ограничены лишь несколькими референтными сферами. Признавая этот их характер следствием языковой личности Иордана, тем не менее можно поставить вопрос и о традициях письменной практики носителей остготского языка в той их части, которые были достаточно образованы или хотя бы грамотны. Можно в этой связи схематично изобразить зафиксированные вербальные репрезентанты проанализированной образности в виде определенной структуры языковой картины мира, характерной для текстового уровня остготского языка того периода, тем более, что многие лексемы зафиксированы в тексте готской Библии, которые ниже приводятся в скобках.

Наиболее развернутый блок этой структуры – тело человека, упомянутое как тело (гот. *leik*), чрево (гот. *wamba, qībus, kilpei*), рука (гот. *arms, handus*), сосцы, язык (гот. *tuggo*), лицо / лик (гот. *ludja, andawleizn, andaugi, wlits, andwairpi*), хребет, зуб (гот. *tunbus*).

Что касается живой природы, то она отражена слабо: лимон, гриб, цветок (гот. *bloma*). Здесь можно заметить, что лимонный лист – это южная деталь, она не могла быть известна готам Балтики или менее южных областей их пребывания. В этом случае можно сказать, что, по-видимому, лексико-семантическая система остготского языка пополнилась лексемой *лимон*.

Остальные блоки такой структуры вербально также малочисленны: неодушевленные объекты – корона; водные массивы – вода (гот. *wato, ahwa*), волна (гот. *wegs*); явления природы – молния (гот. *lauhmuni*), ураган (гот. *skura*); абстрактные явления – вера (гот. *galaubeins*), цвет (юности).

Может возникнуть закономерный вопрос о вероятном заимствовании Иорданом текстовой образности вследствие его знакомства с позднелатинскими сочинениями, особенно при работе над «Гетикой». Если в этой связи сравнить текст Иордана с позднеримскими текстами и, в частности, с работой Павла Орозия «История против язычников», которую Иордан хорошо знал и которую неоднократно цитировал в «Гетике», то следует подчеркнуть, что

в плане образности труд Орозия абсолютно нейтрален, и единственный заметный образ представлен у него следующей метафорой:

(Книга 1. Глава 12. 1.) *At ego nunc cogor fateri, me prospiciendi finis commodo de tanta malorum saeculi circumstantia praeterire plurima, cuncta breuiare. nequaquam enim tam densam aliquando siluam praetergredi possem, nisi etiam crebris interdum saltibus subuolare.* «Но я теперь принужден, чтобы вовремя прилизиться к концу своего рассказа, многие обстоятельства того века, наполненного несчастьями, сократить. Поскольку никогда бы не смог я пройти такой густой лес, если бы часто не перепрыгивал»¹.

Упомянутый факт может стать свидетельством того, что Иордан не следовал принятым образцам текстопостроения, характерным для оригинальных латиноязычных трудов своей эпохи, но шел своим путем и, возможно, был под влиянием интерференции со стороны родного (готского) языка и характерных для него менталитета и ментальности.

Заключение

Предпринятый на базе лингвистической реконструкции анализ позволил выявить некоторые языковые особенности, которые могут расцениваться как текстовые явления, присущие остготскому языку VI в. Исследование исходило из того факта, что наше знание готского языка, ограниченное ресурсами дословного перевода готской Библии, не распространяется на текстовые явления готского языка, свойственные ему в его естественном бытовании. В силу того, что современная лингвистическая реконструкция распространяется и на восстановление текстовых явлений, представилось обоснованным восстановить некоторые из них, просматривающиеся сквозь латинскую вербализацию «Гетики» как проявление готской языковой ментальности. Восстановленные текстовые явления раскрывают характер образности, зафиксированной в метафорах и сравнениях «Гетики», которые отмечены опорой на немногочисленные сферы внешней действительности и отражают ее образное представление. В заключение сформулируем одно из возможных направлений в исследовании готского языка, вполне допустимое в силу разработанности целого ряда положений современной лингвистики. Когнитивные исследования, одним из аспектов которых является изучение ментальности и менталитета во всех их проявлениях, имеют основание развиваться на базе готского языка и принести свои плоды, что нельзя признать излишним, учитывая недостаточность дошедшего до нас готского материала.

Список литературы / References

- Иордан.** О происхождении и деяниях гетов. Getica. М.: Вост. лит., 1960. 436 с.
Jordan. Origin and deeds of the Getae. Getica. Moscow, Izd-vo Vostochnoy literatury, 1960, 436 p. (in Russ.)
- Коньков Д. С.** «Гетика» Иордана – готское историческое предание или конъюнктура эпохи: современное состояние изучения проблемы // Вестн. Том. гос. ун-та. 2012. № 361. С. 70–73.
Konkov, D. S. Jordanes' "GETICA" – a Gothic historical legend or the conjuncture of the epoch: contemporary state of investigating the problem]. *Bulletin of Tomsk State University*, 2012, no. 361, p. 70–73. (in Russ.)
- Колесов В. В.** Русская ментальность в языке и тексте. СПб.: Петербургское востоковедение, 2006. 624 с.
Kolesov, V. V. Russian mentality in language and speech. St. Petersburg, Peterburgskoye Vostokovedeniye, 2006, 624 p. (in Russ.)
- Нерознак В. П.** Реконструкция // Лингвистический энциклопедический словарь. М.: Сов. энциклопедия, 1990. С. 409–410.
Neroznak, V. P. Reconstruction. In: *Linguistic encyclopaedia*. Moscow, Sovetskaya entsiklopedia, 1990, p. 409–410. (in Russ.)

¹ Pauli Orosii Historiae adversum paganos. URL: <http://www.attalus.org/latin/orosius1.html>

- Масленникова Е. М.** Лакунарные разрывы в диалоге культур: к пониманию сущности // Слово и текст: психолингвистический подход: Сб. науч. тр. / Под общ. ред. А. А. Залевской. Тверь: Твер. гос. ун-т, 2008. Вып. 8. С. 100–107.
- Maslennikova, E. M.** Lacunar gaps in the dialogue of culture: to the understanding of the main point]. In: *Word and text: psycholinguistic approach*. Tver, Tver State University, 2008, no. 8, p. 100–107. (in Russ.)
- Попова З. Д., Стернин И. А.** Язык и национальная картина мира. 4-е изд., стереотип. Москва; Берлин: Директ-Медиа, 2015. 101 с.
- Popova, Z. D., Sternin I. A.** Language and national world picture. Moscow, Berlin, Direkt-media, 2015, 101 p. (in Russ.)
- Скржинская Е. Ч.** Иордан и его «Getica» // Иордан. О происхождении и деяниях гетов. Getica. М.: Вост. лит., 1960а. С. 11–61.
- Skrzynska, E. Ch.** Jordanes and his «Getica». In: Jordan. Origin and deeds of the Getae. Getica. Moscow, Vostochnaya literatura, 1960, p. 11–61. (in Russ.)
- Скржинская Е. Ч.** Комментарий // Иордан. О происхождении и деяниях гетов. Getica. М.: Вост. лит., 1960б. С. 185–364.
- Skrzynska, E. Ch.** Commentary [Commentary]. In: Jordan. Origin and deeds of the Getae. Getica. Moscow, Vostochnaya literatura, 1960, p. 185–364. (in Russ.)
- Тураева З. Я.** Лингвистика текста (Текст: структура и семантика): Учеб. пособие. М.: Просвещение, 1986. 127 с.
- Turayeva, Z. Ya.** Text linguistics (Text: structure and semantics). Moscow, Prosveshcheniye, 1986, 127 p. (in Russ.)
- Introducción. In: Jordanes. Origen y gestas de los godos. Madrid, Juan Ignacio Luca de Tena, 2001, p. 4–35.
- Gillett, A.** The Goths and the Bees in Jordanes: A Narrative of no Return. Byzantine narrative. Papers in Honour of Roger Scott. Melbourne, The University Printing Service, 2006, p. 149–163.
- Mierow, Ch. Ch.** Introduction. In: The Gothic history of Jordanes. Princeton, Princeton University Press; London, Humphrey Milford; Oxford University Press, 1915, p. 1–50.
- Peeters, Ch.** Gotisches. *Zeitschrift für vergleichende Sprachforschung*, 1980, Bd. 94, H. 1/2, S. 203–208.
- Scardigli, P. G.** Lingua e storia dei Goti. G. C. Sansoni (ed.). Firenze, 1964, 398 p.
- Swain, B.** Jordanes and Virgil: A case study of intertextuality in the GETICA. *The Classical Quarterly. New Series*, 2010, vol. 60, no. 1, p. 243–249.
- Søby, A.** Christiansen: Cassiodorus, Jordanes and the History of the Goths. In: Studies in a Migration Myth. København, Museum Tusculanum Press, 2002, 391 p.
- Wrede, F.** Über die Sprache der Ostgoten in Italien. Strassburg, Karl J. Trübner, 1891, VII + 208 S.

Материал поступил в редколлегию
Date of submission
16.11.2018

Сведения об авторе / Information about the Author

Трофимова Юлия Михайловна – доктор филологических наук, профессор, профессор кафедры английской филологии Национального исследовательского Мордовского государственного университета им. Н. П. Огарева (ул. Большевикская, 68/1, Саранск, 430005, Россия)

Yulia M. Trofimova, Doctor of Philology, Professor, Professor of English Philology Chair, Ogarev Mordovia State University (68/1 Bolshevistskaya Str., Saransk, 430005, Russian Federation)

primavera49@list.ru

Социокультурный статус формулы извинения *pardon* в английском языке по данным корпусных баз

И. В. Соловьева

*Национальный исследовательский университет «Высшая школа экономики»
Москва, Россия*

Аннотация

Статья посвящена возможностям использования лингвистического корпуса английского языка для формирования межкультурной компетенции. Успешность обучения межкультурной коммуникации зависит как от уровня владения иностранным языком, так и от освоения норм речевого этикета. Сведения об этих нормах преподаватели, не являющиеся носителями языка, могут почерпнуть из учебно-методических комплексов, словарей, справочников, научных исследований. Однако несовпадение рекомендаций поднимает вопрос о поиске более достоверных источников информации о стилистической и социальной маркированности языковых единиц. Лингвистический корпус текстов, будучи репрезентативным, унифицированным и структурированным массивом данных, лучше показывает социокультурную маркированность единиц, так как он отражает реальный узор, а не мнения носителей. В исследовании использован Британский национальный корпус (British National Corpus), представляющий британский вариант английского языка конца XX в. На примере формулы извинения с *pardon* исследуется возможность применять лингвистический корпус как источник сведений о социально-классовой маркированности дискурса. Поскольку перед исследователем стояла задача подтвердить или опровергнуть убеждение, что формула извинения с *pardon* присуща речи низшего и среднего слоя среднего класса, было проведено сравнение частотности ее употребления представителями разных слоев британского общества. По данным нашего корпусного исследования, формула извинения с *pardon* в большей степени присуща речи представителей высших слоев общества, которые пользуются этой формулой в два с лишним раза чаще, чем представители других социальных групп: 303,47 против 139,54 употреблений на миллион. Кроме того, очевидна тенденция к повышению частотности этой формулы извинения в каждом следующем социальном классе, начиная с низшего. Несмотря на ограничения, вызванные полисемией и неполной разметкой материалов, можно говорить о целесообразности применения корпуса как инструмента обучения межкультурной коммуникации.

Ключевые слова

социолингвистическая компетенция, речевой этикет, корпусная лингвистика, социальный статус

Благодарности

Автор выражает благодарность участникам секции «Межпредметные подходы в исследовании межкультурной коммуникации» и руководителю секции канд. филол. наук, проф. И. И. Чироновой за плодотворное обсуждение, а также ценные замечания по сути данного исследования и его перспектив на IV Международной научно-практической конференции «Стратегии межкультурной коммуникации в современном мире: язык, образование, культура», организованной НИУ ВШЭ.

Для цитирования

Соловьева И. В. Социокультурный статус формулы извинения *pardon* в английском языке по данным корпусных баз // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 125–133. DOI 10.25205/1818-7935-2019-17-1-125-133

Language Corpora Data as the Source of Sociocultural Information about *Pardon* Apology

Inna V. Soloveva

*National Research University Higher School of Economics
Moscow, Russian Federation*

Abstract

The article covers the potentials of the English language corpus in terms of developing intercultural awareness, which is the topical issue nowadays, discussed in various domains of humanities. Successful intercultural communication is

contributed by both mastery of language and mastery of speech etiquette. For the speech etiquette standards, non-native speakers teaching foreign languages may refer to the teaching support kits, dictionaries, reference textbooks, and academic research. Communicating with native speakers and reading general fiction may also be of use. Nevertheless, contradictory recommendations raise concern about reliability of information sources on stylistic and cultural marking of various linguistic units. Reflecting the actual usage of linguistic units instead of biased native speakers' opinions and being a representational, unified, and well-structured body of data, a language corpus may be used as a reliable information source for the purposes of teaching a foreign language. For the purposes of this study, British National Corpus was chosen as it represents the British English of the late 20th century. *Pardon* apology formula was used as an example to study the applicability of language corpus data to identify the social class marking of discourse. The research had for a goal validating or disproving the opinion that *pardon* apology formula is inherent to the lower-middle and upper-middle class, therefore its frequency was put to analysis. As result, the language corpus shows that *pardon* apology formula is used by the upper social classes almost twice as frequent as by the other classes (303.47 cases vs 139.54 cases per million). Besides, there is a distinct trend showing that this apology formula is used more frequently by each next social class. Though there are some constraints imposed by multiple meaning and insufficient marking of materials, using the language corpus as a tool of teaching intercultural communication is deemed to be quite practicable.

Keywords

sociolinguistic competence, speech etiquette, corpus linguistics, social status

Acknowledgements

The author would like to show her gratitude to the organizing committee of 4th International Conference *Cross-Cultural Communication Strategies on the Modern World: Language, Education and Culture* (Moscow, Higher School of Economics), the participants of the section *Interdisciplinary Approaches in the Study of Cross-cultural Communication* and the chair prof. I. I. Chironova for the productive discussion of this research and essential feedback on its prospective.

For citation

Soloveva, Inna V. Language Corpora Data as the Source of Sociocultural Information about *Pardon* Apology. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 125–133. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-125-133

Введение

Интерес к вопросам межкультурной коммуникации выражается в росте количества лингвистических исследований, посвященных анализу дискурсивно-прагматических особенностей функционирования формул общения. Находясь на стыке дисциплин, обучение эффективной коммуникации с представителями иноязычных культур не может ограничиваться только рассмотрением языковых формул как средств реализации речевого акта. Успешность обучения межкультурной коммуникации зависит как от уровня владения иностранным языком, так и от освоения норм речевого этикета, т. е. «системы коммуникативных стереотипов, устойчивых выражений, служащих для установления, поддержания и размыкания контакта общающихся» [Формановская, Габдуллина, 2008. С. 4]. Говорящие по-английски инофоны смогут избежать коммуникативных неудач, вызванных различием в специфике национального коммуникативного поведения, благодаря освоению норм английского речевого этикета, что позволит сделать межнациональную коммуникацию на английском языке более эффективной.

Постановка проблемы

Источниками сведений о нормах речевого этикета для учащихся и преподавателей, не являющихся носителями языка, служат учебно-методические комплексы, словари и справочники. Дополнительную информацию о нормах речевого этикета можно получить из специализированных исследований по межкультурной коммуникации, практики живого общения с носителями языка и художественных произведений. В то же время несовпадение рекомендаций становится препятствием на пути изучающих иностранный язык и поднимает вопрос о достоверных источниках информации по стилистической и социальной маркированности языковых единиц. Термин «социальная маркированность» здесь и далее употребляется в со-

ответствии с определением, данным Л. П. Крысиным, а именно как свойство определенных языковых единиц «приобретать функции социальных символов – маркеров принадлежности говорящего к той или иной социальной среде» [2000. С. 26].

Формулы извинения широко используются в естественной коммуникации, являясь средством реализации важной социальной функции, обеспечивающей поддержание гармонии межличностных отношений, в связи с чем они часто оказываются в поле внимания исследователей. Данная статья посвящена анализу употребления формулы *pardon* в английском языке с точки зрения социальной маркированности на материале исследований по английскому речевому этикету в сопоставлении с данными корпусной лингвистики.

Актуальность и новизна исследования

Являются ли формула извинения *pardon* и сочетания типа *to beg one's pardon* социально маркированными? Как коррелируют рекомендации из справочников, монографий и пособий по речевому этикету с информацией лингвистического корпуса? Ответы на эти вопросы будут способствовать развитию методики преподавания речевого этикета, что актуально, во-первых, в связи со стабильно растущим интересом к данному разделу межкультурной коммуникации, во-вторых, с увеличением количества преподавателей и авторов учебников, не являющихся носителями английского языка¹.

Новизна исследования обусловлена применением информации корпуса английского языка в практике обучения межкультурной коммуникации, в частности в обучении английскому речевому этикету. В отличие от предыдущих исследований по вопросам межкультурной коммуникации данная работа основывается не на анализе художественных текстов, ограниченных по охвату реальных ситуаций общения, а на актуальном узусе, зафиксированном в корпусе. Кроме того, исследования лингвистов, базирующихся на лингвистических корпусах, еще не затрагивали вопросы обучения межкультурной коммуникации, что также свидетельствует в пользу новизны данной работы. Таким образом, новизна исследования обусловлена тем, что оно применяет корпусную лингвистику как инструмент для решения задач обучения межкультурной коммуникации.

Цели и задачи исследования

Цель исследования – установить потенциал применения данных корпусной лингвистики для обучения межкультурной коммуникации, в частности охарактеризовать социальную маркированность *pardon* как формулы извинения. Цель исследования определяет задачи: изучить данные корпуса современного английского языка для формулы извинения с *pardon*, определить возможность оценить социальную маркированность языковых единиц и ответить на вопрос о том, является ли *pardon* социально маркированным и, если да, какие слои общества его более склонны употреблять.

Обзор литературы

Импульсом для проведения исследования стала практика обучения английскому речевому этикету взрослых, в ходе которой было обнаружено противоречие между учебно-методическими комплексами и научной, научно-популярной и справочной литературой по межкультурной коммуникации, в частности в рекомендациях по употреблению формул извинения. Одни источники рекомендуют *pardon* как способ принести извинения, в то время как другие

¹ По данным ресурса <http://languageknowledge.eu/>, представляющего результаты исследования *Europeans and their languages* (http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf), 33 % населения Европы говорят по-английски, для 12,91 % из которых английский язык является родным, а для 20,09 % – иностранным. При этом английский язык занимает лидирующую позицию среди прочих иностранных языков: французский как иностранный знают 7,62 % граждан Евросоюза, немецкий – 6,69 %.

его либо игнорируют, либо указывают на социальную маркированность. Далее будут приведены сведения о формуле извинения с *pardon* в учебно-методических комплексах и работах по речевому этикету и межкультурной коммуникации с опорой как на данные языковых корпусов, так и на примеры из художественных произведений.

*Справочники по речевому этикету и учебно-методические комплексы
как источники сведений о социальной маркированности*

Основным источником сведений о формулах вежливости являются справочники по речевому этикету, где *pardon* предлагается в основном для функции переспроса. В справочнике «Русский и английский речевой этикет: сходства и различия», адресованном изучающим английский язык и преподавателям, *pardon* приводится как синонимичный *excuse me* и *I am sorry*. «Вежливая форма извинения без объяснения того, за что просят прощения, возможна и при малом поступке: Прошу прощения! – *I beg your pardon*. Это выражение употребляют и в тех случаях, когда недопонимают сказанное или прерывают говорящего» [Формановская, Габдуллина, 2008. С. 64]. Для последней функции – переспроса – рекомендует формулу с *pardon* справочник «Английский речевой этикет», характеризуя сферу применения *I beg your pardon* как формальную, а *pardon* – как социально маркированную: «This is often used in Britain but tends to be avoided by educated people, who generally prefer **Sorry**» [Шахлай, 2011. С. 32].

В отличие от справочников аутентичные учебники по английскому языку формулу с *pardon* не предлагают. В двух популярных учебно-методических комплексах издательств *Market Leader* by Pearson Education Limited и *English File* by Oxford University Press в тематических разделах, посвященных отработке функционального языка, речевые формулы с *pardon* не приводятся, в том числе в тех ситуациях, когда говорящий переспрашивает или перебивает собеседника, что противоречит приведенной выше информации из справочников.

*Формула извинения pardon
в научных исследованиях и монографиях*

Как монографии, адресованные широкому кругу читателей, так и научные исследования указывают на сниженную социальную маркированность формулы извинения *pardon*. Т. В. Ларина в исследовании «Англичане и русские: язык, культура, коммуникация» пишет: «При переспросе возможны также *Sorry* и *What*, причем последняя формула, столь близкая русскому человеку, употребляется как представителями рабочего класса, так и высшего общества, в то время как *Pardon* предпочитают говорить те, кто принадлежит к низшему и среднему слоям среднего класса» [2013. С. 310]. Подобная точка зрения высказывается в диссертации Е. А. Плетневой, где она утверждает, что «данная формула является ярким языковым маркером принадлежности к низшей прослойке среднего класса британского общества (lower-middle-class), для именованной которой сегодня даже существует особый, уничижительный, на наш взгляд, термин – «Пардония» (*'Pardonia'*)» [2009. С. 17]. Авторы, высказывающие данную точку зрения о классовой маркированности формулы извинения с *pardon*, ссылаются на британского антрополога Кейт Фокс. Не являясь лингвистом, К. Фокс в социально-антропологическом исследовании «*Watching the English: The Hidden Rules of Behaviour*», тем не менее, пишет о речевых социально-классовых маркерах, а формула извинения *pardon* имеет, по ее мнению, четкую маркированность. «This word is the most notorious pet hate of upper and upper-middle classes [...] To the uppers and upper-middles, using such [*pardon*] an unmistakably lower-class term is worse than swearing. Some even refer to lower-middle-class suburbs as *'Pardonia'*. Here is a good class test you can try: when talking to an English person, deliberately say something too quietly for them to hear you properly. A lower-middle or middle-middle person will say, *'Pardon?'*; an upper-middle will say *'Sorry?'* (or perhaps *'Sorry – what?'* or

‘What – sorry?’); but an upper-class or a working-class person will both say, ‘What?’» [Fox, 2014. P. 105–106]².

В то же время другие источники утверждают о смене статуса слова-маркера *pardon*, которое утратило ассоциации с низким социальным уровнем говорящего. Так, в статье Т. А. Ивушкиной, посвященной анализу слов – социальных маркеров на материале современной британской литературы, говорится, что «фраза *I beg your pardon* стала активно использоваться и в высших слоях общества» [2016. С. 50]. Упомянутые исследования основываются на анализе дискурсивно-прагматических особенностей функционирования формул с семантикой извинения в британской культуре на материале художественной литературы, публицистики, фильмов, телесериалов, реальных ситуаций общения. Русскоязычные авторы [Ивушкина, 2016; Ларина, 2013; Плетнева, 2009] поднимают вопрос о сложностях, связанных с подбором материала и определением социально-классовой маркированности языковых единиц.

*Специфика лингвистического корпуса как источника информации
о социальной маркированности языковых единиц*

Одним из источников такой информации могут служить данные корпусной лингвистики, раздела компьютерной лингвистики, «занимающегося разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с использованием компьютерных технологий» [Захаров, 2005. С. 3]. Лингвистический корпус текстов – это репрезентативный, унифицированный и структурированный массив языковых данных, который предоставляет пользователю возможность поиска данных и получения статистической информации. Для английского языка наиболее известны Британский национальный корпус (British National Corpus), Международный корпус английского языка (International Corpus of English), лингвистический Банк английского языка (Bank of English). Сведения из корпусов могут быть достоверным и ценным источником при обучении иностранному языку, так как отражают реальный узус. «Traditional grammars and dictionaries tell us what a word *ought to mean*, but only experience can tell us what a word *is used to mean*. This is why dictionary publishers, grammar writers, language teachers, and developers of natural language processing software alike have been turning to corpus evidence as a means of extending and organizing that experience»³. Лингвистические исследования в области прагматики как в России, так и за рубежом начинают использовать данные лингвистических корпусов. Функция извинения была изучена в англоязычной и русской научной литературе на материале корпусов английского языка [Aijmer, 2014; Скворцова, 2016], однако без фокуса на социально-классовой маркированности единиц.

Обзор литературы позволяет говорить об актуальности исследования социокультурного компонента в связи с вниманием со стороны как отечественных, так и зарубежных ученых. Одни лингвисты рассуждают о социальной маркированности языковых формул с опорой на ограниченный материал, в то время как другие анализируют данные корпусов для представления о закономерностях применения речевых клише, но безотносительно их социальной маркированности. Таким образом, исследование с опорой на дифференцированные в социалингвистическом отношении данные корпуса современного английского языка может

² «У аристократов и у представителей верхушки среднего класса это слово особенно не в чести. [...] По мнению представителей высшего класса и верхушки среднего класса, это явно простонародное словечко хуже бранного выражения. Некоторые даже называют пригороды, в которых обитают представители низов среднего класса, Пардонией. Есть хороший тест на определение классовой принадлежности: беседа с англичанином, умышленно скажите что-нибудь очень тихо, так чтобы вас не расслышали. Выходец из низов или средней части среднего класса переспросит: «Pardon?» – представитель верхушки среднего класса скажет «Sorry?» («Прошу прощения?») или «Sorry – what?» («Простите, что вы сказали?»), а вот человек из высшего общества и рабочий, те оба спросят: «What?» (цит. по: Кейт Фокс «Наблюдая за англичанами. Скрытые правила поведения», Рипол Классик, 2008, переводчик И. Новоселецкая).

³ British National Corpus. URL: <http://www.natcorp.ox.ac.uk/using/index.xml> (дата обращения 01.06.2018).

ответить на вопрос о способности формулы извинения *pardon* быть словом-индексом, сигнализирующем о социально-классовой принадлежности говорящего, а следовательно, позволит решить практические задачи обучения английскому речевому этикету.

Результаты и обсуждение

Исследование отталкивается от гипотезы, предполагающей, что разметка лингвистических корпусов современного английского языка дает возможность определить социальную принадлежность информантов, которые применяют формулу извинения с *pardon* и ее производные. Для исследования был использован наиболее представительный Британский национальный корпус (British National Corpus, BNC), включающий 100 миллионов образцов письменной (90 %) и устной (10 %) речи из разнообразных источников, репрезентирующих британский вариант английского языка конца XX в.

Определение частотности

в зависимости от социально-классовой принадлежности информантов

При запросе по общей базе по слову *pardon*⁴, было получено 1 802 результата в 574 различных текстах из 98 313 429 слов и 4 048 текстов; таким образом, частотность слова *pardon* составляет 18,33 употребления на миллион. Однако для целей данного исследования подходит только один из подтипов поиска *Restricted Range of Spoken Texts and Speakers*, поскольку только он размечен по социально-классовой принадлежности (*Social Class of Respondent*) информантов в соответствии с *NRS social grades*, принятой в Великобритании⁵. Выделяются следующие классы респондентов:

AB – верхний сегмент среднего класса (руководители высшего звена, высококвалифицированные специалисты), средний класс (руководители среднего звена, квалифицированные специалисты);

C1 – низший уровень среднего класса (конторские служащие, управленцы низшего звена);

C2 – квалифицированные рабочие;

DE – рабочий класс (низко- и неквалифицированные рабочие), неработающее население (временно занятые работники, пенсионеры и те, кто живет за счет государственных пособий)⁶.

Результаты

Поскольку перед нами стояла задача подтвердить или опровергнуть мнение о том, что формула извинения-переспроса с *pardon* присуща речи низшего и среднего слоя среднего класса, было проведено сравнение частотности данного слова по этой характеристике, результаты которого таковы (каждому классу соответствует частотность употребления на миллион слов): AB – 303,47; C1 – 235,22; C2 – 158,36; DE – 139,54; Unknown – 88,36. По данным корпуса, формула извинения с *pardon* в большей степени присуща речи верхних слоев общества – с частотностью 2,17 раза. Кроме того, заметна тенденция к повышению частотности этой формулы извинения в каждом следующем социальном классе.

⁴ <http://bncweb.lancs.ac.uk/cgi-bin/bncXML/BNCquery.pl?theQuery=search&urlTest=yes> (дата обращения 04.07.2018).

⁵ NRS social grades - система демографической классификации, принятая в Великобритании. Изначально разработанная National Readership Survey для классификации читателей в середине XX века, она получила широкое распространение в других исследованиях.

⁶ <http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/> (дата обращения 06.07.2018).

*Ограничения применения корпуса
как источника о социально-классовой маркированности*

Следует принимать во внимание ограниченность корпуса как источника данных, связанных с неполной разметкой, которая отражается в категории «unknown». Тем не менее, даже с учетом «серой» зоны, включающей не размеченный по родному языку (см. таблицу) и классовой принадлежности информантов, очевидно, что *pardon* скорее свойственно речи носителей британского варианта английского языка, относящихся к верхнему сегменту общества.

Частотность употребления *pardon* в устной речи с учетом родного языка информантов
(количество употреблений на один миллион, *ipm*)

Frequency of *pardon* apology in oral speech regarding informants' native language (*ipm*)

Родной язык	Класс				
	AB	C1	C2	DE	Unknown
British English	113,47	—	—	47,79	70,84
North American English	—	—	—	—	77,83
Unknown Indian language	—	—	—	—	—
German	—	—	—	—	—
French	—	—	—	—	—
Unknown	326,44	235,22	158,36	144,00	99,70

Еще одно ограничение на применение корпуса как инструмента связано с выделением разных значений одного и того же слова. В связи с этим примеры на поисковый запрос по слову *pardon* в устной речи без ограничений по родному языку или классовой принадлежности были просмотрены индивидуально, в результате чего отсортировано три из 1 277 со значением *pardon* как «помилование, помиловать» (например, в проповеди: *And to our God for he will abundantly pardon*).

Заключение

Несмотря на указанные ограничения, данные Британского национального корпуса свидетельствуют о социально-классовой маркированности формулы извинения с *pardon*, при этом не как присущего речи малообразованных низших слоев населения, по утверждению большинства исследователей, а, напротив, как более часто встречающихся в устной речи представителей верхних сегментов среднего класса.

Что касается применения результатов данного исследования, то на примере формулы извинения с *pardon* было доказано, что данные корпусной лингвистики как инструмент могут быть полезны при обучении межкультурной коммуникации в качестве источника информации о социально-классовой маркированности отдельных этикетных формул. Данные Британского национального корпуса свидетельствуют в пользу уместности формулы извинения с *pardon* в речи образованного человека с достаточно высоким социальным статусом, что способствует решению одной из задач обучения межкультурной коммуникации – формированию социолингвистической компетенции. Знания и умения, необходимые для эффективного использования языка в социальном контексте, выражаются в том числе и в способности выражать свою социальную принадлежность через выбор слов, произношение, грамматиче-

скую правильность речи⁷. Речь выражает и определяет социальную идентичность, при этом в силу природы освоения иностранного языка учащиеся рискуют сформировать такую языковую личность, которая не будет адекватно отражать их классовую принадлежность и социальный статус. Впечатление других говорящих на этом иностранном языке, в том числе и носителей, может кардинально не совпадать с намерением говорящего и приводить к коммуникативным неудачам. При этом надеяться, что, так же как и при освоении родного языка, учащийся будет копировать речевое окружение, не приходится, поскольку на этапе обучения его иноязычное языковое окружение существенно ограничено. Учебники и справочники транслируют нейтральную социальную маркированность речи, данные о словах-индексах исследователи получают из художественной литературы или наблюдения за речевым поведением носителей языка, что не может быть надежным источником в силу ограниченной выборки материала. Данные корпусной лингвистики представляются более ресурсными по масштабу и возможностям охвата для задач формирования социолингвистической компетенции при обучении межкультурной коммуникации.

Список литературы / References

- Захаров В. П.** Корпусная лингвистика: Учеб.-метод. пособие. СПб., 2005. 48 с.
Zakharov, V. P. Corpus Linguistics: Training Manual. St. Petersburg, 2005, 48 p. (in Russ.)
- Ивушкина Т. А.** U / non-U words в зеркале современной английской художественной литературы // Научный диалог. 2016. № 6 (54). С. 42–53.
Ivushkina T. A. U / non-U Words as Mirrored in Modern English Literature. *Nauchnyj dialog*, 2016, no. 6 (54), p. 42–53. (in Russ.)
- Карасик В. И.** Язык социального статуса. М.: Гнозис, 2002. 333 с.
Karasik, V. I. Language of Social Status. Moscow, Gnozis Publ., 2002, 333 p. (in Russ.)
- Крысин Л. П.** Социальная маркированность языковых единиц // Вопросы языкознания. 2000. № 4. С. 26–42.
Krysin, L. P. Social Marking of Language Units. *Voprosy yazykoznaniiya*, 2000, no. 4, p. 26–42. (in Russ.)
- Ларина Т. В.** Англичане и русские: Язык, культура, коммуникация. М.: Языки славянских культур, 2013. 360 с.
Larina, T. V. The Russians and the English: Language, Culture, Communication. Moscow, Yazyki slavyanskikh kul'tur, 2013, 360 p. (in Russ.)
- Плетнева Е. А.** Коммуникативно-прагматический диапазон формул извинения в британской культуре: Автореф. дис. ... канд. филол. наук. Воронеж, 2009. 20 с.
Pletneva, E. A. Communicative and Pragmatic Range of Apology in British Culture. Dissertation abstract. Voronezh, 2009. 20 p. (in Russ.)
- Скворцова Е. Б.** Извинение как первичная функция периферийных формул извинения apologise/ze, forgive me, pardon при исследовании их комбинаторных характеристик // Филологические науки. Вопросы теории и практики. 2016. № 9 (63), ч. 2. С. 154–158.
Skvortsova, E. B. Apology as the Primary Function of Peripheral Forms of Apologizing Apologise/ze, Forgive me, Pardon while Studying their Combinatory Characteristics. *Filologicheskie nauki. Voprosy teorii i praktiki*, 2016, no. 9 (63), part 2, p. 154–158. (in Russ.)
- Формановская Н. И., Габдуллина С. Р.** Русский и английский речевой этикет: сходства и различия: Справочник. М.: Высш. шк., 2008. 111 с.
Formanovskaya, N. I., Gabdullina, S. R. Russian and English Speech Etiquette: Similarities and Differences. Handbook. Moscow, Vysshaya shkola Publ., 2008, 111 p. (in Russ.)

⁷ Вопросы связи языка и социальной принадлежности его носителя оказались в фокусе внимания современных исследователей на рубеже веков, с нарастающими темпами глобализации [Blot, 2003, Coupland, Williams 2003]. «That we judge another's character by reference to language use is not a novel discovery. Language is inescapably a badge of identity... Whenever we open our mouths to speak we provide those who hear us, chosen interlocutors and mere bystanders alike, with a wealth of data, a congeries of linguistic clues others use to position us within a specific social stratum» [Blot, 2003. P. 3].

- Шахлай В. С.** Английский речевой этикет. Минск: Белорус, 2011. 160 с.
Shahlej, V. S. English Speech Etiquette. Minsk, Belarus, 2011, 160 p. (in Russ.)
- Aijmer, K.** Conversational routines in English: convention and creativity. Routledge, New York, 2014, 268 p.
- Blot, R. K.** Language and Social Identity. Praeger, Westport CT, 2003, 292 p.
- Coupland, N., Williams, A.** Investigating Language Attitudes: Social Meaning of Dialect, Ethnicity and Performance. Cardiff, University of Wales Press, 2003, 253 p.
- Fox, K.** Watching the English: The Hidden Rules of Behaviour. London, Hodder & Stoughton, 2014, 583 p.

Материал поступил в редколлегию

Date of submission

08.09.2018

Сведения об авторе / Information about the Author

Соловьева Инна Владимировна, кандидат филологических наук, доцент кафедры английского языка для гуманитарных дисциплин департамента иностранных языков Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

Inna V. Soloveva, Candidate of Philological Sciences, Associate Professor, School of Foreign Languages, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)

isolovyova@hse.ru

ORCID 0000-0003-4357-4664

Лингвистическая дискриминация и лингвистическая апроприация как примеры феномена двойственного сознания в США

А. Н. Себрюк

*Национальный исследовательский университет «Высшая школа экономики»
Москва, Россия*

Аннотация

Несмотря на отмену дискриминационного законодательства в США и последующие меры общественного и правового воздействия, расизм до сих пор остается одной из самых острых проблем американского общества. Современный расизм зачастую выражается в завуалированных и трудно различимых формах (*covert, subtle racism*), которые включают в себя проявление излишней вежливости или специфическое употребление эвфемизмов. Являясь средством выражения основных идеологических позиций, язык также служит мощным инструментом воздействия на общественное сознание. В статье раскрываются способы воспроизводить и укоренять расовую идеологию посредством языка. Оппозиция социальных конструктов «whiteness» и «blackness» непосредственно отражается и в лингвистических явлениях. Например, в таком явлении, как политкорректность, ставшем обязательной частью языковой практики западного общества. Автор также раскрывает содержание популярной в США идеологии расового «дальтонизма» (*colorblind ideology*). В противоположность политике мультикультурализма концепция расового «дальтонизма» не признает никаких различий между расовыми и этническими группами. Однако в американском дискурсе данная идеология стала ассоциироваться со скрытым расизмом, так как приводит к замалчиванию существующих расовых проблем и превращает любое упоминание о расах и расизме в запретную тему. Обосновывается идея о том, что американский английский идеологически связан с категориями расы в сознании его носителей. На примере афроамериканской культуры и афроамериканского английского языка проводится социолингвистический анализ таких феноменов, как «лингвистическая апроприация» и «лингвистическая дискриминация». Диаметральные противоположные проблемы характеризуют американскую действительность: с одной стороны, проблема сохранения языковой и культурной самобытности этнических меньшинств, с другой – их комплексная интеграция в американское общество. Отталкиваясь от концепции «двойственного сознания» Уильяма Дюбуа, автор утверждает, что данный дуализм как одно из важнейших последствий столетий рабства и сегрегации свойственен не только афроамериканскому сообществу, но и евроамериканцам.

Ключевые слова

языковая ситуация, двойственное сознание, «черная» культура, культурная апроприация, идеология расового «дальтонизма» (*colorblind*), афроамериканский английский, языковая идеология, лингвистическая дискриминация, лингвистическая апроприация, междрасовый дискурс, расовые и этнические меньшинства

Для цитирования

Себрюк А. Н. Лингвистическая дискриминация и лингвистическая апроприация как примеры феномена двойственного сознания в США // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 134–148. DOI 10.25205/1818-7935-2019-17-1-134-148

Linguistic Discrimination and Linguistic Appropriation as the Examples of Double Consciousness in the U.S.

Anna N. Sebryuk

*National Research University Higher School of Economics
Moscow, Russian Federation*

Abstract

Despite the abolishment of discriminatory laws and practices in the United States and the subsequent social and legal measures, racism continues to be one of the most acute problems of American society. Modern racism is expressed

in new specific forms, often veiled and difficult to distinguish (covert, subtle racism), which include excessive politeness or the specific use of euphemisms. As a means of expressing basic ideological positions, language also serves as a powerful tool to influence public consciousness. The article reveals the ways to reproduce and enforce racial ideology through language. The opposition of social constructs «whiteness» and «blackness» is directly reflected in linguistic phenomena. For example, in the phenomenon of political correctness, which has become an indispensable part of the language practice in Western society. The author also reveals the content of colorblind ideology. In contrast to the policies of multiculturalism, the concept of colorblindness does not recognize any differences between racial and ethnic groups. However, in American discourse, this ideology is often associated with covert racism, since it leads to the silencing of existing racial problems and makes any mention of races and racism a taboo subject. It is suggested that American English is ideologically linked to the categories of race in the minds of its speakers. Using the example of African-American culture and African American English (AAE), the author conducts a sociolinguistic analysis of such phenomena as «linguistic appropriation» and «linguistic discrimination». Diametrically opposed problems characterize the American reality: on the one hand, the problem of preserving the linguistic and cultural identity of ethnic minorities, on the other hand, their integration into the American society. Drawing on the concept of “double consciousness” by W. E. B. Dubois, the author argues that this dualism, as one of the most important consequences of slavery and segregation, characterizes both African Americans and European Americans.

Keywords

language situation, African American English (AAVE), discrimination; racism, Black culture, double consciousness, colorblind ideology, language ideology, linguistic discrimination, linguistic appropriation, interracial discourse, racial and ethnic minorities

For citation

Sebryuk, Anna N. Linguistic Discrimination and Linguistic Appropriation as the Examples of Double Consciousness in the U.S. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 134–148. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-134-148

Введение

Парадоксы американской межрасовой политкорректности

В последние годы тема межрасовых конфликтов в США вновь приобрела особую актуальность. В противовес всколыхнувшему всю страну общественному движению против насилия в отношении афроамериканцев Black Lives Matter («Жизни черных важны») белыми оппонентами была инициирована кампания White Lives Matter («Жизни белых важны»). Политика мультикультурализма, идеалом которой является параллельное сосуществование разных культур, соседствует с общественным движением «Только английский» (The English-only movement), активно выступающим за установление единых норм американского литературного языка для белых и чернокожих. Различные политические и социально-экономические меры, призванные избавить страну от расистских предрассудков, пока не увенчались успехом. Повышенный интерес к изучению феноменов *whiteness* и *blackness* в американской академической среде (которые до сих пор не обрели эквивалент на русском языке) отражает попытку привлечь внимание к скрытым формам расизма и причинам его сохранения в обществе [Kendall, 2013. Р. 172], несмотря на официальную пропаганду политкорректной этики межрасового общения. Социальная напряженность на расовой почве обуславливает также необходимость рассмотрения вопросов, связанных с черной культурной идентичностью и обоснованностью заимствования евроамериканцами атрибутов культуры других народов.

Цель данного исследования состоит в том, чтобы рассмотреть такие разнонаправленные социолингвистические процессы, как лингвистическая дискриминация и лингвистическая апроприация на основе современных взаимоотношений между белым населением США и афроамериканцами, характеризующихся, на наш взгляд, социокультурным неравенством. В статье доказывается, что американский английский идеологически связан с категориями расы в сознании его носителей.

В современной социолингвистике особое место уделяется вопросу о языке как социальной практике, устанавливающей властные отношения между представителями различных расовых и классовых групп.

Язык играет важнейшую роль в выстраивании идеальной картины мира, хранении и передаче идеологием; именно через язык происходит самоидентификация этнической группы. Он использовался и используется как средство идеологического воздействия и идеологической борьбы. Система взглядов и идеалов, закреплённая в языке, и эмоционально-оценочная окраска языковых единиц находятся в прямой связи с нравственным и социально-экономическим состоянием общества. В этой связи С. Г. Тер-Минасова отмечает: «Язык и идеология взаимодействуют друг с другом: идеология оказывает большое влияние на язык, но и язык влияет на идеологию» [2000. С. 121]. Важно лишь понимать, как подчеркивал Рамсей, что идеологический аспект языка, о котором так часто говорится, заключается не в характеристиках и свойствах языковой системы, которая является автономной, а в использовании ресурсов языка [Rumsey, 1990. P. 356–357]. Для П. Бурдьё то, как люди используют язык, напрямую связано с тем, что они хотят сказать и что они собой представляют в морально-этическом и социокультурном смыслах. Поскольку данный процесс предполагает выбор среди бесчисленных возможностей языковой системы, то намерения говорящего демонстрируются через использование тех или иных вариантов [Bourdieu, 1999. P. 12–13].

Не является исключением и повседневный американский межрасовый дискурс. Исследования наглядно раскрывают его способы воспроизводить и укоренять расовую идеологию, так как он неизбежно материализует и общественное, и индивидуальное сознание. «Язык, – справедливо утверждал Э. Сепир, понимая под языком *language*, т. е. языковую деятельность, – это “путеводитель” в социальной действительности... А “реальный мир”, т. е. представления о социальной реальности, в значительной мере неосознанно строится на основе унаследованных языковых стереотипов той или иной социальной группы [2003. С. 130–131].

В данном исследовании предпринимается попытка разобраться со спецификой таких явлений, как лингвистическая дискриминация и апроприация, через концепцию двойственного сознания (*double consciousness*), предложенную афроамериканским историком, социологом и гражданским активистом У. Дюбуа в известной работе «Души черного народа». В двойственном сознании заключена сущность «двуединства», характерного для жизни и культуры афроамериканцев. Как полагает У. Дюбуа, это особая форма идентичности афроамериканцев, согласно которой, они смотрят на себя глазами белых, [других], через призму жалости или презрения, ощущая себя одновременно и американцами, и африканцами. Определяющей чертой их жизни является парадоксальное условие бытия между двумя идентичностями, а также включения и одновременного исключения из американского общества. Бытие между «черный» и «американец» обуславливает постоянное противоречие и сопровождает афроамериканцев всю жизнь («It is a peculiar sensation, this double-consciousness, this sense of always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity. One ever feels his twoness, – an American, a Negro; two warring souls, two thoughts, two unreconciled strivings; two warring ideals in one dark body, whose dogged strength alone keeps it from being torn asunder») ¹. История американского африканца – это, таким образом, вечный внутренний конфликт, это история борьбы двух систем ценностей, двух идеалов, двух культур.

Очевидно, что ощущение своей чужеродности, «инаковости» во многом еще сегрегированном американском обществе порождает напряженность взаимоотношений афроамериканцев с белым населением. Цвет кожи до сих пор остается нерушимым барьером, отделяющим мир чернокожих американцев от мира белых, несмотря на значительное улучшение материального и социального положения афроамериканцев в последние десятилетия.

Некоторые авторы (чернокожие и белые) склонны даже винить в таком положении не конкретные культурно-исторические и социально-экономические условия формирования межрасовых сообществ, а базовую ментальность всех «европейцев» с ее известной морально-этической символизацией черного цвета. Так, У. Дюбуа убежден, что черный цвет кожи вос-

¹ См.: *Du Bois, W. E. B. The Souls of Black Folk*. URL: <http://www.gutenberg.org>.

принимался белыми американцами как свидетельство зла и порочности, в то время как белый цвет символизировал чистоту и добродетель. В труде «Африка. Очерк по истории африканского континента и его обитателей» он утверждает, что расизм как негативное отношение к расовому типу возник именно на этом культурно-семиотическом основании: «Впервые в истории в понятии “негр” стали неразрывно связывать цвет кожи и расовую принадлежность, рабское положение с деградацией. Белая раса изображалась как “чистая” и высшая; черная – как “нечистая”, неполноценная и, безусловно, низшая...» [Дюбуа, 1961. С. 197].

Однако такой однозначный подход к цветовой символике разделяется далеко не всеми исследователями. К примеру, А. Ф. Фефелов считает, что в современном американском дискурсе в стремлении максимально нейтрализовать расовый дискурс с помощью введения политкорректной лексики происходит упрощенное декодирование оппозиции «белое / черное». Существовавшая ранее европейская семантика белого и черного трактуется крайне ограниченно – зачастую только с точки зрения расовой оппозиции в ее североамериканском контексте. Наличие в языке словосочетаний с прилагательным «black» с негативной коннотацией не стоит интерпретировать как действительность социального дискурса, так как оно необязательно обуславливает их фактическое употребление в речи в конкретных социокультурных ситуациях. Речь, несомненно, точнее отражает общественное сознание говорящих людей и их социальные проблемы, чем лексика языка, закрепленная в словарях. Как подчеркивает А. Ф. Фефелов, нельзя забывать, что традиционная метафорика белого и черного возникла вне колониального периода и не является отражением расовой оппозиции [2017. С. 55, 63]. Следовательно, требуются дифференцированные знания о семантике цветовой символики и ее функционировании в социокультурном поле.

Для афроамериканского этноса США проблема самоидентификации до сих пор является центральной. Идея Ф. Фанона в одной из его самых известных работ «Black Skin, White Masks» созвучна с мыслями У. Дюбуа. Столетия рабства нанесли афроамериканцам культурную травму, отняв у них человеческий статус, лишив чувства человеческого достоинства. Афроамериканцы должны были пересмотреть свое прошлое, для того чтобы реконструировать коллективную идентичность. Он подытоживает сложные пути построения самоидентичности чернокожих, тесно связанные с последствиями расизма в США, следующим образом: «...the black man has two dimensions. A Negro behaves differently with a white man and with another Negro. That this self-division is a direct result of colonialist subjugation is beyond question...» [Fanon, 2008. P. 17].

Однако реальная расовая и межрасовая ситуация в США гораздо сложнее: как чернокожие смотрят на себя с двух противоположных сторон, так и белые американцы воспринимают афроамериканцев двояко. С одной стороны, в сознании белых американцев афроамериканцы ассоциируются с чувственностью, артистичностью и духовностью, с другой – с преступностью, бедностью и девиантным поведением. Такая противоречивость отношения обусловлена не столько расовой метафорикой белого и черного цвета, сколько спецификой американских протестантских ценностей, наиболее характерными из которых являются эгалитаризм и индивидуализм. С одной стороны, белые американцы активно поддерживают афроамериканцев в борьбе за равенство и сочувствуют им, с другой – согласно присущему многим американцам пуритано-протестантскому мировоззрению, высоко ценятся трудолюбие и целеустремленность, отчасти поэтому правительственные программы поддержки меньшинств неустанно вызывают критику с их стороны [Шурыгина, 2017. С. 110]. Такая двойственность ведет к напряжению в общении с афроамериканцами. Нуждаясь в разрядке, в зависимости от ситуации напряженный человек ведет себя или преувеличенно позитивно, или преувеличенно негативно [Stephan, 2000. P. 23–25].

Со времен рабства в Северной Америке для белых американцев было экономически выгодно считать чернокожих если не животными, то, по крайней мере, неполноценными представителями человеческой расы. Такое восприятие, поддерживаемое американскими теологами, учеными, политиками и философами того времени позволяло им распоряжаться

выходцами из Африки как товаром без угрызений совести. Однако, несмотря на это, к примеру, поколениями белые американцы доверяли воспитание своих детей чернокожим няням «Мамму», что противоречит их же расистским убеждениям, что афроамериканцы не способны на что-то большее, чем физический труд, чем примитивная «черная» работа.

Двойственность сознания находит отражение и в новых, завуалированных формах расизма. После отмены так называемых законов Джима Кроу, дискриминация по отношению к афроамериканцам зачастую трансформируется в так называемый скрытый расизм (*covert racism*), который включает в себя проявление излишней вежливости или специфическое употребление политкорректных языковых средств.

Психологи С. Гертнер и Дж. Довидио считают, что одной из форм скрытого расизма является аверсивная (от *aversive* – «питающий отвращение»). Она свойственна тем, кто, заявляя об отсутствии у себя каких-либо расовых предубеждений и выражая сочувствие к чернокожим как к жертвам несправедливости и дискриминации, одновременно ощущает дискомфорт в компании афроамериканцев из-за личной неосознаваемой неприязни и зачастую даже испытывает страх или отвращение, хотя и старается это скрыть [Dovidio et al., 2005. P. 11–12].

Другой формой современного расизма считается амбивалентный расизм (И. Кац, Р. Хасс), который характеризуется наличием как позитивного, так и негативного отношения к какому-либо расовому меньшинству. Амбивалентные расисты переживают эмоциональный конфликт между позитивными и негативными чувствами в отношении к стигматизируемой расовой группе [Whitley, 2009. P. 84–85].

Еще одно проявление двойственности сознания белого населения заметно в явлении «white guilt» (т. е. «белая вина»), описанном Ш. Стилом. По его мнению, современная Америка перешла из эры «превосходства белых» (от «White Supremacy») в эпоху «белой вины», подразумевая под ней индивидуальную или коллективную вину, которую испытывают некоторые белые люди за вред, причиненный расистским отношением белого населения к этническим меньшинствам как в прошлом, так и в настоящее время. Примером «белой вины» часто являются попытки принятия культур и философии других этнических групп в ущерб себе, т. е. игнорируя при этом свои собственные корни и интересы. Стил считает, что феномен «белой вины» является следствием скрытой расистской идеологии [Steele, 2007. P. 12–14]. Публичная демонстрация антирасистских убеждений не то же самое, что реальная помощь чернокожим людям, она не является даже свидетельством отношения к ним как к полноценным людям. Поскольку многие люди в современной Америке отрицают само существование расизма, попытки решить эту проблему, о которой так много говорят, принимают у них форму культивирования вины белых, а не реальной борьбы с институциональными и историческими последствиями расизма.

Идеология расового «дальтонизма» в современной межрасовой коммуникации

В рамках нашего исследования отдельного рассмотрения заслуживает относительно новая расовая идеология «дальтонизма» (*colorblind ideology*), которая в настоящее время парадоксально ассоциируется с аверсивным и амбивалентным расизмом [Awad et al., 2005. P. 1385].

Идеологической целью установки расового «дальтонизма» является построение общества будущего, где люди не будут дифференцировать расы и станут безразлично относиться к любым характерным чертам и факторам расовой или этнической принадлежности в общении друг с другом. Наиболее ярко цель идеологии была сформулирована в контексте движения за гражданские права и международных антирасистских движений 1950-х и 1960-х гг. Центральным положением идеологии послужила идея Мартина Лютера Кинга о том, что когда-нибудь о людях будут судить по «содержанию их характера», а не по «цвету их кожи» [Sears, 2000. P. 6].

Однако в американском межрасовом дискурсе принцип идеологии «дальтонизма» начал вызывать все более негативное отношение, поскольку, как это ни удивительно, стал ассоциироваться с расистским мировоззрением, отрицанием расизма и негативным отношением к антирасистской политике. Главным аргументом критиков является то, что продвижение идеологии расового «дальтонизма» приводит к замалчиванию существующих расовых проблем и превращает любое упоминание о расах и расизме в табу, запретную тему, которую не принято обсуждать открыто. Дело в том, что отрицание расовых и культурных различий не влияет на фактическое положение дел и социальную дифференциацию. Невозможно быстро решить проблемы расовых предубеждений и неравенства путем превращения этой темы в табу [Bonilla-Silva, 2017. Р. 77–78].

На самом деле понятие «раса» не является пустым, поскольку, как показывает реальное положение дел в США, расовая принадлежность действительно влияет на возможности, отношения, доход и многие другие аспекты жизни представителей меньшинств. Более того, белые американцы, которые вряд ли испытывают какие-либо неудобства из-за своей расы, могут открыто игнорировать или отрицать наличие расизма в американском обществе и оправдывать существующий социальный порядок, поскольку чувствуют себя комфортнее. Однако меньшинства, которые регулярно сталкиваются с трудностями из-за своей расовой или этнической принадлежности, абсолютно иначе смотрят на идеологию расового «дальтонизма», так как она создает общество, которое отрицает их негативный расовый опыт и отвергает их культурное наследие и самобытность [Richeson, Nussbaum, 2004. Р. 417–419]. Путем игнорирования расы игнорируется нынешний расизм, который препятствует социальному равенству. В свою очередь, сведение к минимуму последствий расизма ведет к отсутствию поддержки государственных проектов, направленных на борьбу с расовой несправедливостью [Bonilla-Silva, 2017. Р. 95].

Итак, идеология расового «дальтонизма» может интерпретироваться как скрытый расизм и отличается от открытых форм расизма тем, что она не поддерживает негативные установки и не пропагандирует ненависть по отношению к определенным расовым группам, но, распространяя ошибочные представления о понятии «раса», стремясь «отменить» его, она искажает реальную историю межрасовых отношений в США.

Культурная апроприация

Противоречивое отношение в американском обществе также вызывает уникальный вариант английского языка, на котором говорят афроамериканцы, их маркер этничности и средство самоидентификации. Существующие среди белого населения стереотипы об афроамериканцах и пренебрежительное к ним отношение распространяются и на их язык: они говорят на упрощенном и искаженном английском якобы потому, что от природы ленивы, не дисциплинированы и тупы.

Афроамериканский английский изначально сложился как форма общения среди африканских рабов, привезенных в Новый Свет из разных африканских стран между XVI и XIX в. Лишенные материального и духовного наследия, для сохранения своей исконной африканской самобытности и в целях самозащиты от безграничной жестокости своих хозяев, рабы были вынуждены выработать особый язык, который не понимали бы рабовладельцы и который бы стал своего рода символом их сопротивления.

На протяжении своей истории афроамериканский английский назывался по-разному: Negro Nonstandard English, Black English, Black English Vernacular и Ebonics. Названия менялись в связи с изменениями в расовом дискурсе. На сегодняшний момент в американской лингвистике принято использовать название African American English (AAE), т. е. афроамериканский язык.

В непрофессиональной среде AAE часто называют «плохим английским» [Alim, Baugh, 2007. Р. 33]. Как отмечалось ранее, «African-American Vernacular English is stigmatized be-

cause its speakers are stigmatized» [Spears, 1999. P. 9]. Его «стигматизированный» статус был впервые оспорен группой исследователей во главе с американским лингвистом У. Лабовым в 1970-х гг. Они доказали, что афроамериканский английский – это самостоятельная лингвистическая система, обладающая определенными правилами, грамматикой, фонетикой и лексикой. Следовательно, язык афроамериканцев не дает оснований говорить о каких-либо нарушениях в когнитивном развитии его носителей, как считалось ранее [Labov, 1972. P. 202].

Дальнейшее изучение афроамериканского английского лишь подтверждало это заключение. По мнению А. Спирса, «African-American English is not a collection of slang terms, it is not used only by teenagers who spend much time in the streets, it is not a corruption of anything – it has its own history and did not spring historically from any one kind of English. It is a product of the historical blending of mostly English features but also some features of West African languages» [Spears, 2008. P. 523].

Со сложным положением афроамериканцев в американском обществе и противоречивым статусе афроамериканского английского связаны два противоположных языковых явления: лингвистическая апроприация и лингвистическая дискриминация.

Явление лингвистической апроприации (linguistic appropriation) нельзя рассматривать в отрыве от понятия культурной апроприации, из которого оно проистекает. Стоит заметить, что широко используемое в современном западном дискурсе понятие «культурная апроприация» крайне мало затрагивается в российской науке. Тем не менее в последние годы в СМИ и академической среде США разгорается полемика вокруг термина «культурная апроприация» в контексте культуры коренного населения Америки и афроамериканцев.

Согласно Оксфордскому словарю английского языка, культурная апроприация – это «тихое», молчаливое (культурное) заимствование или некорректное присвоение обычаев, практик, идей и т. д. одного народа / общества членами другого народа / общества, как правило, занимающего доминирующее положение («the unacknowledged or inappropriate adoption of the customs, practices, ideas, etc. of one people or society by members of another and typically more dominant people or society»).

Как показывает европейская история, нередко заимствование атрибутов других культур происходило довольно тяжело, а иногда, по мнению афроамериканцев США, и «насиленно». Необычность и экзотичность в глазах европейцев, принадлежащих к доминантной культуре, определенных элементов других, «подчиненных», культур, обуславливает появление моды на них. Доминантная культура заимствует видимые культурные коды других культур, пренебрегая сохранением культурного контекста и обесценивая их. Превратившись в модные тренды и став востребованными, заимствованные или присвоенные атрибуты «нижних» культур приносят финансовую прибыль носителям доминирующей культуры.

Этот процесс эксплуатации «слабых» культур и их «поглощение» доминантными обуславливает тенденцию многих американских исследователей рассматривать культурную апроприацию как крайне негативное явление, в результате которого происходит не взаимообогащение двух культур, а фактический «захват» достижений одной культуры другой.

Как считают некоторые культурологи, такая форма культурного присвоения является привилегией белых, их бессознательной формой угнетения, поддерживающей межкультурное неравенство, поскольку они присваивают себе элементы культуры представителей этнических меньшинств². Д. Поттер также полагает, что культурная апроприация причиняет вред культуре чернокожих США, он рассматривает ее как одну из скрытых форм расизма. Присвоение атрибутов афроамериканской культуры и их дальнейшая ассимиляция в домини-

² См.: Reyes, J. Cumberland County College. Cultural Appropriation & White Privilege. URL: <https://www.academia.edu>.

рующем обществе, зачастую с искажением первоначального смысла, явно не способствует упрощению процесса самоидентификации афроамериканцев³.

Существенное теоретическое влияние на рассмотрение взаимоотношений между культурами и языками через призму господства / подчинения и постколониальных исследований оказал французский социолог П. Бурдьё. Он утверждал, что словесные обмены между людьми происходят на языковых рынках, которые определяют количество символических и материальных ресурсов. Под «рынком», или «полем», П. Бурдьё подразумевал структурированное социальное пространство позиций, в котором позиции и их отношения между собой определены дистрибуцией различных видов ресурсов, или «капитала». Языковые рынки порождают социальную стратификацию, оценивая речевые акты людей, занимающих доминирующее положение в обществе, при этом обесценивая язык социально-экономически угнетённых групп [Bourdieu, 1999. Р. 38–29].

Афроамериканцы и коренные жители Америки, на протяжении столетий угнетаемые белым населением, выступают теперь против неавторизованного использования элементов своего национального фольклора, языка, религиозной символики, одежды, стиля причёсок и головных уборов, музыки, танцев и пр. Американские поп-звезды, дизайнеры и художники, обвиненные в культурной апроприации, вынуждены извиняться и оправдываться перед представителями тех культур, чьи атрибуты и символика были ими использованы [Лозинская, 2017].

Следует заметить, что и в Оксфордском, и в Кембриджском словаре в статье о культурной апроприации приводятся примеры, связанные именно с африканской культурой. «Some see his use of African music as cultural appropriation» (CDAE)⁴. «His dreadlocks were widely criticized as another example of cultural appropriation» (OED)⁵.

Лингвистическая апроприация и лингвистическая дискриминация

Лингвистическая апроприация в США исторически представляла собой важный инструмент белого расизма, способствующий моделированию добродетельного образа белых... одновременно с очернением и стереотипизацией представителей этнических сообществ, у которых заимствовались те или иные атрибуты [Bucholtz, 1999. Р. 444–445]. Таковы, например, ранние образы европейских исследователей и путешественников как просвещённых людей и умышленное унижение коренных жителей «открываемых» континентов до уровня полуживотных-полудикарей.

С точки зрения критической расовой теории, развиваемой сейчас в США, вхождение слов или их значений, особенно из этнических говоров, в систему доминантного языка является ничем иным, как «кражей». Американский лингвист Дж. Хилл дала этому явлению название «linguistic appropriation» и описала его следующим образом: «In linguistic appropriation, speakers of the target language (the group doing the borrowing) adopt resources from the donor language, and then try to deny these to members of the donor language community. But they also achieve it indirectly, by reshaping the meaning of the borrowed material into forms that advance their own interest, making it useless or irrelevant, or even antithetical, to the interests of the donor community. This reshaped meaning may then be imposed on donor speakers» [Hill, 2008. Р. 158].

Итак, лингвистической апроприацией называется тип сложного лингвокультурного заимствования, которое включает в себя «кражу» языкового «капитала» какого-либо специфического лингвоэтнического меньшинства доминирующей нацией; нередко влечёт за собой искажение оригинального значения заимствованных элементов.

³ См.: Potter, D. Cultural Appropriation as Masked Racism: Humor, Hip-Hop and The Tonight Show Starring Jimmy Fallon. URL: <https://www.academia.edu>.

⁴ Cambridge Dictionary of American English. URL: <https://dictionary.cambridge.org>.

⁵ Oxford English Dictionary. URL: <https://en.oxforddictionaries.com>

Заимствования в современном английском белых американцев, подпадающие под понятие лингвистической апроприации, идут преимущественно из следующих трех источников: American Indian languages, African American English и US Spanish.

Лингвистическая апроприация опирается на манипулятивные техники, которые хорошо известны в рекламе и пропаганде. Слова и культуремы трактуются как товар и как собственность, а их значение определяется теми, в чьем распоряжении они оказались. «Захваченное» слово претерпевает изменение значения, приобретая то, которое отвечает интересам заимствующей группы. Для насаждения этих своих значений говорящие на языке-реципиенте должны иметь доминирующее положение по сравнению с говорящими на языке-доноре. Процессу этой лингвистической апроприации во многом способствует контроль над такими общественными институтами, как медиа, учебные заведения и органы законодательной власти, которые отвечают за циркуляцию информационных ресурсов. К лингвистической апроприации причастны, как утверждает Хилл, элиты, к которым она относит ученых, преподавателей, журналистов, писателей, «людей с влиянием», способных перевести в реальную практику расистскую идеологию [Hill, 2008. P. 158–159].

В современной жизни США суть процесса лингвистической апроприации состоит в том, что, «заимствуя» у афроамериканцев, индейцев или испаноязычных американцев, белая Америка «наращивает ценности» американской идентичности, одновременно еще более маргинализируя группы доноров. Как говорит Хилл, такое культурное «заимствование» регулярно наблюдается и в отношении языка чернокожих, позволяя белым американцам как бы «примерять на себя» их культуру с помощью афроамериканского английского языка и сленга, отвергая при этом культурные и социальные основы реальной жизни чернокожих [Ibid. P. 175].

Такая апроприация часто создает эффект пародирования, ироничного комментария различных лингвокультурных фактов жизни афроамериканцев. Данный эффект хорошо виден в следующих примерах «заимствований», в которых не только исказилось первоначальное значение единицы, но и произошло усиление негативных стереотипов в отношении чернокожих.

1. **Bae** (сокращенная форма от «*babe*» – «дорогой», «милый»).

В то время как в общеамериканский лексикон слово вошло только в последние несколько лет, большинство афроамериканцев США используют его в течение десятилетий. В афроамериканском дискурсе эта форма используется для выражения ласки и привязанности к кому-либо («What's up, bae?», «That's my bae»). Данное слово получило распространение с 2013 г. через Твиттер и Инстаграм афроамериканцев, хештеги #baecaughtmesleepin и #cookingforbae. Среди остальных американцев этномаркированная форма *bae* стала широко известна после выхода песни афроамериканского певца Фаррелла Уильямса «Come Get It Bae»⁶. Его значение в стандартном американском английском было быстро искажено до степени бессмысленности: форма *bae* стала применяться даже по отношению к неодушевленным предметам (например, к еде).

2. **Ratchet** – сленговое слово в хип-хопе, которое в его первоначальном смысле относится к грубой, «неотесанной» женщине. Имеет как положительную, так и негативную коннотацию. До популяризации в американском сленге слово использовалось чернокожими преимущественно в значении «getting excited, partying, going hard». Однако после выхода ряда песен, ставших популярными, слово *ratchet* получило более широкое значение и приобрело в американском английском уничижительную сексистскую коннотацию (о грубых женщинах, особенно афроамериканках). Оно усиливает распространение стереотипа об афроамериканской женщине как неуклюжей, необразованной, чрезмерно шумной на публике, вульгарно сексуальной. Неутихающие споры вокруг уместности использования этого слова

⁶ См.: Powell, A. Comments about the Word “Bae” from a YouTube Discussion Thread about Pharrell’s “Come Get It Bae” Record. URL: <http://pancocojams.blogspot.com>.

побудили американскую гражданскую активистку и писательницу запустить в 2012 г. общественную кампанию «Bury the Ratchet», в ходе которой ряд лидеров афроамериканского сообщества анализировали многочисленные негативные стереотипизированные медиаобразы чернокожих женщин в американских СМИ ⁷.

3. **Uppity** («спесивый, наглый»). Согласно этимологическому словарю, слово появилось в 1880 г. и использовалось афроамериканцами в адрес других чернокожих, которые казались им наглыми и нахальными (ED) ⁸. Затем в речи белых американцев оно обрело расистскую коннотацию, так как подразумевало чернокожего, который «не знает своего места».

4. **Thug** («головорез, бандит, гангстер»). Представителями афроамериканского сообщества слово приравнивается к расистскому оскорблению. В американском дискурсе слово преимущественно использовалось в хип-хоп культуре. Знаменитый хип-хоп исполнитель Тупак Шакур популяризировал выражение «thug life» «житуха-бандюгана», которое превратилось в целое социальное движение в начале 90-х гг. XX в. Фраза возникла как защитная реакция на расистские обвинения со стороны белых в том, что афроамериканцы – люди второго сорта, умеющие только грабить, продавать наркотики и убивать. Считается, что Шакур призвал к прекращению насилия и воспитанию детей в соответствии с новыми правилами, стремясь к тому, чтобы дети жили лучше, чем их родители. М. Джефрис в работе «Thug Life: Race, Gender, and the Meaning of Hip-Hop» пишет: «The concept of the thug underwent a... transformation, from signifying disgust, rebellion, and nihilism to evoking coolness and power» [Jeffries, 2011. P. 156–157].

В речи афроамериканцев *thug* – это отчасти хулиган, отчасти герой, демонстрирующий несогласие со сложившимся положением дел и проявляющий своего рода устойчивость перед лицом расизма. Однако, получив распространение в общеамериканском английском, *thug* стало применяться в отношении к чернокожим преступникам, выходцам из неблагополучных районов, например, в адрес мародеров и уличных грабителей. В современном же дискурсе слово считается политкорректной заменой для «nigger».

5. **Shuck and jive (Shuckin' and jivin')**. Появившись в среде рабов на юге США в 1870-е гг., выражение изначально означало «притворство, дурачество и обман» [Major, 1994. P. 378]. Чернокожие рабы весело пели и танцевали, очищая кукурузные початки от шелухи (*shucking corn*). Эта манера подачи себя, наряду с лукавством при общении с рабовладельцами, стала частью уловок для того, чтобы избежать физического или психологического наказания, а нередко и спасти собственную жизнь.

В современном языке употребляется для обозначения увиливания, хитрости, особенно в общении с каким-либо авторитетом или начальником. Фраза вызывает споры и обвинения в расизме, так как обычно употребляется по отношению к афроамериканским деятелям, например к Бараку Обаме. В 2008 г. генеральный прокурор Нью-Йорка Эндрю Куомо подвергся резкой критике со стороны общественности за свое высказывание «You can't shuck and jive at a press conference» в адрес Барака Обамы, который в то время баллотировался от Демократической партии в качестве кандидата против Хиллари Клинтон. В 2012 г. бывший кандидат в вице-президенты от Республиканской партии Сара Пэйлин также вызвала полемику, когда заявила: «Obama's shuck and jive ends with benghazi lies». В 2013 г. Раш Лимбо назвал стратегию президента Барака Обамы по Сирии «shuck and jive» ⁹.

6. **Go all ghetto**. Негативное отношение афроамериканцев к апроприации атрибутов и символики их культуры нашло отражение в таких понятиях, как «wigger», «wannabe black», и в выражении «go all ghetto».

Wigger, wigger и *white nigger* обозначают белого человека, который демонстрирует манеры и стиль одежды, ассоциирующиеся с афроамериканской культурой, и «щеголяет» в своей

⁷ Black Youth Project. Angela Davis Launches 'Bury The Ratchet' Campaign to Fight Negative Images of Black Women. URL: <http://blackyouthproject.com>.

⁸ Online Etymology Dictionary. URL: <https://www.etymonline.com>.

⁹ Reeve, E. Was It Racist for Palin to Accuse Obama of 'Shuck and Jive'? URL: <https://www.theatlantic.com>.

речи словечками из афроамериканского английского. *Wigger* является объединением слов *белый* (white) и уничижительного термина *ниггер* (nigger). В наши дни слово считается оскорбительным и расистским, отражающим стереотипы белых об афроамериканцах. Оно выступает в качестве синонима «white trash» («белого мусора») и может использоваться в уничижительном смысле, подразумевая неудачную попытку культурной апроприации со стороны белого населения [Kitwana, 2006. P. 42].

Аналогом его является выражение «wannabe black» («подражатель черным»). Согласно UD¹⁰, обозначает белого человека, который пытается вести себя как черный (to act black), например использовать сленг и носить приспущенные штаны. Wannabe black – это тот, кто старается подражать чернокожим и выглядит при этом фальшивым и неестественным.

Фраза «go all ghetto», как и другие выражения, содержащие слово «ghetto» в современном американском дискурсе, указывает на негативное отношение к афроамериканцам. Слово *ghetto* («a quarter of a city in which members of a minority group live especially because of social, legal, or economic pressure» – MWD¹¹) стало синонимично всем другим уничижительным словам, предназначенным для высмеивания чернокожих людей за то, как они себя ведут, что носят и как говорят. В UD даются следующие определения современных значений данного слова: «Ghetto is a derogatory term used towards individuals who lack the standards of manners and ethics», «anything that is jury rigged, broken in some way, or otherwise of generally poor quality». Как видно из современных трактовок, слово «ghetto» обрело крайне отрицательную коннотацию.

Противоположным явлением по отношению к лингвистической апроприации можно назвать языковую дискриминацию (linguistic discrimination, linguicism). Языковой дискриминацией называют несправедливое обращение к индивиду исключительно на основе использованного им языка, а именно дискриминацию в отношении кого-либо на основе того, как он говорит или пишет [Murillo, Smith, 2011. P. 147]. Например, в зависимости от акцента, специфики лексико-грамматических и синтаксических средств в речи говорящего могут сформироваться суждения о его финансовом положении, образовании, социальном статусе и характере. В Словаре лингвистических терминов приводится следующее определение лингвистической дискриминации: «Нарушение гражданских и/или экономических прав личности или этнической группы по причине незнания или плохого знания языка численно или социально доминирующего на данной территории этноса» (СТЛ, 2010).

Американский исследователь Д. Макуортер считает проблему лингвистической дискриминации особенно злободневной именно в США. Поскольку некоторые афроамериканцы говорят на определенном «нестандартном» варианте английского языка, который рассматривается как неправильный и искаженный, они часто являются объектами лингвистической дискриминации [McWhorter, 2011. P. 37]. ААЕ нередко воспринимается американским обществом как признак ограниченного интеллекта или плохого образования. Он никогда не получал социального престижа стандартного языка, что приводит к широко распространенному заблуждению о том, что это грамматически неполноценная форма английского языка. «Central to this ideology is the belief that Black English is ‘bad grammar’, ‘slang’, and a ‘pseudo-language’» [Ronkin, Karn, 1999. P. 362]. Кроме того, как и многие другие нестандартные диалекты (особенно креольские), ААЕ иногда называют «ленивым», «маргинальным» или «плохим» английским языком.

В американской прессе регулярно появляются публикации о дискриминации по языковому признаку со стороны учителей и потенциальных работодателей. Диглоссия не признается официально допустимой в американских учебных заведениях. В то время как в некоторых странах мира, таких как Финляндия, Италия, Марокко, нестандартные грамматические или лексические формы объясняются региональным / географическим происхождением говора-

¹⁰ Urban Dictionary. URL: <https://www.urbandictionary.com>.

¹¹ Merriam Webster Dictionary. URL: <https://www.merriam-webster.com>.

щих, в США их характеризуют с точки зрения интеллектуальных способностей и достижений.

Лингвистическая дискриминация обусловлена существованием стандартной идеологии языка. Лингвистические идеологии – представления о самом языке, формируемые политическими и экономическими интересами, – могут способствовать воспроизводству и распространению стереотипов [Lippi-Green, 2012. P. 95–96]. Функции лингвистических идеологий разнообразны. Они рационализируют и оправдывают понятия общества о структуре его языка (или языков) и способах его использования. Они подразделяют языковые структуры и способы их использования на «хорошие» и «плохие», правильные и неправильные, связывая их с людьми, которых считают хорошими или плохими, моральными или аморальными [Fairclough, 2001. P. 14].

Как считает Дж. Хилл, политический или экономический интерес в понятии «лингвистической идеологии» совершенно очевиден в случае с афроамериканцами. Например, когда белый оратор осуждает афроамериканский английский как безграмотный сленг, можно допустить, что это осуждение продвигает интересы говорящего к расовому статус-кво, к той самой белой привилегии. «Language ideologies represent themselves as forms of common sense»; «they rationalize and justify the forms and functions of text and talk» [Hill, 2008. P. 33–34].

Афроамериканцы вынуждены соответствовать европейско-американскому обществу, чтобы добиться успеха в жизни, и данное соответствие в конечном счете означает «искоренение черного языка... и принятие языковых норм белого среднего класса» [Smitherman, 1986. P. 173].

Одной из форм лингвистической дискриминации является лингвистическое профилирование (Linguistic profiling), представляющее собой практику выявления социальных характеристик личности на основе слуховых (фонетических) ориентиров, в частности диалекта и акцента. Автор понятия, американский лингвист Джон Боу считает, что в начале телефонного разговора идентификация этнической принадлежности человека по звуку его голоса порождает предвзятое отношение по признаку расы. Согласно его исследованию, было зафиксировано, что если голос звонящего идентифицируется как принадлежащий афроамериканцу или мексиканцу, то со стороны некоторых потенциальных работодателей, агентов по недвижимости, кредитных сотрудников и поставщиков услуг следует расовая дискриминация. Задолго до того, как они смогли оценить способности, достижения, кредитный рейтинг, трудовую биографию абонентов, они прекращали общение исключительно на основании звучания голоса. «Those who sound white get the appointment» [Baugh, 2017. P. 2–3].

Заключение

Исследуя проблемы, связанные с языковой идеологией западного общества, в особенности американского, нельзя не коснуться феномена политкорректности, охватившей все сферы жизни в США и вызывающей крайне противоречивые чувства. Несомненна положительная сторона идеи, которая легла в основу явления, а именно стремление уважать чувства и достоинство индивида и соблюдать его права во всех сферах жизни. С другой стороны, новые культурно-поведенческие и языковые нормы часто граничат с абсурдностью, не только ограничивая права человека на свободу слова, но и приводя к искажению смыслов и к дезинформации. Говоря об особенностях современной расовой политкорректности, нельзя не отметить недостаток четкой дифференциации между расовыми и расистскими употреблениями так называемых «неполиткорректных» слов [Фефелов, 2014] и нечеткость определения, какие из них маркированы и содержат расистские значения, а какие несут нейтральные обозначения различных рас и этносов, населяющих США [Фефелов, 2017. С. 63].

Политкорректность и идеология расового «дальтонизма» (colorblind) в США, призванные бороться с расизмом, поддерживают равенство в языке, но не на практике, приводя к замалчиванию проблем, стоящих перед обществом.

Определяющей чертой жизни афроамериканцев является парадоксальное условие бытия между двумя идентичностями. Двойственное сознание свойственно и белому населению. Оно проявляется в противоречивом отношении к афроамериканцам: от восхищения и желания им подражать до пренебрежения и дискриминации.

Асимметричные взаимоотношения между белым населением и афроамериканцами находят прямое отражение в языковых явлениях.

Лингвистическая апроприация как одна из форм культурной апроприации воспринимается в США как продолжение привилегии белого населения. Белая Америка не ссылается должным образом на первоисточники, игнорирует исторический контекст и зачастую лишь закрепляет стереотипы об афроамериканцах.

Язык афроамериканцев становится объектом постоянной критики со стороны общественности, а также основанием для лингвистического профилирования и лингвистической дискриминации, поскольку его связывают не с маркером этничности и средством самоидентификации его носителей, а с их уровнем образования и интеллектуальными способностями.

Список литературы / References

- Дюбуа У. Э. Б.** Африка. Очерк по истории африканского континента и его обитателей. М.: Иностр. лит., 1961. 358 с.
Du Bois, W. E. B. An Essay Toward a History of the Continent of Africa and Its Inhabitants. Moscow, Inostrannaya literatura Publ., 1961, 358 p. (in Russ.)
- Лозинская А.** Косички как преступление. Почему звезд обвиняют в расизме // Газета.ру. 2017. URL: https://www.gazeta.ru/lifestyle/style/2017/06/a_10741403.shtml (дата обращения 20.08.2018).
Lozinskaya, A. Pigtales as a Crime. Why are Celebrities Accused of Racism? *Gazeta.ru*. 2017. URL: https://www.gazeta.ru/lifestyle/style/2017/06/a_10741403.shtml (retrieved 20.08.2018). (in Russ.)
- Сепир Э.** Статус лингвистики как науки // Языки как образ мира. М.: ООО «Издательство АСТ»; СПб.: Terra Fantastica, 2003. С. 127–138.
Sepir, E. The status of linguistics as a science. Languages as an image of the world. Moscow, St. Petersburg, 2003, p. 127–138. (in Russ.)
- Тер-Минасова С. Г.** Язык и межкультурная коммуникация. М.: Слово/Slovo, 2000. 261 с.
Ter-Minasova, S. G. Language and Intercultural communication. Moscow, Slovo/Slovo, 2000, 261 p. (in Russ.)
- Фефелов А. Ф.** Расовые и расистские сигналы в детских романах Марка Твена в свете политкорректной правки // Вестн. Новосиб. гос. ун-та. Серия: Лингвистика и межкультурная коммуникация. 2014. Т. 12, вып. 2. С. 40–49.
Fefelov, A. F. Dialectics of Race and Racism Markers in “The Adventures of Tom Sawyer” and “The Adventures of Huckleberry Finn”. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2014, vol. 12, no. 2, p. 40–49. (in Russ.)
- Фефелов А. Ф.** Типы реализации переносной семантики цветообозначений black / white в мейнстримных СМИ США // Вестн. Новосиб. гос. ун-та. Серия: Лингвистика и межкультурная коммуникация. 2017. Т. 15, № 4. С. 52–67.
Fefelov, A. F. Semantic Patterns of Interpreting the Symbolic Meanings of Color Terms Black/White in Some US Mainstream Media. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2017, vol. 15, no. 4, p. 52–67. (in Russ.)
- Шурыгина Е. Н.** Особенности культурного концепта AMERICAN MATERIALISM как одного из основополагающих аспектов американской картины мира (на материале историко-публицистической литературы и современных англоязычных словарей) // Вестн. Новосиб. гос. ун-та. Серия: Лингвистика и межкультурная коммуникация. 2017. Т. 15, № 2. С. 106–117.
Shurygina, E. N. The Features of the Cultural Concept “American Materialism” as One of the Most Significant Aspects of the American Worldview (On the Basis of Historical and Publicistic Texts and Modern Dictionaries of the English Language). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2017, vol. 15, no. 2, p. 106–117. (in Russ.)

- Alim, H. S., Baugh, J.** Talkin' black talk: Language, Education, and Social Change (Multicultural Education Series). New York, Teachers College Press, 2007, 208 p.
- Awad, G., Cokley, K., Ravitch, J.** Attitudes toward affirmative action: A comparison of color-blind versus modern racist attitudes. *Journal of Applied Social Psychology*, 2005, vol. 35 (7), P. 1384–1399.
- Baugh, J.** Linguistic Profiling and Discrimination. The Oxford Handbook of Language and Society, 2017.
- Bonilla-Silva, E.** Racism without Racists Color-Blind Racism and the Persistence of Racial Inequality in America. Rowman & Littlefield, 2017, 376 p.
- Bourdieu, P.** Language and Symbolic Power. Harvard University Press, 1999, 320 p.
- Bucholtz, M.** You da man: Narrating the racial other in the production of White masculinity. *Journal of Sociolinguistics*, 1999, no. 3 (4), p. 443–460.
- Dovidio, J. F., Glick, P., Rudman, L. A.** Introduction: Reflecting on The Nature of Prejudice: Fifty Years after Allport. In: On the nature of prejudice: fifty years after Allport / Ed. by J. F. Dovidio, P. Glick, L. A. Rudman. Oxford, Blackwell Publ., 2005. p. 1–16.
- Fairclough, N.** Language and power. Harlow, Pearson Education, 2001, 240 p.
- Fanon, F.** Black Skin, White Masks. New York, Grove Press, 2008, 206 p.
- Hill, J.** The Everyday Language of White Racism. John Wiley & Sons, 2008, 240 p.
- Jeffries, M.** Thug Life: Race, Gender, and the Meaning of Hip-Hop. Chicago, University of Chicago Press, 2011, 263 p.
- Kendall, F.** Understanding White privilege: creating pathways to authentic relationships across race. New York, Routledge, 2013, 215 p.
- Kitwana, B.** Why White Kids Love Hip Hop: Wankstas, Wiggers, Wannabes, and the New Reality of Race in America. Civitas Books, 2006, 240 p.
- Labov, W.** Language in the Inner City: Studies in the Black English Vernacular. Philadelphia, University of Pennsylvania Press, 1972, 412 p.
- Lippi-Green, R.** English with an Accent: Language, Ideology and Discrimination in the U.S. New York, Routledge, 2012, 360 p.
- Major, C.** Juba to jive: a dictionary of African-American slang. Penguin Books, 1994, 548 p.
- McWhorter, J. H.** What language is (and what it isn't and what it could be). New York, Gotham Books, 2011, 240 p.
- Murillo, L. A., Smith, P. H.** "I will never forget that": Lasting effects of language discrimination on language-minority children in Colombia and on the U.S.-Mexico border. *Childhood Education*, 2011, no. 87 (3), p. 147–153.
- Richeson, J., Nussbaum, R.** The impact of multicultural versus color-blindness on racial bias. *Journal of Experimental Psychology*, 2004, no. 40, p. 417–423.
- Ronkin, M., Karn, H.** Talking Back, Talking Black: Truths About America's Lingua Franca. *Journal of Sociolinguistics*, 1999, no. 3 (3), p. 360–380.
- Rumsey, A.** Wording, Meaning, and Linguistic Ideology. *American Anthropologist*, 1990, vol. 92, no. 2, p. 346–361.
- Spears, A.** Race and Ideology: Language, Symbolism, and Popular Culture. Wayne State University Press, 1999, 242 p.
- Spears, A.** Pidgins / Creoles and African American English. The Handbook of Pidgins and Creoles. Malden, Blackwell Publishers, 2008, p. 512–542.

- Smitherman, G.** Talkin and testifyin: The Language of Black America. Detroit, Wayne State University Press, 1986, 285 p.
- Steele, S.** White Guilt: How Blacks and Whites Together Destroyed the Promise of the Civil Rights Era. Harper Perennial, 2007, 208 p.
- Stephan, W. G., Stephan, C. W.** An integrated threat theory of prejudice. Reducing prejudice and discrimination. Mahwah, Lawrence Erlbaum Associates Publishers, 2000, p. 23–45.
- Whitley, B.** The Psychology of Prejudice and Discrimination. Wadsworth Publ., 2009, 720 p.

Список словарей / List of Dictionaries

СЛТ – **Жеребило Т. В.** Словарь лингвистических терминов. Назрань: Пилигрим, 2010. 486 с.
CDAE – Cambridge Dictionary of American English. URL: <http://dictionary.cambridge.org>
MWD – Merriam-Webster Dictionary. URL: <https://www.merriam-webster.com>
ED – Online Etymology Dictionary. URL: <https://www.etymonline.com>
OED – Oxford English Dictionary. URL: <https://en.oxforddictionaries.com>
UD – Urban Dictionary. URL: <https://www.urbandictionary.com>

*Материал поступил в редколлегию
Date of submission
31.08.2018*

Сведения об авторе / Information about the Author

Себрюк Анна Набиевна, кандидат филологических наук, доцент кафедры английского языка для гуманитарных дисциплин департамента иностранных языков Национального исследовательского университета «Высшая школа экономики» (ул. Мясницкая, 20, Москва, 101000, Россия)

Anna N. Sebryuk, PhD in Philology, associate professor, School of Foreign Languages, Department of English Language for the Humanities, National Research University Higher School of Economics (20 Myasnitskaya Str., Moscow, 101000, Russian Federation)

sebryuk.anna@gmail.com

Динамика образа войны в русском языковом сознании (психолингвистический аспект)

Р. А. Кафтанов

*Институт филологии СО РАН
Новосибирск, Россия*

Аннотация

Статья посвящена анализу изменений, связанных с образом войны в русском языковом сознании, на материале психолингвистических ассоциативных экспериментов. Результаты анализа могут применяться в комплексной гуманитарной экспертизе, направленной на мониторинг этнополитической ситуации в рамках своевременной диагностики зарождающихся конфликтов. Сравнительно-сопоставительный анализ русских ассоциативных баз, полученных в разные годы, позволил вскрыть динамику связей, образуемых словом *война* на макроуровне (в ядре языкового сознания) и на микроуровне (в ассоциативном поле, объединяющем реакции на слово-стимул *война*). Сравнение с материалами иноязычных ассоциативных баз (британской, американской, французской и испанской) позволило выявить специфику смысловой структуры общерусских образов языкового сознания, а также универсальные образы, характерные для русского языкового сознания и языкового сознания европейских народов. С помощью ассоциативного эксперимента, проведенного с российскими военнослужащими, выявлены профессиональные особенности связей, образуемых словом *война*. В языковом сознании военных слово *война* образует большое количество ассоциативных связей, тогда как в бытовом сознании это понятие не обладает высокой степенью актуальности. Во всех анализируемых ассоциативных базах к наиболее частотным ассоциатам, вызванным словом *война*, относятся слова *мир* и *смерть*. При этом если в европейских ассоциативных базах разных лет неизменным остается высокая частотность реакции *мир*, образующей вместе со стимулом *война* универсальную дихотомию, то в языковом сознании русских в течение последних сорока лет на первое место по частотности вышла ассоциация «война – смерть». Различия в степени устойчивости дихотомии «война – мир» в языковом сознании русских и европейцев могут быть связаны с изменчивостью смысловой наполненности обоих членов дихотомии, о чем свидетельствуют результаты их этимологического анализа, предпринятого в ходе данного исследования.

Ключевые слова

война, мир, структура ассоциативного поля, универсальное и профессионально специфичное, русские ассоциативные базы данных и словари

Для цитирования

Кафтанов Р. А. Динамика образа войны в русском языковом сознании (психолингвистический аспект) // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 149–160. DOI 10.25205/1818-7935-2019-17-1-149-160

Dynamics of the Image of War in the Russian Language Consciousness (Psycholinguistic Aspect)

Roman A. Kaftanov

*Institute of Philology SB RAS
Novosibirsk, Russian Federation*

Abstract

The article is devoted to the psycholinguistic analysis of changes the image of war has been undergoing in the Russian language consciousness. The research of the word *war* on the basis of psycholinguistic associative experiments can open access to the understanding of meanings given to the word by typical native speakers, as well as to the understanding of images that arise in their linguistic consciousness. A comparative analysis of the Russian associative bases

obtained in different years made it possible to reveal the dynamics of connections formed by the word *war* at the macro level (in the core of linguistic consciousness) and at the micro level (in the associative field including reactions to the word-stimulus *war*). Comparison of the data of Russian and European (British, American, French and Spanish) associative bases exposed the common and the Russian-only of the semantic structure of the image of war in linguistic consciousness of nations under analysis. An associative experiment made with Russian military servicemen helped reveal professional features of the associations evoked by the word *war*, primarily, a higher degree of relevance of the concept *war* evidenced by a great number of associative connections. In all associative bases under analysis, the most frequent reactions caused by the stimulus *war* are the words *peace* and *death*. However, the European associative bases demonstrate stability of the high frequency of reaction *peace* in the course of years, while in the Russian linguistic consciousness the *war* – *death* association has become more frequent in the last forty years. This phenomenon may be related to the variability of the meaningfulness of the both members of the universal dichotomy *war* – *peace*, as evidenced by the results of their etymological analysis undertaken in this research. The word *peace* in various languages originally meant “harmony, integrity”, than “the absence of war”. This original meaning seems to be more correlated to the primary meaning of the concept of war in Romance and Germanic (“confusion”), than to the historic denotation of the term to “hunting, stalking” in Slavic and Baltic languages. The results of the psycholinguistic analysis do not reveal any signs of heightened militarism in the Russian mass consciousness. The war is perceived by native Russian speakers as a negative phenomenon, which instills confidence in the predominance of peace-loving traits in the Russian national character.

Keywords

war, peace, language consciousness, structure of the associative field, universal and professionally specific, Russian associative databases and dictionaries

For citation

Kaftanov, Roman A. Dynamics of the Image of War in the Russian Language Consciousness (Psycholinguistic Aspect). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 149–160. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-149-160

Введение

Война сопровождает человечество на протяжении всей его истории (хотя некоторые исследователи указывают на факты существования обществ без войн [Turnbull, 1987; Fry, Söderberg, 2013]). Американский физик Р. Смолли считает, что война – шестая из десяти проблем, с которыми человечество столкнется в ближайшие 50 лет (после энергетического кризиса, истощения водных ресурсов, проблемы продовольственного снабжения, разрушения окружающей среды, роста бедности) [Smalley, 2005. P. 413–414]. Неслучайно феномен войны волнует представителей различных научных областей: историков, социологов, антропологов, психологов, философов.

В социологии и психологии широкое распространение получили идеи о естественности войны, которая «вытекает из самой природы вещей, имеет под собой твердую биологическую основу, и на практике ее едва ли можно избежать» [Фрейд, 1992. С. 335]. Другие доводы в пользу естественности войны как проявления природы человека приводятся в [Joost, Meerloo, 2009; Šmihula, 2013]. Противоположная точка зрения состоит в том, что война имеет исторические и социальные корни. Подобные идеи высказываются в [James, Friedman, 2006]. Среди причин, порождающих войны, исследователи видят: рост населения и нехватку ресурсов [Мальтус, 1993], сочетание в некоторых регионах большого количества молодых мужчин с нехваткой постоянной мирной работы [Fuller, 2003; 2004], экономическую конкуренцию между странами [Фромм, 1994; Fearon, 1995].

Война является одной из важнейших тем философских исследований. Некоторые мыслители оправдывали войну, как необходимость: Г. В. Ф. Гегель [1978. С. 229], Ф. Ницше [1990. С. 34], В. С. Соловьев [1999. С. 11], Н. А. Бердяев [2012. С. 240], Ф. М. Достоевский [1981] и др. Гуманистическая критика войны связана с именами Эразма Роттердамского [Трактаты о вечном мире, 2003. С. 30] и И. Канта [1966. С. 264] (см. также [Орехов, 2014. С. 63]). Идеи пацифизма получили свое философско-этическое развитие в произведениях Л. Н. Толстого, А. Швейцера, Д. Л. Андреева, Б. Рассела, О. Хаксли и др.

Таким образом, феномен войны входит в круг объектов исследования гуманитарных наук. Данные этих наук в сочетании с данными психолингвистики могут послужить основой ком-

плексной гуманитарной экспертизы, направленной на мониторинг этнополитической ситуации. Такой мониторинг способствовал бы своевременной диагностике зарождающихся конфликтов. Кроме того, комплексная гуманитарная экспертиза может быть «направлена на поиск путей оптимизации гибких ресурсов развития в системе образования» [Шапошникова, 2017. С. 124] с целью формирования у подрастающих поколений ассертивных поведенческих и мировоззренческих моделей, повышающих адаптивность к экстремальным ситуациям.

Решение этих задач предполагает изучение сознания человека, в том числе психолингвистическими способами. Один из таких способ заключается в овнешнении сознания другого человека с помощью языковых знаков (слов, свободных и устойчивых словосочетаний, предложений, текстов). «Совокупность образов сознания, формируемых и овнешняемых при помощи языковых средств» [Тарасов, 2000. С. 26], в отечественной психолингвистике получило название «языковое сознание». Поскольку «сознание человека всегда этнически обусловлено» [Леонтьев, 1993. С. 20], языковое сознание имеет национально-культурную специфику.

Один из методов исследования языкового сознания и его образов состоит в анализе ассоциативно-вербальной сети, полученной при помощи ассоциативного эксперимента. Ассоциативно-вербальная сеть позволяет системно и целостно моделировать языковую картину мира носителя языка [Черкасова, 2014. С. 406]. Следовательно, с помощью ассоциативно-вербальной сети исследователь может «выявить влияние особенностей культуры на образы языкового сознания ее носителя» [Уфимцева, 2011. С. 236–237].

Психолингвистическое исследование слова *война* может открыть доступ к пониманию того, какие смыслы вкладывают в это слово носители данного языка и культуры, какие типичные образы возникают в их языковом сознании.

Для исследования этого вопроса мы использовали следующие источники:

- САНРЯ – «Словарь ассоциативных норм русского языка» А. А. Леонтьева, на основе данных ассоциативного эксперимента, проводившегося с 1969 по 1972 г.;
- РАС – «Русский ассоциативный словарь» Ю. Н. Караулова и др., на основе данных ассоциативного эксперимента, проводившегося с 1986 по 1997 г.;
- ЕВРАС – Русский региональный ассоциативный словарь-тезаурус Г. А. Черкасовой и Н. В. Уфимцевой, на основе данных ассоциативного эксперимента, проводившегося в европейской части России с 2008 по 2013 г.;
- СИБАС – Русская региональная ассоциативная база данных И. В. Шапошниковой и А. А. Романенко, на основе данных ассоциативного эксперимента, проводившегося в азиатской части России с 2008 по 2013 г.

Эти ассоциативные базы могут рассматриваться в качестве эталонных для анализа общерусского языкового сознания в синхронно-диахронном аспекте [Шапошникова, 2017].

В 2015 г. в одном из высших военных учебных заведений, находящихся на территории Сибирского федерального округа, нами был проведен ассоциативный эксперимент среди курсантов с целью выявления особенностей профессионального языкового сознания. Результатом этого эксперимента стал подкорпус ассоциаций военных в системе СИБАС (далее – ПВАС).

Во временном плане указанные источники можно представить как словари трех поколений жителей нашей страны: САНРЯ – словарь первого поколения, РАС – второго, ЕВРАС, СИБАС и ПВАС – третьего.

Ассоциативные базы САНРЯ, РАС, ЕВРАС и СИБАС позволяют выделить ядро русского языкового сознания. Ядро объединяет слова, которые «имеют наибольшее число связей, т. е. вызваны в качестве ответов на наибольшее число стимулов» [Караулов, 2000. С. 194], представляют особое значение для испытуемого как личности и отражают «самые емкие понятия, связь с которыми имеет максимальную вероятность воспроизведения» [Залевская, 1990. С. 149].

Ассоциативные связи слова «война» по данным психолингвистических экспериментов

Данные ассоциативных экспериментов показали, что слово *война* либо не входит в ядро языкового сознания (САНРЯ, ЕВРАС, СИБАС), либо расположено близко к границе ядра (как в РАС, где это слово имеет ранг 73¹). На основании этого факта мы можем сделать вывод о том, что понятие, выраженное словом *война*, не обладает высокой степенью актуальности в сознании рядового носителя русского языка.

Однако это суждение справедливо и в отношении носителей других языков, о чем свидетельствуют материалы ассоциативных экспериментов. Так, в ядре английского языкового сознания, полученном А. А. Залевской [1981] на материале ЕАТ слово *war* имеет ранг 70. В ядре французского языкового сознания, выведенном из ассоциативного словаря DAF, слово *guerre* имеет ранг 56. В испанской ассоциативной базе (Санчес Пуиг и др., 2001) слово *guerra* вообще не входит в ядро языкового сознания. Таким образом, относительно невысокая степень актуальности понятия «войны» для обыденного сознания имеет универсальный характер.

На фоне этого резко выделяется ядро ассоциативных доминант ПВАС, где слово *война* имеет ранг 3. Такой высокий ранг свидетельствует в пользу актуальности этого слова и выражаемого им понятия для тех, чья профессиональная деятельность связана с войной.

Во всех рассматриваемых нами ассоциативных базах слово *война* было использовано в качестве одного из стимулов. Полученное в результате ассоциативное поле (далее – АП) ВОЙНА представляет интерес для нашего исследования, так как каждое отдельное АП «является отражением тех знаний, которые стоят за словом в данной культуре» [Уфимцева, 2015. С. 84].

В табл. 1 приведены характеристики АП ВОЙНА в САНРЯ, РАС, ЕВРАС и СИБАС. Характеристики и содержание АП ВОЙНА в ПВАС приведены в табл. 2. В списках реакций после каждой реакции стоит цифра полужирного начертания, обозначающая количество испытуемых, ответивших данным словом на стимул ВОЙНА. Несколько разных реакций с одним и тем же индексом частотности расположены в алфавитном порядке и отделяются друг от друга точкой с запятой.

Мы полагаем, что сравнительно-сопоставительный анализ ассоциативных полей, созданных вокруг одного и того же стимула, позволяет не только вскрыть различия, обусловленные профессиональной принадлежностью испытуемых (гражданские или военные), но и обнаружить изменения, которые претерпели образы и акцентуации смыслов, связанные с понятием «война», в течение 40 лет (с момента публикации САНРЯ).

Семантика высокочастотных реакций на стимул ВОЙНА во всех рассматриваемых ассоциативных базах указывает на то, что в массовом сознании война характеризуется негативными оценками, является противоположностью миру и источником бедствий, в первую очередь, смерти: высокой устойчивостью обладают реакции *мир* и *смерть*. Между тем сравнительно-сопоставительный анализ данных табл. 1 и 2 позволяет выявить ряд различий в семантике и интенсивности высокочастотных реакций АП ВОЙНА в рассматриваемых источниках.

1. В ассоциативных базах САНРЯ и РАС самая частотная реакция – *мир*, тогда как в ЕВРАС, СИБАС и ПВАС самая частотная реакция – *смерть*.

2. Наблюдается снижение частотности отсылок к Великой Отечественной войне в словарях третьего поколения. Мы уже не видим реакции *Отечественная* среди высокочастотных, как это наблюдается в САНРЯ и РАС.

¹ Здесь и далее величина ранга обратно пропорциональна частотности реакции. Другими словами, ранг 1 присваивается слову, вызванному максимальным количеством стимулов.

Таблица 1

Ассоциативное поле ВОЙНА в русских ассоциативных базах

Table 1

Frequency of reactions *peace* and *death* to stimulus *war* in Russian associative bases

Показатель	САНРЯ	РАС	ЕВРАС	СИБАС
Всего реакций	205	631	535	498
Различные реакции	82	211	160	159
Одиночные реакции	57	146	105	116
Устойчивые реакции	148	485	430	382
Высокочастотные реакции (свыше 1 % АП)	мир 56 ; ужас 11 ; Отечественная 9 ; страшная 7 ; не нужна; ужасная 6 ; жестокая; народная; разруха; разрушительная; смерть 4 ; горе; захватническая; кровь; ненужная; страшно 3	мир 92 ; миров 51 ; и мир 41 ; смерть 28 ; Отечественная 18 ; страшная 17 ; ужас 14 ; мировая 13 ; жестокая 12 ; горе; народная 9 ; кровь; страх 7	смерть 83 ; мир 62 ; миров 38 ; и мир 33 ; кровь 21 ; ужас 13 ; плохо 12 ; оружие 11 ; боль; зло 10 ; страх 9 ; мировая 8 ; жестокая; разруха; убийство 6	смерть 77 ; мир 53 ; миров 37 ; и мир 35 ; кровь 26 ; зло 15 ; страх 11 ; горе 10 ; боль; оружие; убийство 8 ; плохо 7 ; жестокая; ужас 6

Таблица 2

Ассоциативное поле ВОЙНА в подкорпусе военных ассоциаций в системе СИБАС

Table 2

Frequency of reactions *peace* and *death* to stimulus *war* in non-Russian associative bases

Показатель	ПВАС
Всего реакций	666
Различные реакции	222
Одиночные реакции	150
Устойчивые реакции	516
Высокочастотные реакции (свыше 1 % АП)	смерть 83 ; мир 65 ; миров 52 ; зло 30 ; и мир; плохо 23 ; победа 19 ; страх 14 ; мировая 12 ; скоро 8 ; кровь 7

3. Наблюдается постепенное снижение частотности эпитетов, характеризующих войну. Если в САНРЯ война в сознании испытуемых описана как *страшная 7*; *ужасная 6*; *жестокая 4*; *разрушительная 4*; *ненужная 3*, для испытуемых РАС война *страшная 17* и *жестокая 12* (количество высокочастотных эпитетов сократилось), то в СИБАС и ЕВРАС она

только *жестокая* 6, а у военных вообще отсутствуют оценочные эпитеты среди высокочастотных реакций.

4. Одна из отличительных особенностей ПВАС – высокая частотность реакции *победа* 19 при низкой частотности подобных реакций в других ассоциативных базах.

Не имея возможности в рамках одной статьи проанализировать выявленные нами все различия, рассмотрим первое из них – смена рангов реакций *мир* и *смерть*. Сравнивая ассоциативные базы трех поколений, можно заметить постепенное выдвижение реакции *смерть* на первое место. Если в САНРЯ реакция *смерть* расположена недалеко от периферии (4 реакции), то в РАС ее частотность значительно возросла (28 реакций), чтобы затем, ко времени составления ЕВРАС, СИБАС и ПВАС, занять ядро АП, оттеснив на второе место реакцию *мир*. По всей видимости, в сознании россиян за последние 40 лет произошли какие-то процессы, в результате которых война стала более прочно ассоциироваться со смертью.

Частотность реакции *мир* связана с дихотомией «война – мир», универсальность которой подтверждается иноязычными ассоциативными словарями. Самая частотная реакция во всех АП, соответствующих русскому АП ВОЙНА переводится на русский язык словом *мир* (как отсутствие войны): в британской ассоциативной базе (ЕАТ) – *peace* 41 (из 99 реакций АП), во французской (DAF) – реакция *paix* 152 (из 520), в испанской (Санчес Пуиг и др., 2001) – *paз* 155 (из 585). Второе слово в этих же источниках соответствует русскому слову *смерть*: *death* 9, *mort* 53, *muerte* 47. Ассоциативные эксперименты рассматриваемых нами иноязычных источников проводились не только в разных странах, но и в разное время (британский ЕАТ по времени соответствует САНРЯ, испанская ассоциативная база – РАС, французский DAF – ЕВРАС и СИБАС). Мы считаем, что это служит дополнительным обоснованием универсальности дихотомии «война – мир». Частотность реакции со значением «мир» настолько высока, что от нее значительно отстоит следующая по частотности реакция со значением «смерть». На этом фоне особенно заметны изменения, произошедшие в русском языковом сознании к моменту проведения экспериментов ЕВРАС и СИБАС, приведшие к тому, что два слова поменялись рангами в АП.

На рис. 1 показано соотношение частотности реакций *мир* и *смерть* в русских ассоциативных базах. Цифры по оси ординат соответствуют процентному соотношению между частотностью данной реакции и общим количеством реакций в АП. На рис. 2 показаны аналогичные данные иноязычных источников: британского (ЕАТ), французского (DAF), испанского (Санчес Пуиг и др., 2001).

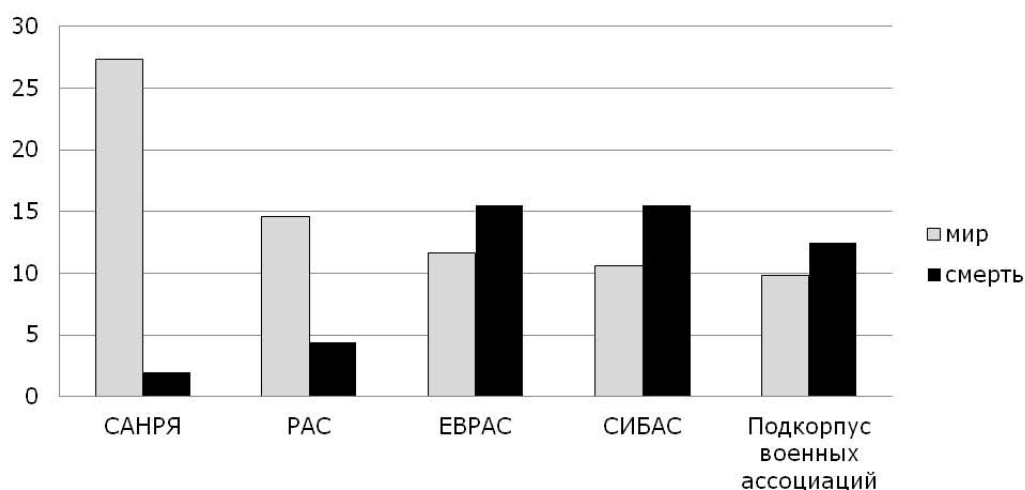


Рис. 1. Частотность реакций *мир* и *смерть* на стимул ВОЙНА в русских ассоциативных базах

Fig. 1. Frequency of reactions *peace* and *death* given to stimulus WAR in Russian associative bases

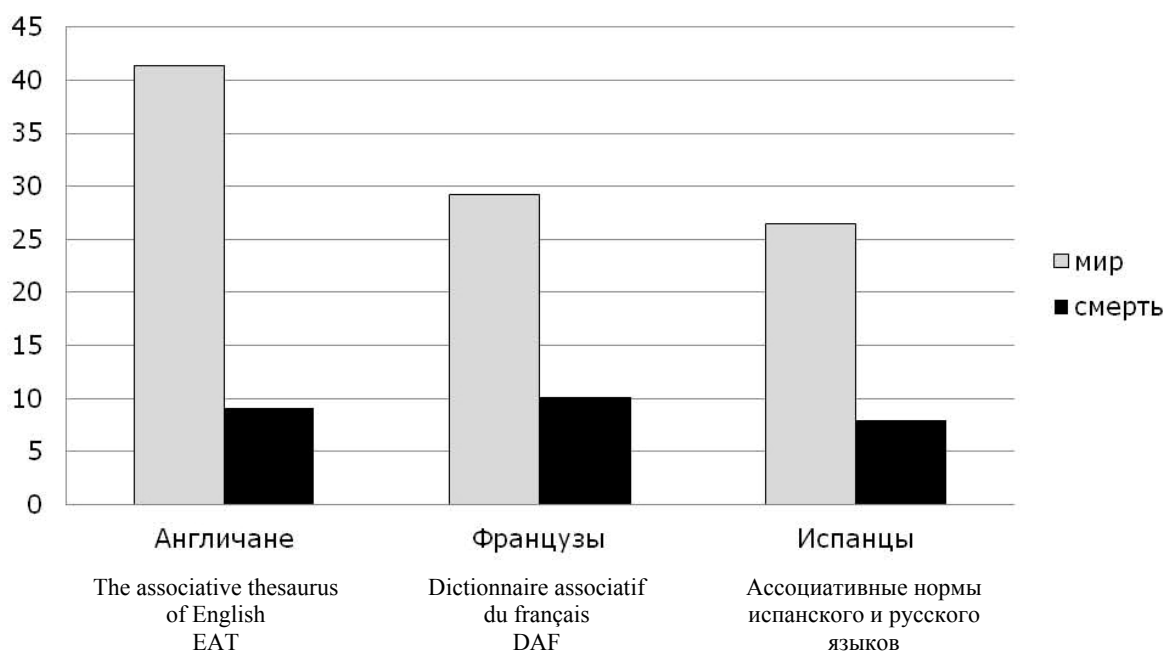


Рис. 2. Частотность реакций *мир* и *смерть* на стимул *ВОЙНА* в иноязычных ассоциативных базах

Fig. 2. Frequency of reactions *peace* and *death* given to stimulus *war* in non-Russian associative bases

Этимологический анализ членов дихотомии «война – мир»

В отличие от других оппозиций, выделяемых в ядре русского языкового сознания («плохо – хорошо», «нет – есть», «хороший – плохой», «жизнь – смерть», «мужчина – женщина», «большой – маленький», «день – ночь» и др.), дихотомия «война – мир» характеризуется изменчивостью смысловой наполненности ее членов.

Этимологический анализ слова «мир»

Так, понятие «мир», согласно Ю. С. Степанову, в древности было «довольно далеко от современного понимания мира в дипломатическом смысле как отсутствия войны или прекращения войны; оно относилось, прежде всего, к “ладу” и “добрососедству” между “своими”» (Степанов, 2004. С. 88).

В современном русском языке существует два слова *мир*, считающихся омонимами: 1) *мир* – вселенная, система мироздания как целое; 2) *мир* – согласные отношения, спокойствие, отсутствие войны, ссоры. Ю. С. Степанов считает первичным фактом, относящимся к константам культуры, совмещение двух этих значений в одном слове. При этом в русской культуре такое совмещение сохраняется в большей степени, чем в других европейских культурах, где эти смысловые ряды разделены более резко. «Ядром этого соединения является концепт “Свой” в противопоставлении “Чужим, чужому”. “Мир” в древнейших культурах индоевропейцев – это то место, где живут люди “моего племени”, “моего рода”, “мы”, место, хорошо обжитое, хорошо устроенное, где господствует “порядок”, “согласие между людьми”» (Степанов, 2004. С. 86–87).

Эти рассуждения подкрепляются этимологическим анализом слова *мир*. Согласно Ю. С. Степанову, это слово происходит от праиндоевропейского корня **mei-* «укреплять». Зубные расширители этого корня привели к возникновению слова *место* (из **moit-to*), которое первоначально означало «укрепленное и освященное место». Сонантные расширители

дали рождение словам *мир* (из **mei-* + *-r-*) и *мил*, *милый* (из **mei-* + *-l-*). Таким образом, этимологический анализ показывает, что ядро понятия «мир» включает в себя понятия «свои», «обитатели своей земли», «люди, связанные родственными, договорными или дружескими отношениями» (Степанов, 2004. С. 89).

Аналогичные понятия послужили основой латинского слова *pax*, которое означает не только «мир, как отсутствие войны», но также «спокойствие, мирный договор, милость (богов)» [Дворецкий, 2003. С. 559]. Слово *pax* происходит от праиндоевропейского корня **pak-* «укреплять, скреплять», изначально вбиванием кола, сваи; обнесением забором, оградой; в дальнейшем – договором, соглашением. Близкое отношение между понятиями «мир, отсутствие вражды» и «мир, хорошо обустроенное пространство» воплотилось в первые века нашей эры в термине *Pax romana* «римский мир», указывающий одновременно на территориальное образование (Рим и колонии) и на основу его стабильного существования (клятвы мира, которые приносили покоренные народы Риму). Во французском языке слово *pax* приобрело вид *paix*, в испанском – *paz*.

В середине XII в. слово *peis* (норманнский вариант старофранцузского слова *pais*) пришло в английский язык со значением «гражданское спокойствие», «мир в душе», вытеснив исконно английские *frið*, *sib*, *griþ*, *seht* (OED). В эпоху Великого сдвига гласных в английском языке (XV–XVI вв.) корневой гласный, подвергаясь общему процессу сужения звуков, стал произноситься на «ступеньку» выше с учетом перемещения языка в вертикальной плоскости: [ɛ:] → [e:] → [i:] [Шапошникова, 2009. С. 248]. В результате, к XVI в. это слово приобрело современное написание.

И русское слово *мир*, и латинское слово *pax* широко использовались в ареале христианской культуры в качестве элементов устойчивых словосочетаний (*Мир вам! С миром! Мир вашему дому! Pax vobis(cum)! Requiescat in pace*). В христианском понимании мир «есть нормальное благодатное состояние человеческой души, освобожденной от рабства греху, <...> состояние примиренности с Богом», которое психологически выражается во внутреннем порядке души (Основы социальной концепции Русской Православной Церкви, 2000). Теоретически, такое состояние человек может испытывать в любых условиях, даже в обстановке интенсивных боевых действий. Отсюда следует, что христианское понимание мира имеет лишь частичное отношение к дихотомии «война – мир». Слова *мир* и *pax* использовались в качестве перевода древнегреческого слова εἰρήνη, этимологию которого исследователи возводят к глаголу εἶρω «сплести, соединять» (Strong), что также указывает на первичность смыслового ряда «согласие, договор». Слово εἰρήνη применялось для перевода еврейского слова *шалом*, трехбуквенная основа которого *ш-л-м* (шин-ламед-мем) «означает результат действия совершенствования, т. е. “совершенство, завершенность”». Это не просто мир как отсутствие войны, а мир как достижение совершенства, гармонии, цельности» [Лемельман, 2000. С. 38]. Таким образом, первичность смысла «согласие, договор, цельность, гармония», вкладываемое в слово *мир*, наблюдается не только в индоевропейской семье. Арабский эквивалент еврейского корня – سلم «С-Л-М» (присутствующий в слове *ислам*) – используется не столько в контексте войны и мира, сколько в значении «спокойствия, безопасности, подчинения себя воле Всевышнего» [Churchill, Lewis, 2008. P. 8].

Этимологический анализ свидетельствует в пользу того, что слово *мир* в различных языках относительно недавно стало употребляться в основном современном значении «отсутствие войны».

Этимологический анализ слова «война»

Слово *война* происходит от праславянского **vojьna* – соединения корня *вой* «воин» и суффикса *-ьна* (> *на*). Исследователи связывают праславянское слово с праиндоевропейским корнем **ueia-*, **uei-* «стремиться, помогать, гнаться». Эти смыслы засвидетельствованы в родственных языках: литовском *vejti*, *vyti* «гнать(ся), преследовать», санскрите *vēti* «находится позади, преследует, ведет (войско), гонит, наставляет», авестийском *vayeiti* «гонит, преследует», латинском *venor*, *-āri* «охотиться», древнеисландском *veiðr* «охота», древне-

верхненемецком *weida* «охота», греческом *ἔμαι* «стремлюсь, желаю», ирландском *fiad* «дичь» (Фасмер, 2009. С. 334–335; Преображенский, 2010. С. 90–91).

Английское слово *war* заимствовано из норманнского варианта старофранцузского языка (в древнеанглийском использовались слова *wig*, *guð*, *heado*, *hild*, *gewin*). Норманнское слово *werre* «война» имело старофранцузскую параллель *guerre* «трудность, спор, соперничество, бой, война». Во французский язык (как и в испанский) это слово пришло из франкского **werra*. Романские народы стали использовать германское слово вместо латинского *bellum* «война», возможно, в стремлении избавиться от омонимии со словом *bello* «красивый». Франкское слово связано с прагерманским **werz-a-* (от него произошло современное немецкое *verwirren* «запутывать, сбивать с толку»), которое, в свою очередь, восходит к праиндоевропейскому корню **wers-* «запутывать, смешивать» (OED).

Этимологический анализ показывает, что в романских и германских языках понятие «война» изначально связано со смысловым рядом «путаница, смешение». В то же время противоположные смыслы «гармонии, целостности» этимологически наполняют понятие «мир». В славянских, балтийских, индоиранских языках этимологически понятие «война» слабее коррелирует с понятием «мир», так как наполнено смыслами «охота, преследование». Возможно, в этом кроется одна из причин того, что дихотомия «мир – война» не проявляет такой прочной устойчивости в русском языковом сознании, как это наблюдается в языковом сознании европейцев, где рассматриваемая оппозиция неизменно сохраняет высокую степень актуальности на протяжении нескольких лет.

Историческому изменению подверглась не столько семантическая наполненность слова *война*, сколько отношение к выражаемому этим словом явлению. «Отношение к войне как к занятию, достойному настоящего мужчины, было присуще большинству обществ, которые в процессе своего исторического развития проходили стадию родоплеменного строя и военной демократии» [Басов, 2005. С. 92]. Широкое распространение представлений о войне как о зле, равно как и ценность мира, актуализировались лишь в XX в.

Ассоциативные эксперименты САНРЯ и РАС проводились на фоне официальной советской риторики борьбы за мир, а затем обнадеживающего окончания холодной войны. Однако последующие исторические события показали, что перспектива мира без войны по-прежнему остается отдаленной. Возможно, получившее в широких слоях населения ощущение иллюзорности общего мира повлияло на то, что в сознании носителей русского языка война стала чаще восприниматься в отрыве от своего антонима и в большей степени ассоциироваться со злом, которое она несет, – смертью. В ассоциативно-вербальной сети ЕВРАС, СИБАС и ПВАС это отразилось в ослаблении интенсивности реакции *мир* и усилении частотности реакции *смерть*.

Заключение

Наше исследование показало, что, хотя слово *война* относится к базовому лексическому фонду, в сознании средней языковой личности это понятие не является актуальным. Исключение составляют военные профессионалы, для которых слово *война* относится к ядру ассоциативных доминант, т. е. образует большое количество связей. В целом, анализ ассоциативно-вербальной сети не выявляет каких-то патологически воинственных настроений. Война воспринимается носителями русского языка как негативное явление, что вселяет уверенность в преобладании миролюбивых черт в национальном характере.

Список литературы / References

- Басов И. И. Западноевропейское рыцарство XII–XV вв. в евразийском историко-культурном контексте: этика противоборства (опыт сравнительно-исторического исследования): Дис. ... канд. ист. наук: 07.00.03. Армавир, 2005. 259 с.
Basov, I. I. Western European chivalry in XII–XV centuries in the Eurasian historical and cultural context: the ethics of confrontation (the experience of comparative historical research). Cand. Phil. Sci. Diss. Armavir, 2005, 259 p. (in Russ.)

- Бердяев Н. А.** Философия неравенства / Сост. и отв. ред. О. А. Платонов. М.: Ин-т рус. цивилизации, 2012. 624 с.
Berdyaev, N. A. Philosophy of inequality. Moscow, Institut russkoj tsivilizatsii, 2012, 624 p. (in Russ.)
- Гегель Г. В. Ф.** О научных способах исследования естественного права, его месте в практической философии и его отношении к науке о позитивном праве // Политические произведения. М.: Наука, 1978. С. 185–275.
Hegel, G. W. F. On the Scientific Ways of Treating Natural Law, on its Place in Practical Philosophy, and its Relation to the Positive Sciences of Right. In: Politicheskie proizvedeniya. Moscow, Nauka, 1978. P. 185–275. (in Russ.)
- Достоевский Ф. М.** Полн. СОБР. соч. и писем: В 30 т. / АН СССР, ИРЛИ (Пушкинский Дом); редкол.: В. Г. Базанов (отв. ред.) и др. Л.: Наука. Ленингр. отд-ние, 1981. Т. 22: Дневник писателя за 1876 год. Январь–апрель. 407 с.
Dostoevskiy, F. M. Complete works and letters. Leningrad, Nauka, 1981, vol. 22, 407 p. (in Russ.)
- Залевская А. А.** О комплексном подходе к исследованию закономерностей функционирования языкового механизма человека // Психолингвистические исследования в области лексики и фонетики. Калинин: Калинин. гос. ун-т, 1981. С. 28–44.
Zalevskaya, A. A. On an integrated approach to the study of the laws governing the functioning of the human language mechanism. In: Psiholingvisticheskie issledovaniya v oblasti leksiki i fonetiki. Kalinin, Kalinin State University, 1981. P. 28–44. (in Russ.)
- Залевская А. А.** Слово в лексиконе человека: психолингвистическое исследование. Воронеж: Изд-во Воронеж. ун-та, 1990. 208 с.
Zalevskaya, A. A. Word in the human lexicon: psycholinguistic research. Voronezh, Voronezh University Publ., 1990, 208 p. (in Russ.)
- Кант И.** Соч.: В 6 т. М.: Мысль, 1966. Т. 6. 743 с.
Kant, I. Works. Moscow, Mysl', 1966, vol. 6, 743 p. (in Russ.)
- Караулов Ю. Н.** Активная грамматика и ассоциативно-вербальная сеть. М.: ИРЯ РАН, 1999. 180 с.
Karaulov, Yu. N. Active grammar and associative-verbal network. Moscow, Institute of Russian Language RAS, 1999, 180 p. (in Russ.)
- Караулов Ю. Н.** Показатели национального менталитета в ассоциативно-вербальной сети // Языковое сознание и образ мира: Сб. ст. / Отв. ред. Н. В. Уфимцева. М.: ИЯ РАН, 2000. С. 191–206.
Karaulov, Yu. N. Indicators of national mentality in the associative-verbal network. In: Yazykovoe soznanie i obraz mira. Ed. by N. V. Ufimtseva. Moscow, Institute of Linguistics RAS, 2000, p. 191–206. (in Russ.)
- Лемельман Й.** Практическая грамматика иврита для всех. Хеврон Израиль: Ахаз, 2000. Ч. 1. 128 с.
Lemelman, J. Practical grammar of Hebrew for all. Hebron, Israel, Ahaz, 2000, part 1, 128 p. (in Russ.)
- Леонтьев А. А.** Языковое сознание и образ мира // Язык и сознание: парадоксальная рациональность. М.: ИЯ РАН, 1993. С. 16–21.
Leontiev, A. A. Language consciousness and the image of the world. In: Yazyk i soznanie: paradoksal'naya ratsional'nost'. Moscow, Institute of Linguistics RAS, 1993, p. 16–21. (in Russ.)
- Мальтус Т. Р.** Опыт о законе народонаселения. Петрозаводск: Петроком, 1993. 136 с.
Malthus, T. R. An Essay on the Principle of Population. Petrozavodsk, Petrokom, 1993, 136 p. (in Russ.)
- Ницше Ф.** Так говорил Заратустра // Ницше Ф. Соч.: В 2 т. М.: Мысль, 1990. Т. 2. С. 5–237.
Nietzsche, F. Thus Spoke Zarathustra. Moscow, Mysl', 1990, p. 5–237. (in Russ.)
- Орехов А. М.** «Вечный мир» или «вечная война»? (И. Кант vs К. Клаузевиц) // Пространство и время. 2014. № 3 (17). С. 62–66.
Orehov, A. M. «Eternal peace» or «eternal war»? (I. Kant versus C. Clausewitz). *Prostranstvo i vremya*, 2014, no. 3 (17), p. 62–66. (in Russ.)
- Соловьев В. С.** Спор о справедливости: Сочинения. М.: ЭКСМО-Пресс, Харьков: Фолио, 1999. 190 с.
Soloviev, V. S. The dispute about justice. Moscow, EKSMO-Press, Kharkov, Folio, 1999, 190 p. (in Russ.)
- Тарасов Е. Ф.** Актуальные проблемы анализа языкового сознания // Языковое сознание и образ мира: Сб. ст. / Отв. ред. Н. В. Уфимцева. М.: ИЯ РАН, 2000. С. 24–32.
Tarasov E. F. Actual problems of the analysis of language consciousness. In: Yazykovoe soznanie i obraz mira. Ed. by N. V. Ufimtseva. Moscow, Institute of Linguistics RAS, 2000, p. 24–32. (in Russ.)

- Трактаты о вечном мире / Сост. И. С. Андреева, А. В. Гулыга. СПб.: Алетея, 2003. 398 с.
Traktaty o vechnom mire [Treatises on the Eternal Peace]. Comp. by I. S. Andreeva, A. V. Gulyga. St. Petersburg, Aleteya, 2003, 398 p. (in Russ.)
- Уфимцева Н. В.** Языковое сознание: динамика и вариативность. М.: Ин-т языкознания РАН, 2011. 252 с.
Ufimtseva, N. V. Language consciousness: dynamics and variability. Moscow, Institute of Linguistics RAS, 2000, 252 p. (in Russ.)
- Уфимцева Н. В.** Значение слова как отражение социокультурной реальности // Вопросы психолингвистики. 2015. № 3 (25). С. 83–92.
Ufimtseva, N. V. The meaning of the word as a reflection of sociocultural reality. *Voprosy psiholingvistiki*, 2015, no. 3 (25), p. 83–92. (in Russ.)
- Фрейд З.** Неизбежна ли война? // По ту сторону принципа удовольствия. М.: Прогресс, 1992. С. 325–337.
Freud, S. Why war? In: *Beyond the Pleasure Principle*. Moscow, Progress, 1992, p. 325–337. (in Russ.)
- Фромм Э.** Анатомия человеческой деструктивности. М.: Республика, 1994. 447 с.
Fromm, E. The Anatomy of Human Destructiveness. Moscow, Respublika, 1994, 447 p. (in Russ.)
- Черкасова Г. А.** Динамическая ассоциативно-вербальная модель языкового сознания носителей русского языка // Новые информационные технологии в автоматизированных системах. 2014. № 17. С. 405–414.
Cherkasova, G. A. Dynamic associative-verbal model of linguistic consciousness of native speakers of the Russian language. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2014, no. 17, p. 405–414. (in Russ.)
- Шапошникова И. В.** История английского языка: Учеб. пособие. Новосибирск: Новосиб. гос. ун-т, 2009. 506 с.
Shaposhnikova, I. V. History of the English language. Manual. Novosibirsk, Novosibirsk State University, 2009, 506 p. (in Russ.)
- Шапошникова И. В.** Мотивационная база реформы образования и профессиональная идентификация русской языковой личности // Вопросы психолингвистики. 2017. № 4 (34). С. 124–145.
Shaposhnikova, I. V. Motivational basis of the education reform and professional identification of the Russian language personality. *Voprosy psiholingvistiki*, 2017, no. 4 (34), p. 124–145. (in Russ.)
- Churchill, B. E., Lewis, B. E.** Islam: The Religion and the People. FT Press, 2008, 256 p.
- Fearon, J. D.** Rationalist Explanations for War. *International Organization*, 1995, no. 49 (3), p. 379–414.
- Fry, D. P., Söderberg, P.** Lethal Aggression in Mobile Forager Bands and Implications for the Origins of War. *Science*, 2013, vol. 341, p. 270–273.
- Fuller, G.** The Youth Factor: The New Demographics of the Middle East and the Implications for U.S. Policy. Washington, D.C., The Saban Center for Middle East Policy at the Brookings Institution, 2003, 42 p.
- Fuller, G.** The Youth Crisis in Middle Eastern Society. Clinton Township, Michigan, Institute for Social Policy and Understanding, 2004, 13 p.
- James, P., Friedman, J.** Globalization and Violence. London, Sage Publications, 2006, vol. 3: Globalizing War and Intervention, 359 p.
- Joost, A. M., Meerloo, M. D.** The Rape of the Mind: The Psychology of Thought Control, Menticide, and Brainwashing. Progressive Press, 2009, 326 p.
- Šmihula, D.** The Use of Force in International Relations. Bratislava, Veda, 2013, 309 p.
- Smalley, R. E.** Future Global Energy Prosperity: The Terawatt Challenge. *MRS Bulletin*, 2005, vol. 30, p. 412–417.
- Turnbull, C.** The Forest People. New York, Touchstone Books, 1987, 320 p.

Список источников / List of Sources

- Дворецкий И. Х.** Латинско-русский словарь. М.: Рус. яз.-Медиа, 2003. 846 с.
Dvoretsky, I. Kh. Latin-Russian Dictionary. Moscow, Rus.yaz.-Media, 2003, 846 p. (in Russ.)
- ЕВРАС – **Уфимцева Н. В., Черкасова Г. А.** Русский региональный ассоциативный словарь-тезаурус ЕВРАС. URL: <http://iling-ran.ru/main/publications/evras> (дата обращения 24.08.2018).

- Russian regional associative dictionary-thesaurus EVRAS, 2014. G. A. Cherkasova, N. V. Ufimtseva (Comp.). URL: <http://iling-ran.ru/main/publications/evras> (accessed 24.08.2018). (in Russ.)
- Основы социальной концепции Русской Православной Церкви. VIII. Война и мир. М., 2000. URL: <https://mospat.ru/ru/documents/social-concepts/viii/> (дата обращения 24.08.2018).
Fundamentals of the social concept of the Russian Orthodox Church. VIII. War and peace. Moscow, 2000. URL: <https://mospat.ru/ru/documents/social-concepts/viii/> (accessed 24.08.2018). (in Russ.)
- Преображенский А. Г.** Этимологический словарь русского языка: А-О. 2-е изд. М.: ЛКИ, 2010. 720 с.
Preobrazhenskiy, A. G. Etymological dictionary of the Russian language: A-O. Moscow, LKI, 2010, 720 p. (in Russ.)
- РАС – Русский ассоциативный словарь / Сост. Ю. Н. Караулов, Ю. А. Сорокин, Е. Ф. Тарасов, Н. В. Уфимцева, Г. А. Черкасова. М.: Астрель, 1994, Т. 1; 1998. Т. 2. URL: <http://www.thesaurus.ru/dict/dict.php> (дата обращения 24.08.2018).
Russian associative dictionary. Yu. N. Karaulov, Yu. A. Sorokin, E. F. Tarasov, N. V. Ufimceva, G. A. Cherkasova. Moscow, Astrel', 1994, 1998. URL: <http://www.thesaurus.ru/dict/dict.php> (accessed 24.08.2018). (in Russ.)
- Санчес Пуиг М., Караулов Ю. Н., Черкасова Г. А.** Ассоциативные нормы испанского и русского языков. Москва, Мадрид: Азбуковник, 2001. 496 с.
Sanchez Puig, M., Karaulov, Yu. N., Cherkasova, G. A. Associative norms of Spanish and Russian languages. Moscow, Madrid, Azbukovnik, 2001. 496 p. (in Russ. and Spanish)
- САНРЯ – Словарь ассоциативных норм русского языка / Под ред. А. А. Леонтьева. М.: МГУ, 1977. 192 с.
Dictionary of associative norms of the Russian language. Ed. by A. A. Leontiev. Moscow, Moscow State University, 1977. 192 p. (in Russ.)
- СИБАС – Русская региональная ассоциативная база данных (2008–2015) / Сост. И. В. Шапошникова, А. А. Романенко. URL: <http://adictru.nsu.ru> (дата обращения 24.08.2018).
SIBAS – Russian regional associative database (2008–2015). I. V. Shaposhnikova, A. A. Romanenko (Comp.). URL: <http://adictru.nsu.ru> (accessed: 24.08.2018). (in Russ.)
- Степанов Ю. С.** Константы: Словарь русской культуры. 3-е изд., испр. и доп. М.: Академический проект, 2004. 992 с.
Stepanov, Yu. S. Constants: Dictionary of Russian culture. Moscow, Akademicheskij proekt, 2004. 992 p. (in Russ.)
- Фасмер М.** Этимологический словарь русского языка. В 4 т. Т. 1: А–Д: ок. 4000 слов / Пер. с нем. и доп. О. Н. Трубачева. 4-е изд., стереотип. М.: Астрель: АСТ, 2009. 588 с.
Fasmer, M. Etymological dictionary of the Russian language. In 4 vols. Moscow, Astrel': AST, 2009, vol. 1, 588 p. (in Russ.)
- DAF – Dictionnaire associatif du français (2008–2010). Comp. by M. Debrenn, N. V. Ufimtseva, A. A. Romanenko. URL: <http://dictaverf.nsu.ru> (accessed: 24.08.2018).
- EAT – **Kiss, G., Armstrong, C., Milroy, R.** The associative thesaurus of English. Edinburgh: Univ. of Edinb., MRC Speech and Communication Unit, 1972. URL: <http://rali.iro.umontreal.ca/rali/?q=en/Textual%20Resources/EAT> (accessed: 24.08.2018).
- OED – Online Etymology Dictionary. URL: <https://www.etymonline.com> (accessed: 24.08.2018).
- Strong, J.** The Exhaustive Concordance of the Bible, 1890. URL: <http://www.sacrednamebible.com/kjvstrongs/index2.htm> (accessed: 24.08.2018).

Материал поступил в редколлегию
Date of submission
03.09.2018

Сведения об авторе / Information about the Author

Кафтанов Роман Анатольевич, аспирант Института филологии СО РАН (ул. Николаева, 8, Новосибирск, 630090, Россия)

Roman A. Kaftanov, post-graduate student, Institute of Philology SB RAS (8 Nikolaev Str., Novosibirsk, 630090, Russian Federation)

rk76@ngs.ru

Правила оформления текста рукописи

Авторы представляют статьи на русском языке объемом от 0,5 печатного листа (20 тыс. знаков, шрифт 14, межстрочный интервал 1,5) до 1 авторского листа (40 тыс. знаков) включая иллюстрации (1 иллюстрация форматом 190×270 мм = $\frac{1}{6}$ авторского листа, или 6,7 тыс. знаков). Публикации, превышающие указанный объем, допускаются к рассмотрению только после индивидуального согласования с ответственным редактором.

Требования к оформлению основного текста и иллюстративных материалов

К рукописи необходимо приложить сведения об ученой степени, ученом звании, должности и месте работы, а также контактный телефон, электронный и почтовый адрес автора.

Обязательным требованием является наличие *резюме статьи* на русском и английском языках, а также авторский перевод названия статьи на английский язык, ключевые слова на двух языках, индекс УДК (Универсальной десятичной классификации).

Образец оформления статьи

УДК 81 + 811.131.1 + 811.161.1

Русская и итальянская абстрактная адъективная метафоризация

М. А. Кузьмина

*Новосибирский государственный университет
Новосибирск, Россия*

Аннотация

Статья посвящена сопоставительному семантическому анализу направлений русской и итальянской адъективной метафоризации на основе номинаций «абстрактных» признаков объектов. Выявляются наиболее регулярные направления метафорического переноса, отмечаются общие и специфические черты метафорообразования в двух языках.

Ключевые слова

адъективная метафора, метафоризатор, направление метафоризации

Благодарности

Исследование проведено за счет гранта Российского научного фонда (проект № 14-50-00036)

The Russian and Italian Abstract Adjectival Metaphorization

Maria A. Kuzmina

*Novosibirsk State University
Novosibirsk, Russian Federation*

Abstract

The paper presents semantic comparison and analysis of Russian and Italian adjectival metaphorization tenors, the vehicles and metaphors both being abstract nominations. By revealing of the most regular tenors, common and specific metaphorization features are demonstrated.

Keywords

adjectival metaphor, vehicle, metaphorization tenor

Acknowledgements

The work was supported by the Russian Science Foundation, project 14-50-00036

Основной текст статьи

Список литературы / References

Список словарей / List of Dictionaries

Список источников / List of Sources

*Материал поступил в редколлегию
Date of submission
05.12.2018*

Сведения об авторах / Information about the Authors

Подпись автора (авторов)

Библиографические ссылки: в тексте в квадратных скобках указывается фамилия автора, год издания. Например: [Розен, 1969. С. 5]. В конце статьи помещается список литературы в алфавитном порядке без нумерации. Библиографическое описание публикации включает: фамилии и инициалы авторов (всех, независимо от их числа), полное название работы, а также издания, в котором опубликована (для статей), город, название издательства, год издания, том (для многотомных изданий), номер, выпуск (для периодических изданий), объем публикации (количество страниц – для монографии, первая и последняя страницы – для статьи). Ссылки на архивные документы оформляются в виде сноски (текст сноски располагается внизу страницы). Библиографические источники не нумеруются!

Образцы составления библиографического описания

Авторская монография:

Ильиш Б. А. История английского языка. М.: Лит. на иностр. яз., 1958. 366 с.

Коллективная монография

(все авторы должны быть упомянуты):

Суперанская А. В., Сталтмане В. Э., Подольская Н. В., Султанов А. Х. Теория и методика ономастических исследований: Моногр. М.: Наука, 1986. 298 с.

Статья в сборнике:

Черкасова Г. А. Русский ассоциативный тезаурус: компьютерная технология создания и издания // Этнокультурная специфика языкового сознания: Сб. науч. ст. / Под ред. Н. В. Уфимцевой. М., 2003. С. 181–190.

Статья в ученых записках (ученых трудах):

Скрипка А. С. К датировке некоторых типов сарматского оружия // Учен. зап. Ленингр. пед. ин-та. 1977. Вып. 2. С. 60–77.

Статья в журнале:

Кириллов Д. А. Образ выборов в языковом сознании российской молодежи. Опыт сравнительного исследования на материале РАС и свободного ассоциативного эксперимента // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2008. Т. 6, № 2. С. 17–24.

Автореферат:

Яньшин П. В. Психосемантический анализ категоризации цвета в структуре сознания субъекта: Автореф. дис. ... д-ра психол. наук. М., 2001. 42 с.

Рецензия:

Панин Л. Г. [Рецензия] // Сибирский филологический журнал. 2003. № 3–4. С. 245–247. Рец. на кн.: Турбин Г. А., Шулежкова С. Г. Старославянский язык: Учеб. пособие. М.: Изд-во Моск. гос. ун-та, 2002. 145 с.

При подготовке иллюстративного материала просим учесть, что рисунки следует предоставлять в форматах .gif, .jpg, .tif отдельными файлами.

Допускается создание таблиц и диаграмм в WinWord и Excel (обязательно приложить исходный файл в формате .xls), обязательно прилагать файлы используемых (нестандартных) шрифтов (.ttf), кегль шрифта в надписях не должен быть меньше 9. Максимальное поле изображения 190 × 270 мм.

Все вопросы, связанные с изменением и уточнением текста в процессе редакторской правки, должны сниматься авторами в ходе переписки по электронной почте в строго определенных для этого редколлегией сроки. Нарушение сроков ведет к отказу в опубликовании статьи. *Переписка традиционной почтой не осуществляется.*

Требования к содержанию публикуемых материалов

Требования к теме исследования, заявленной в статье. Как правило, выбор аспирантом темы публикации соотносится с темой его кандидатского диссертационного исследования. Прежде чем приступить к написанию статьи, необходимо выяснить, является ли разрабатываемая аспирантом тема актуальной и новой для филологических наук. В решении этого вопроса аспиранту не следует полагаться только на мнение научного руководителя и сотрудников своего НИИ или вуза. Необходимо тщательно проверить самому, какое количество работ уже было выполнено по данной теме отечественными и зарубежными исследователями. Для уточнения ситуации с актуальностью и новизной темы нужно обратиться к авторитетным библиографическим ресурсам (rsl.ru, rnl.ru, elibrary.ru, loc.gov и т. д.) и с помощью всех ключевых слов выявить круг уже выполненных исследований. Если, например, соматическим компонентам в составе английских фразеологизмов, в том числе социолектных, уже посвящено более 150 структурно-семантических исследований, то соискателям ученых степеней стоит переключиться на другие темы.

Требования к объектной и предметной новизне исследования. Новизна лингвистического исследования может быть объектной и предметной.

Требования к описанию объекта исследования. Автор статьи в результате выполненного им исследования должен иметь наиболее полное и системное представление об изученном им объекте. Границы объекта должны быть предельно четко очерчены в статье. Например, если речь идет о языковых единицах, то должно быть указано, к какому языку или языкам какого исторического периода или периодов относятся данные единицы, каково их общее количество, из каких именно источников, письменных или устных, они были взяты и каким методом были собраны. При отсутствии у объекта точных квалификативных характеристик выводы исследования нельзя считать репрезентативными.

Требования к характеристике предмета исследования. Об объекте, даже хорошо изученном в ряде направлений, можно получить ценную новую информацию при новом подходе к его изучению. Из содержания статьи специалистам должно быть ясно, в чем именно состоит предметная новизна выполненной работы.

Требования к формулировке названий статьи. Формулировка названия статьи должна четко отражать объектные и предметные характеристики исследования. Слишком широкие формулировки, не соответствующие содержанию работы (например «Морфологические категории: коммуникативный аспект интерпретации»), вводят читателей в заблуждение.

Требования к обоснованности и достоверности научных положений и выводов, представленных в статье. Степень обоснованности и достоверности научных положений и выводов любой лингвистической работы обуславливается репрезентативностью исследовательской картотеки и применением необходимых методов исследования. В настоящее время репрезентативность большинства лингвистических исследовательских картотек должна подкрепляться компьютерными корпусными данными (коллекции ссылок на лингвистические корпусы можно посмотреть, например, на сайте <http://www.uow.edu.au/~dlee/CBLLinks.htm>). Перечни методов лингвистических исследований представлены в учебной и специальной литературе.

Требования к соотношенности полученных в исследовании новых выводов с целью и задачами, заявленными в статье. Последовательно перечисленные в конце статьи выводы исследования (в порядке убывания их значимости) должны быть скоррелированы с заявленными в начале статьи целью и задачами.

Требования к точному и последовательному использованию терминов в научном тексте. Системность – отличительная черта научного знания. Субъективный эссеизм, компилятивность, слишком вольное обращение с терминами или полное игнорирование специальной терминологии свидетельствуют о непрофессионализме автора. Все необходимые для изложения термины должны быть системно сгармонизированы и употреблены автором статьи только однозначно. В статьях, посвященных металингвистическим проблемам (истории лингвистической терминологии, неоднозначности толкования терминов в различных научных школах, фиксации терминов в специальных словарях и т. д.), должны быть представлены

только новые для отечественных специалистов сведения. Введение оригинальной авторской терминологии должно быть объективно обосновано.

Требования к использованию цитат. Все приведенные в тексте статьи явные и скрытые цитаты должны иметь ссылки. Реферативность изложения, обилие цитат и другие признаки «вторичного текста» не позволяют рассматривать некоторые статьи как оригинальные и самостоятельные произведения научного стиля.

Требования к отражению в статье информации о личном участии автора в исследовании. Из содержания статьи должно быть понятно, в чем именно состоит личное участие автора или авторского коллектива в получении научных результатов. В теоретических исследованиях автор (или авторы) не должен «компоновать» конспекты чужих трудов в соответствии с замыслом своей работы, а должен выходить на качественно новый уровень самостоятельной рефлексии.

Требования к отражению в статье прикладного значения полученных результатов исследования. В тексте статьи должны быть указаны сферы использования результатов, полученных автором.

Соответствие текста статьи требованиям к научному стилю и оформлению. Текст статьи должен соответствовать требованиям, предъявляемым к текстам этого жанра. Информацию о жанровой специфике научных статей можно почерпнуть из учебных пособий по научному стилю речи. Требования к оформлению статей в журнале «Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация» публикуются на соответствующих страницах в каждом номере журнала. Просим авторов обращать внимание на объем присылаемых рукописей (не менее 20 000 знаков) и недопустимость тезисного изложения материала.

Передавая рукопись статьи (произведение) в редколлегию журнала, автор тем самым предоставляет ей право использования передаваемых материалов в составе журнала следующими способами: обнародование, воспроизведение, распространение, доведение произведения до всеобщего сведения путем размещения в сети Интернет, публичный показ, а также перевод на иностранные языки, включая те же действия относительно переведенного произведения, на территории всех государств, где произведение подлежит правовой охране.

Доставка материалов

Представляемые в редакцию материалы можно передать лично (комната 1269, новый корпус НГУ) или переслать по электронной почте.

Адрес редакционной коллегии
серии «Лингвистика и межкультурная коммуникация»

Кафедра истории и типологии языков и культур
ул. Пирогова, 1, Новосибирск, 630090, Россия

Тел.: (383) 363 42 23
E-mail: lingua@vestnik.nsu.ru

Журнал распространяется по подписке,
подписной индекс 18285 в каталоге ОАО «Роспечать»