

© 2011 г. А.А. КРЕТОВ, И.А. МЕРКУЛОВА, В.Т. ТИТОВ*

ПРОБЛЕМЫ КВАНТИТАТИВНОЙ ЛЕКСИКОЛОГИИ СЛАВЯНСКИХ ЯЗЫКОВ

Статья посвящена выявлению существенных черт лексико-семантической системы славянских языков. Авторы исследуют функциональную, синтагматическую, парадигматическую и эпидигматическую стратификацию славянской лексики, производят сравнения славянских словарей по доминанте парадигматического ряда и составляют карту лексико-семантического пространства славянских языков.

Бурное развитие типологии началось с фонетического уровня, распространилось на морфологический и синтаксический и почти не затронуло лексико-семантического уровня. В этом отношении показателен проект «Языки мира», выполняемый Институтом языкознания РАН: сведения о лексике в описании языка скудны и ориентированы на внешнюю лингвистику: исконная/заимствованная лексика, источники заимствованной лексики, способы образования новых слов (это уже не столько лексикология, сколько словообразование), географическая, социальная и функциональная стратификация лексики.

Начиная с середины XX века проблемы лексико-семантической типологии волнуют славистов и сочувствующих: Н.И. Толстой [Толстой 1963; 1966], В.А. Москович [Москович 1965]; Б.Ю. Городецкий [Городецкий 1969], А. Гудавичюс [Гудавичюс 1985], А.К. Супрун [Супрун 1983; 1988], В.Г. Гак [Гак 1989], Х. Томмола [Томмола 2003], Е.В. Рахилина [Рахилина, Прокофьева 2004; Майсак, Рахилина 2007].

При этом основное внимание уделяется либо внешнелингвистическим характеристикам лексики (А.К. Супрун), либо фрагментам лексико-семантической системы (Н.И. Толстой, А. Гудавичюс, Е.В. Рахилина и И. Прокофьева, Т.А. Майсак и Е.В. Рахилина).

Лексический состав языка, как известно, измеряется сотнями тысяч единиц, поэтому представляется исследователю чем-то необозримым, трудно поддающимся систематизации и сопоставлению. Более того, высказывается мнение, согласно которому «...в сравнении с фонологией и грамматикой, лексика представляет собой более слабую и нечеткую, в большей мере вероятностную систему, значительно меньше структурированную. Поэтому, в отличие от фонетико-фонологической и грамматической типологии, в лексической типологии отсутствуют и, по-видимому, невозможны классификации языков» [Мечковская 2000: 106].

Практическим опровержением этого мнения является, в частности, книга А.В. Дыбо [Дыбо 1996] и монография [Майсак, Рахилина 2007]. Мнение, высказанное Н.Б. Мечковской, не учитывает иерархической и полевой организации словаря, в котором важнейшей является оппозиция ядро // периферия. Ядро лексики устойчиво и структурировано. Слаба (изменчива) и нечётка – периферия. Ядро мало по размеру, периферия – безгранична. Ядром владеет каждый индивид, периферией – все общество, обслуживаемое данным языком. Поэтому в лексической типологии классификации языков возможны, но в основу их должно быть положено ядро лексико-семантической системы. Таким

* Исследование выполнено при финансовой поддержке РГНФ проект № 07-04-00036а.

образом, важнейшим шагом на пути к лексико-семантической типологии является выделение ядер лексико-семантических систем языков мира. Ее решению помогает применение параметрического анализа лексики, подробно описанного в работах В.Т. Титова и примененного им к материалу романских языков [Титов 2002; 2005]. Есть все основания считать накопленный на романском материале опыт приложимым также и к славянским языкам. «Пилотный» проект такого исследования был представлен авторами в рамках тематического блока на XIV Международном съезде славистов (Охрид, Македония) [Кретов, Меркулова 2008].

Для получения параметрических характеристик славянской лексики необходимо было решить ряд сопутствующих задач:

- 1) создание электронных версий инославяно-русских словарей;
- 2) создание баз данных на основе полученных словарей;
- 3) получение количественной информации по каждому из параметров исследуемых языков;
- 4) применение методики параметрического анализа лексики к материалу каждого из исследуемых языков;
- 5) получение карты лексико-семантического пространства славянских языков.

Для того чтобы выявить типологически существенные черты лексико-семантической системы предполагается достаточным проанализировать тысячу наиболее «весомых» единиц лексико-семантической системы. Выделение этой «ядерной тысячи», или этого «тысячного ядра», и составляет задачу квантитативной лексикологии. Отобраны эти слова могут быть по ряду параметров: функциональному (употребительность) – оценивается по длине слова; синтагматическому (широкая сочетаемость) – оценивается по количеству фразеосочетаний со словом; эпидигматическому (многозначность) – оценивается по количеству значений у слова; парадигматическому (вхождение в многочисленные синонимические ряды) – оценивается по частоте слов, использованных в словарных дефинициях. Следует отметить, что парадигматический параметр даст вес не слов, а значений. Если у слова несколько значений, то в первую очередь выбирается основное значение, а во-вторую – значение с максимальным весом.

По каждому из параметров отбирается 1000 слов с максимальным весом, определяемым по формуле:

$$Pr_i = \frac{\sum r - R_{1-i}}{\sum r}, \quad (1)$$

где $\sum r$ – сумма единиц всех рангов, R_{1-i} – сумма единиц от первого до данного, а Pr_i – вес единиц i -го ранга.

Формула (1) позволяет определить вес каждой из единиц на всем корпусе словаря. Более того, вес лексико-семантической единицы по каждому из параметров зависит от количества в данном словаре единиц с такой же или лучшей параметрической характеристикой, что делает параметрический анализ системным в полном смысле этого слова, поскольку вес каждой единицы зависит от количества и качества всех остальных единиц словаря.

Параметрический вес для каждого из членов ядра определяется как сумма его весов по каждому из четырех указанных выше параметров. Выделение лексико-семантических ядер исследуемых языков проводилось на основе анализа «малых» двуязычных словарей, которые содержат основную коммуникативно значимую лексику. В качестве источников исследования был обработан целый корпус словарей (см. список этих словарей в конце этой статьи).

Данный список содержит не только двуязычные словари, так как опыт показал, что малые инославяно-русские словари, безусловно, оптимальны лишь по одному из параметров – функциональному. В остальных случаях при анализе столь близкородственных языков, какими являются славянские, лучшие результаты дает обращение к словарям синонимов (парадигматика), фразеологии (синтагматика) и одноязычным толковым

словарям (эпидигматика – многозначность). Привлечение соответствующих источников помогает корректировать результаты.

Для обработки источника № 27 участникам проекта предварительно потребовалось перевести метаязыковую часть с немецкого на русский язык, для источника № 24 – с английского на русский язык, для источника № 25 – с польского на русский. Кроме того, был проведен параметрический анализ двух словарей македонского языка (№ 14 и № 20) и получены два ядра македонской лексики, соответственно. Для русского и словацкого языков пришлось получить также по два ядра. Для русского: первый раз – с использованием данных словаря Ожегова (№ 15), второй раз – с использованием данных русско-финского словаря (Ф-параметр), словаря Ушакова (№ 19: Д-параметр), а также электронных фразеологического и синонимического словарей русского языка (С- и П-параметры соответственно). Для словацкого языка синтагматический и эпидигматический параметры определялись с использованием словаря № 22.

С помощью компьютерной программы MergOn (автор – преподаватель кафедры программного обеспечения и администрирования информационных систем ВГУ Н.В. Огаркова) были получены большое (содержащее слова, вошедшие в ядро по трем параметрам) и малое (содержащее слова, вошедшие в ядро по каждому из четырех параметров) ядра, а также периферия большого ядра (слова, вошедшие в ядро по двум параметрам) лексики 14 славянских языков: белорусского, болгарского верхнелужицкого, кашубского, македонского, нижнелужицкого, польского, русского, сербохорватского, словацкого, словенского, старославянского, украинского и чешского языков.

Использование в качестве источников информации о лексико-семантической системе языка словарей требует обоснования. Как известно, язык не дан лингвистам в непосредственном наблюдении. Это всегда научный конструкт. На протяжении последних двухсот лет считалось бесспорным, что язык описывается комплектом из грамматики и словаря. Но в последнее время приходится слышать все больше сомнений в достоверности словарей. Порою вопрос ставится даже так: «А что вы описываете, язык или словарь?». При этом предполагается, что язык – это реальность, а словарь – непонятно что. Интересно, что вопроса: «А что вы описываете, язык или грамматику?» обычно не ставят, хотя он столь же правомерен.

Это обстоятельство требует разъяснений. Язык как лингвистическая реальность ненаблюдаем. В непосредственном наблюдении лингвисту даны тексты: либо в виде процесса, развертывающегося во времени (речь), либо в виде результата (собственно текст: устный или письменный). Множество текстов дает некий корпус текстов, которому приписывается свойство репрезентативности. Следует сказать, что любой корпус – сколь бы велик он ни был – в силу своей закрытости и конечности будет неполон и случаен по отношению к языку. Любой корпус будет лишь ограниченным набором следов работы языка. Грамматики и словари – это производные от корпусов текстов, положенных в их основание. Таким образом, язык-реальность ненаблюдаем, корпус как отражение работы языка всегда неполон и случаен, а грамматики и словари являются отражением этого корпуса текстов. Следовательно, любой словарь и любая грамматика является лишь отражением отражения языка (корпуса текстов). Альтернативой исследования словарей может послужить лишь исследование корпусов, что принципиально ситуации не изменит: ведь сами словари взяты из тех же (или иных), но корпусов текстов. Разница лишь в том, что корпусным лингвистам придется самим проделать ту работу, какую до них и за них уже выполнили лексикографы. Таким образом, единственным реальным источником информации о лексико-семантической системе языка для лексикологов являются словари. Как любая человеческая продукция, словари могут быть более и менее удачны и совершенны. Но следует ли из этого, что за словарями не стоит никакой лингвистической реальности и словарям верить нельзя? Вековая практика создания и использования словарей говорит об обратном: словари выполняют свою функцию, дают достоверную информацию о языке. Но поскольку любой словарь (и любая грамматика) – лишь отражение отражения языка, они могут расходиться в своих показаниях. Специальное параметрическое исследование лексики, проведенное Т.А. Казаковой на материале немец-

ких словарей, показало, что «состав лексико-семантических ядер, получаемых в ходе анализа словарей различных объемов, совпадает на 64%» [Казакова 2008: 5], т.е. почти на 2/3. Одна треть приходится на «субъективный фактор». Таким образом, у нас есть основание полагать, что по крайней мере на 2/3 словари отражают языковую реальность. Кроме того, установлено – чем ближе к ядру, тем выше стабильность словаря: «удовлетворительные прогнозы касаются наиболее существенных в системном отношении лексико-семантических единиц: для большого ядра ЛСГ ошибка прогноза не превышает 29%, для малого – 13%, для доминанты – 6%» [Кретов 2006: 213–214].

Представим полученные результаты параметрического анализа лексики славянских языков.

1. ФУНКЦИОНАЛЬНАЯ СТРАТИФИКАЦИЯ СЛАВЯНСКОЙ ЛЕКСИКИ

Функциональный аспект лексики принято оценивать по (абсолютной или относительной) частоте появления слова в тексте. Эта оценка очень сильно зависит от содержания текста, которое, в свою очередь, через создателя и адресатов привязано ко времени и месту. Более объективна косвенная оценка функциональной активности слова по его длине, предложенная В.Т. Титовым, основывающаяся на зависимости длины слова от его частотности, выявленной Дж.К. Ципфом [Zipf 1949]. Проанализируем функциональное ядро лексики славянских литературных языков, ориентируясь, прежде всего, на данные о длине слова в инославяно-русских словарях.

Исследования показывают, что большие словари размывают длину слова как типологическую характеристику языка, т.к. разрастаются в основном за счет иносистемных слов или за счет словосочетаний, оформленных как одно слово. Для типологической характеристики лексики по длине пригоднее малые и средние словари. В них специфика лексико-семантической системы данного языка представлена в более чистом виде. Для характеристики словаря по функциональному параметру оптимальными являются словари объемом в 8000–12 000 слов, а наименее информативными – большие.

Исходя из этих соображений, мы и осуществили отбор словарей для сопоставительного анализа. Соотношение их объемов представлено в последней строке Табл. 1. Как видим, за двумя исключениями, привлекаемые словари относятся к малым (между 5 и 15 тыс. слов). По необходимости полабский представлен словарем сверхмалого типа, а македонский язык – словарем среднего типа.

Распределение слов по длине представлено в таблице 1.

Лишь в полабском максимум приходится на слова длиной в 5 звуков. В остальных языках он приходится на 7 звуков (македонский, русский, чешский, словацкий, сербохорватский, польский, болгарский, кашубский) или на 8 звуков (белорусский, словенский, украинский, старославянский, лужицкие языки). Как свидетельствует предварительный анализ русских словарей разного объема, различие между этими величинами не принципиально.

В целом, распределения в славянских словарях сходны. Выделяется лишь полабский материал в силу своей малочисленности и разговорной природы.

2. СИНТАГМАТИЧЕСКАЯ СТРАТИФИКАЦИЯ СЛАВЯНСКОЙ ЛЕКСИКИ

Непосредственная информация о богатстве синтагматики слов в двуязычных словарях отсутствует. Косвенным показателем богатства синтагматических связей славянской лексики является количество фразеологизмов, имеющих в заромбовой части словарной статьи. Как показывает практика, малые словари зачастую весьма «скупо» используют заромбовую часть, поэтому для некоторых языков приходилось привлекать информацию из большого двуязычного или фразеологического словаря.

Как показывает проведенное исследование, наибольшую синтагматическую активность в славянских языках имеют слова со значением частей тела человека (глаз, голо-

Распределение слов по длине в словарях славянских языков¹

Звуков	МАК	БЕЛ	СЛВН	УКР	РУС	ЧЕШ	СЛВЦ	СЕРБ	ПОЛ	БОЛ	СТ-СЛ	ПОЛАБ	КАШУБ	НЛУЖ	ВЛУЖ
1	9	2	6	1	6	8	9	6	3	12	0	0	7	1	5
2	85	11	7	1	29	85	29	36	33	44	1	2	51	31	34
3	552	103	249	152	283	339	269	323	209	253	28	105	343	287	333
4	1705	359	638	495	628	731	585	673	564	709	339	406	880	714	1013
5	3791	1058	1308	1055	1319	1544	1358	1268	1057	1278	524	673	1692	1479	2458
6	4979	1575	1724	1462	1571	1730	1621	1431	1353	1384	873	577	2043	2111	3766
7	5451	2092	2014	1753	1763	2006	1863	1560	1505	1502	1094	408	2485	2524	4908
8	5196	2116	2147	2001	1514	1711	1719	1538	1490	1328	1316	303	2378	2804	5514
9	4266	1954	1740	1767	1274	1274	1282	1194	1136	967	1244	158	1582	2140	4852
10	2721	1549	1331	1454	955	887	930	800	852	688	909	77	983	1595	3760
11	1573	1059	771	1086	745	451	488	518	627	367	584	19	422	915	2421
12	735	660	462	609	493	222	274	222	369	218	304	8	151	477	1441
13	286	382	207	349	318	113	122	108	216	76	167	4	60	230	770
14	118	208	72	171	208	67	67	52	127	56	80		16	104	359
15	60	93	30	60	114	19	29	17	67	20	52		7	44	181
16	12	30	5	26	57	6	13	5	42	9	25		1	17	72
17	5	14	2	13	22	2	3	1	12	3	13			13	23
18	8	4	1	6	13	2	2		3	3	8			3	13
19	0	1	1	3	8		1		2	3	5				5
20	2	2	1	0	4				1	2	1			1	4
21	0		1	2	3						2				1
22	0			1										1	5
23	1														
Сумма	31 555	13 272	12 717	12 467	11 327	11 197	10 664	9752	9668	8922	7569	2740	13 101	15 491	31 933

¹ Данная таблица, как и таблица 3, содержит ту базовую информацию, на которой строится все исследование.

ва, рука), времени и его отрезков (год, время, час), а также бытийные глаголы (быть, иметь, давать).

Таким образом, можно подсчитать суммарный синтагматический вес славянских слов: *рука* 13,92, *дать* 13,815, *иметь* 13,668, *голова* 13,515, *время* 12,71, *глаз* 12,599, *час* 12,551, *ряд* 11,808, *год* 11,745, *быть* 10,638. Обратим внимание на то, что больше всего слов из этого вошло в малое ядро старославянского языка – 6 (слова со значениями 'быть', 'время', 'год', 'рука', 'голова', 'иметь'). Далее следует русский язык, в ядре которого 4 слова из данного списка (*дать*, *час*, *быть*, *время*). Наш материал наталкивает на предположение, что ядро славянской фразеологии было заложено еще в старославянский период и унаследовано языками, имеющими с ним преемственность письменной традиции. С этой точки зрения нуждаются в серьезной проверке свидетельства болгарского словаря-источника.

3. ПАРАДИГМАТИЧЕСКАЯ СТРАТИФИКАЦИЯ СЛАВЯНСКОЙ ЛЕКСИКИ

Вхождение слов в многочисленные синонимические ряды является системным показателем парадигматических связей лексических единиц. Эта информация может быть извлечена из двуязычных словарей в результате анализа дефиниций: квантитативного, описанного В.Т. Титовым [Титов 2004], или позиционного, предложенного несколько позднее [Кретов, Титов 2006]. Суть обоих подходов состоит в максимальном использовании информации, содержащейся в двуязычных словарях.

При позиционном анализе, применяемом нами, синонимами признаются слова, если дефиниции каких-то из их значений совпадают полностью или в начальной части (например, в первых трех полноточных словах).

Если главная идея, на которой основан параметрический анализ лексики, состоит в кумулятивной организации словаря: по мере продвижения от периферии к ядру устойчивость лексико-семантической системы увеличивается, то логично предположить, что это распространяется и на устройство синонимии.

Системный вес лексических значений, объединяющих синонимические ряды, убывает по мере убывания их размерности. Например, синонимический ряд, объединяющий семь слов, в чешском словаре всего один, а синонимических рядов, состоящих из двух слов, – 973.

Чем важнее функция в системе, тем надежнее она дублируется, чем важнее то или иное значение для носителей языка, тем больше синонимов оно имеет. Следовательно, размерность синонимического ряда может быть мерой важности данного значения для носителей данного языка. На какие же значения в качестве самых важных для носителей славянских языков указывают наши словари-источники? Представим информацию по первым двум полноточным словам, находящимся в «верхушке» парадигматических ядер славянских языков. Самый многочисленный синонимический ряд можно назвать доминантным значением, а второй – вице-доминантным значением.

Для русского языка это значение 'умереть', объединяющее 46 синонимов; вице-доминанта – 'ударить' 44 синонима. Для украинского языка значение 'бить' – 79 синонимов; на втором месте сразу два глагола: *охоплювати* ('захватывать, обнимать') и *сказати* ('сказать, произнести'), имеющие по 63 синонима. Для белорусского языка – это значение 'ударить' – 62 синонима; на втором месте ряд со значением 'хороший' – 54 синонима. Для чешского – доминантным является значение 'отрезать'. Для словацкого языка вверху оказались сразу 5 синонимических рядов, которые содержат по 4 слова. Их объединяют значения: 'движение', 'ездить', 'рабочий', 'упасть', 'центр'. Сходная ситуация характеризует польский язык – 3 ряда по 4 синонима ('жидкий', 'нарушать', 'удар'). Для сербского словаря первые полноточные слова, образующие синонимические ряды из четырех слов, имеют значение 'вырасти' и 'кудрявый'. Правда, в обоих случаях перед нами не синонимы, а словообразовательные варианты. Для словенского языка лидирует слово 'охватывать', входящее в дефиниции 8 слов. Далее следует ряд, объединенный дефиницией 'охватить'. Для македонского языка 12 синони-

Стратификация слов с максимальным синтагматическим весом в славянских языках

Лексема	Мак	Бел	Слвн	Укр	Рус	Чеш	Слвк	Серб	Пол	Болг	СтСл	Кашуб	нЛуж	вЛуж
Рука	1,000	1,000	0,996	1,000	1,000	1,000	0,998	0,999	0,999	0,996	0,980	0,954	0,999	0,999
Дать, давать	1,000	1,000	0,996	0,999	0,995	1,000	0,999	0,978	0,934	0,997	0,991	0,954	0,988	0,984
Иметь	0,979	0,997	1,000	0,930	0,982	0,915	1,000	0,978	0,987	0,999	0,904	0,999	0,999	0,999
Голова	1,000	1,000	0,998	1,000	1,000	0,999	0,805	0,895	1,000	0,922	0,904	0,996	0,997	0,999
Время	0,996	—	1,000	0,997	1,000	0,998	0,994	0,998	1,000	1,000	0,904	0,825	0,999	0,999
Глаз	1,000	1,000	0,979	1,000	0,999	0,998	0,983	0,993	—	0,922	0,904	0,825	0,997	0,999
Час	0,996	1,000	0,998	0,998	0,998	0,972	0,805	0,998	0,934	0,993	0,998	—	0,863	0,998
Ряд	1,000	—	0,991	0,986	0,982	—	0,983	1,000	0,975	0,993	0,904	0,999	0,997	0,998
Год	—	0,984	1,000	0,992	0,991	0,915	0,998	0,998	0,999	0,922	—	0,996	0,952	0,998
Быть	—	—	1,000	0,999	0,997	—	0,942	0,895	0,934	0,922	1,000	0,953	0,999	0,997

мов объединяет доминанта 'сходить с ума'. По крайней мере 6 различных корней в этом ряде представлены. Для болгарского языка первые места занимают ряды со значением 'окружать' и 'окружить' – 3 разных корня. Для старославянского языка – значения 'больной', 'единосущный', 'неразумный', 'опечалиться', 'чудо' объединяют по 6 синонимов. Для кашубского языка доминантным по парадигматическому весу является синонимический ряд со значением 'нищий'. Для нижнелужицкого это значение – 'помнить, вспомнить', образующее ряд из 12 синонимов, для верхнелужицкого 'болтать, тараторить', дающее синонимический ряд в 6 синонимов. Доминанты синонимии славянских языков имеют сходство. Например, значение 'ударить', 'удар', 'бить' объединяют русский, украинский и польский языки.

Важнейшими значениями в обследованных словарях являются: 'удар, ударять, бить, упасть, чудо, умереть, окружить, отрезать, рабочий, жидкий, движение, ездить, охватить'.

Показательно, что большая часть значений указывает на угрозу для жизни или здоровья (физического или психического) или аномалии ('чудо').

4. ЭПИДИГМАТИЧЕСКАЯ СТРАТИФИКАЦИЯ СЛАВЯНСКОЙ ЛЕКСИКИ

Под эпидигматическими связями, вслед за автором этого термина Д.Н. Шмелевым [Шмелев 1973], мы понимаем «деривационные отношения в широком смысле», включающие межсловную (словообразование) и внутрисловную («семантическое словообразование», порождающее многозначность) деривацию, трактуемую как «глубинная ось языка», обеспечивающая в диахронии – преемственность и устойчивость, а в синхронии – гибкость и динамику и сообщающая языку практические неограниченные номинативные возможности.

Косвенным показателем эпидигматической активности слов в двуязычных словарях является их многозначность. Чем больше значений у слова, тем богаче его эпидигматические связи и, следовательно, тем надежнее оно укоренено в лексико-семантической системе языка. Поскольку для образования семантемы (системы семем) многозначного слова требуется значительное время, то многозначные и сверхмногозначные слова являются своего рода центрами системного порядка и скрепами, объединяющими отдельные слова в систему. Богатство

значений – еще и показатель многофункциональности слова, а следовательно, – его функциональной динамичности. Таким образом, многозначность сочетает динамику и устойчивость.

Как показало исследование, ядра межсловной и внутрисловной деривации пересекаются на 70%. Такое огрубление, во-первых, вынужденно, а во-вторых, оправданно, ибо обеспечивает получение достоверных результатов.

Распределение слов по числу значений в словарях славянских языков представлено в таблице 3 (на основании проведенного анализа в качестве наиболее взвешенного источника сведений о многозначности в русском языке нами избран словарь Ушакова, а для украинского языка – Большой толковый словарь).

Исследуя многозначность славянской лексики, мы допускаем неадекватность данных двуязычных словарей, которая при необходимости может корректироваться сведениями одноязычных толковых словарей.

Определим такой типологически важный показатель, каким является индекс многозначности (ИМ), определяемый как среднее количество значений, приходящихся на слово в данном словаре. Его значения для словарей славянских языков таковы:

кашуб	рус	пол	бел	СтСл	лужН	слвн	укр	мак	лужВ	серб	бол	слвц	чеш
1,632	1,361	1,326	1,317	1,3	1,287	1,271	1,265	1,256	1,21	1,182	1,158	1,155	1,092

Самым многозначным из славянских словарей оказывается кашубский, а самым бедным многозначностью – чешский. Высокая многозначность кашубского языка вызывает сомнения и нуждается в проверке. Одной из причин высокого значения ИМ в кашубском словаре может быть приписывание в нем статуса леммы формам одного слова. Исследование показало, что ИМ практически не зависит от размера словаря. Это обстоятельство позволяет в целом с доверием отнестись к ИМ славянских языков как к их типологической характеристике.

Если значения ИМ славянских языков (за исключением кашубского) располагаются в интервале 1,36–1,09, то значения живых романских языков, установленные ранее В.Т. Титовым, располагаются практически в том же интервале 1,28–1,14. Возможно, совпадение интервалов ИМ в романских и славянских словарях свидетельствует об их типологической близости и генетическом родстве – принадлежности индоевропейской языковой семье.

5. СРАВНЕНИЕ СЛАВЯНСКИХ СЛОВАРЕЙ ПО ДОМИНАНТЕ ПАРАМЕТРИЧЕСКОГО РЯДА

Параметрический анализ позволяет представить словарь в виде конуса, вершиной которого является одно-единственное слово. Полученное таким образом слово является не просто словом: его позиция является производной и равнодействующей от всех остальных слов словаря, взятых по четырем системнообразующим параметрам. Поэтому семантические координаты доминанты лексико-семантической системы обладают не меньшей информативностью, чем, скажем, координаты высшей точки Джомолунгмы. Даже имея на карте координаты одних пиков гор-восьмитысячников, мы можем в общих чертах представить себе очертания горной системы. Аналогично, и слова-доминанты дают представление о своеобразии лексико-семантической системы. Пик – точка, вобравшая в себя всю гору. Доминанта – слово, вобравшее в себя весь словарь.

Так, например, из шести романских языков четыре (латинский, испанский, португальский, румынский) имеют доминанту «дать», один (итальянский) – доминанту «делать» и один (французский) – доминанту «сладкий, мягкий» [Титов 2004: 160]. Параметрический анализ лексики в силу своей принципиальной системности позволяет обнаружить типологический потенциал уже в доминанте (самом системно «весомом» слове) словаря. Доминанта словаря – слово, набравшее максимальный пара-

Распределение слов в зависимости от числа значений (в относительных величинах)

Язык/ Знач	БЕЛ	БОЛ	МАК	ПОЛ	РУСС	СЕРБ	СЛВЦ	СЛВН	СтСлав	УКР	НЛУЖ	ВЛУЖ	ЧЕШ	КАШУБ
40					0,000012					0,000005				
36										0,000010				
34										0,000005				
32										0,000005				
30										0,000005				
28					0,000012									
27	0,000017													
26														
25														
24					0,000012		0,000020			0,000005				
23					0,000012					0,000010				
22			0,000034							0,000005				
21					0,000012									
20							0,000020			0,000005				
19					0,000012					0,000020				
18					0,000036					0,000015				
17	0,000017		0,000034		0,000024					0,000035				
16	0,000067				0,000024					0,000035				
15			0,000034		0,000012					0,000025				
14	0,000050		0,000101		0,000059		0,000040			0,000059				
13	0,000133				0,000095					0,000064		0,000031		
12	0,000133		0,000101		0,000142		0,000020			0,000143				
11	0,000283		0,000034		0,000118		0,000160		0,000126	0,000203				0,001145
10	0,000183		0,000304		0,000178		0,000160			0,000282		0,000031		0,000687
9	0,000467		0,000439	0,000113	0,000533		0,000260		0,001509	0,000455		0,000031		0,001985
8	0,000600		0,000608	0,000339	0,000841		0,000620		0,000629	0,000801	0,000065	0,000031		0,001908
7	0,001300		0,001249	0,000790	0,001422		0,000780		0,001132	0,001275	0,000258	0,000094		0,003817
6	0,002650		0,002869	0,002709	0,002831		0,001300		0,002390	0,002605	0,000646	0,000219		0,009770
5	0,005850	0,001024	0,006751	0,003500	0,005769	0,000205	0,002820	0,000394	0,005282	0,005526	0,001808	0,001315	0,000179	0,016640
4	0,014750	0,002275	0,018530	0,014337	0,013386	0,002358	0,007340	0,004725	0,015721	0,011990	0,009038	0,004353	0,000714	0,036486
3	0,038350	0,020585	0,061194	0,045157	0,043891	0,018145	0,020680	0,034179	0,038737	0,030500	0,041895	0,026869	0,008574	0,074422
2	0,134100	0,105538	0,191683	0,156807	0,167603	0,137365	0,073680	0,187352	0,117721	0,104692	0,163385	0,134720	0,071805	0,188459
1	0,801050	0,870579	0,716036	0,776247	0,762966	0,841927	0,892100	0,773350	0,816753	0,841222	0,782906	0,832305	0,918728	0,664682
Слов:	60 000	8793	29 627	8858	84 414	9755	50 000	12 698	7951	202 327	15 491	31 933	11 197	13 101

Лексико-семантические доминанты славянских словарей

Ранг:	1	2	3	4	5
Язык	доминанта	вице-доминанта			
старославянский	быти	годъ	слово	мѣра	доухъ
сербский	ред	час	број	доба	дођи
болгарский	ред	лют	лице	курс	седа
македонский	има	ум	чини	род	жица
русский	дать	час	бить	дом	мир
украинский	іти	лад	вік	дати	бити
белорусский	біць	бок	ісці	вада	браць
словацкий	pán	muž	hrať	vôľa	vec
чешский	dílo	kraj	míra	bod	mysl
словенский	delo	grd	biti	blok	domać
нижнелужицкий	biś	graś	wześ	rěd	braś
верхнелужицкий	ćah	hrać	tołc	trěc	ćeńki
польский	ciąć	ująć	bieg	łeb	być
кашубский	mæc	vid	pôn	křiv	łono

метрический вес (по сумме четырех параметров). Следующее за ним слово является вице-доминантой.

Нами был проведен сопоставительный и уже не количественный, а качественный анализ доминант славянских языков. Для наглядности возьмем не только доминанты, но и слова последующих четырех рангов.

Как видим, в лексическом аспекте типологически едины оказываются болгарский и сербский языки, у которых совпадают доминанты (*ред*), к ним близок украинский с вице-доминантой *лад* и нижнелужицкий, имеющий существительное *rěd* на четвертом месте в малом ядре. Идеей движения объединяются украинский (*іти*), польский (*ciąć*, *bieg*), верхнелужицкий (*ćah*), белорусский (*ісці*). Примыкает к ним сербский язык, в котором слово со значением движения замыкает пятерку самых «весомых» слов. Бытийность, доминирующая в старославянском (*быти*, *годъ*), тесно связана с понятием времени: *час* (русский, сербский), *вік* (украинский), *доба* (сербский). *Дать*-языками можно считать русский (доминанта) и украинский (4 ранг). При этом польский, нижнелужицкий и белорусский имеют в числе обследованных слов антонимичные лексемы *ująć* 'взять, отнять' и *wześ* 'брать, взять, отнимать', *braś* 'брать, взять, хватать', *brać* 'брать', а македонский имеет доминанту обладания *има*. Вице-доминанты совпадают у нижне- и верхнелужицких языков (*graś*, *hrać*). К ним примыкает словацкий '*hrať*' (3 ранг). Идеей труда объединены ядра лексиконов чешского (*dílo*), словацкого (*robit'*), словенского (*delo*), белорусского (*рабіць*), македонского (*чини*) и словацкого (*vec*) языков. Совпадают также верхушки параметрических ядер у кашубского и словацкого языков, в которые попало слово со значением 'мужчина'.

Если сгруппировать семантически близкие слова, т.е. сложить параметрические веса слов со значением 'работа, труд', 'делать, работать', можно определить, какая из доминант в этих словарях самая весома. Таким образом, доминантой доминант славянских словарей являются слова со значением 'дело, труд, работать', если принять во внимание, что 'бить' обозначает неотъемлемую и существенную часть трудового процесса, доминантность труда у славян станет еще очевиднее. Вице-доминантным значением является значение движения ('ход, идти'). Третью позицию занимают слова со значением упорядочения мира: 'ряд, очередь, полоса'. Все три указанных типа значений позволяют утверждать, что синтетической доминантой лексико-семантических систем славянских языков является динамика, направленная на упорядочение мира.

6. КАРТА ЛЕКСИКО-СЕМАНТИЧЕСКОГО ПРОСТРАНСТВА СЛАВЯНСКИХ ЯЗЫКОВ

Для составления карты лексико-семантического пространства славянских языков учитывалась информация о словах, вошедших в параметрическое ядро по трем параметрам. Размерность таких множеств колеблется от 141 до 983 слов (таблица 5):

Таблица 5

Размерность параметрических ядер славянских языков

№	ЯЗЫКИ	Кол-во слов
1	Белорусский	983
2	Русский	817
3	Словацкий	702
4	Македонский	669
5	Старославянский	401
6	Кашубский	395
7	Польский	360
8	Сербский	312
9	Украинский	307
10	Чешский	283
11	Болгарский	263
12	Нижнелужицкий	194
13	Словенский	163
14	Верхнелужицкий	141

Далее анализировалась семантика этих ядер, высчитывалось количество совпадений по дефинициям для каждой пары языков. Для определения близости лексико-семантических систем двух языков была использована следующая формула: число совпадений дефиниций делилось на среднее арифметическое от ядер двух сравниваемых языков.

Результаты представлены в таблице 6.

В итоге с помощью компьютерной программы, разработанной программистом кафедры типологической и прикладной лингвистики Воронежского госуниверситета М.В. Катовым, была получена следующая карта.

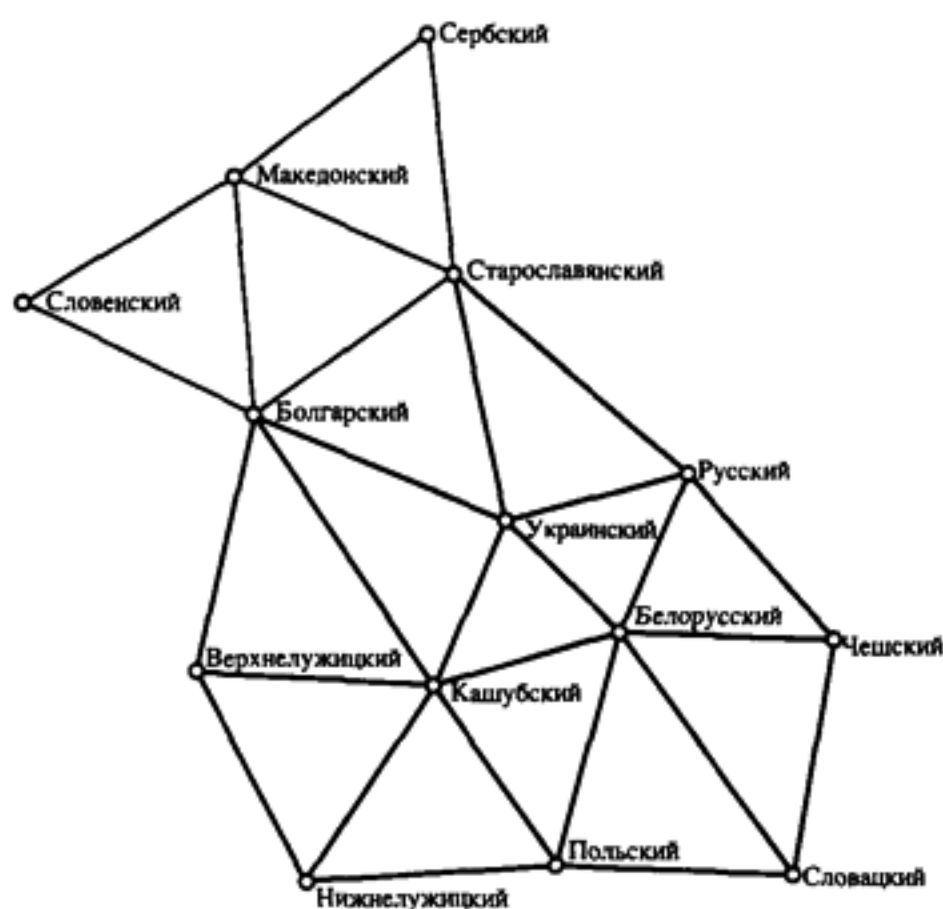


Рис. 1. Лексико-семантическая карта славянских языков (по коэффициенту близости больших параметрических ядер лексики)

Коэффициенты близости лексико-семантических систем славянских языков (бинарные совпадения)

	Нижнелуж.	Верхнелуж.	Польский	Чешский	Русский	Украин.	Болгар.	Сербский	Старослав.	Словенский	Македон.	Белорус.	Словацкий	Кашуб.
Нижнелужицкий		0,1851	0,1445	0,1468	0,1702	0,1158	0,0832	0,0830	0,0849	0,0728	0,1126	0,2719	0,1708	0,2003
Верхнелужицкий	0,1851		0,1311	0,1462	0,1405	0,0893	0,1040	0,0971	0,0736	0,0789	0,0938	0,2316	0,1561	0,1866
Польский	0,1445	0,1311		0,1886	0,2496	0,1440	0,0639	0,0683	0,1147	0,0848	0,0939	0,3611	0,1865	0,2613
Чешский	0,1468	0,1462	0,1886		0,2266	0,1186	0,0842	0,0807	0,1141	0,0673	0,0800	0,2648	0,1883	0,2456
Русский	0,1702	0,1405	0,2496	0,2266		0,3619	0,1260	0,1109	0,1490	0,0918	0,0998	0,4063	0,1210	0,1987
Украинский	0,1158	0,0893	0,1440	0,1186	0,3619		0,0667	0,0711	0,1435	0,0468	0,1030	0,4400	0,1375	0,3675
Болгарский	0,0832	0,1040	0,0639	0,0842	0,1260	0,0667		0,1600	0,1669	0,1268	0,1802	0,2220	0,1309	0,2736
Сербский	0,0830	0,0971	0,0683	0,0807	0,1109	0,0711	0,1600		0,1667	0,1179	0,1733	0,2387	0,1271	0,2093
Старославянский	0,0849	0,0736	0,1147	0,1141	0,1490	0,1435	0,1669	0,1667		0,1605	0,1784	0,2402	0,1380	0,1373
Словенский	0,0728	0,0789	0,0848	0,0673	0,0918	0,0468	0,1268	0,1179	0,1605		0,1586	0,2063	0,1411	0,1362
Македонский	0,1126	0,0938	0,0939	0,0800	0,0998	0,1030	0,1802	0,1733	0,1784	0,1586		0,2985	0,1352	0,1017
Белорусский	0,2719	0,2316	0,3611	0,2648	0,4063	0,4400	0,2220	0,2387	0,1966	0,2063	0,2985		0,2592	0,3381
Словацкий	0,1708	0,1561	0,1865	0,1883	0,1210	0,1375	0,1309	0,1271	0,1380	0,1411	0,1352	0,2592		0,1703
Кашубский	0,2003	0,1866	0,2613	0,2456	0,1987	0,3675	0,2736	0,2093	0,1373	0,1362	0,1017	0,3381	0,1703	

Как видим, южно-, восточно- и западнославянские языки образуют три континуума, объединенных между собой. Самым компактным является восточнославянский континуум. Вполне компактен и континуум южнославянских языков. Отсутствие прямой связи старославянского языка со словенским обоснованно и логично: старославянский язык обслуживал православие, которое в Словении оказалось вытесненным католичеством и протестантизмом. Западнославянский континуум организован тремя парами языков: чешский/словацкий, польский/кашубский и верхнелужицкий/нижнелужицкий. Большая близость (наличие непосредственной связи) нижнелужицкого с польским и кашубским понятна. Менее удачна на карте мнимая удаленность верхнелужицкого от чешского и словацкого, а двух последних – от южнославянских.

Анализируя лексико-семантическую карту славянских языков, следует иметь в виду ее условность и неполноту: наличие 14-ти сопоставляемых языков предполагает построение 14-мерного объекта, который мы вынуждены проецировать на двухмерную плоскость, неизбежно (и вынужденно) огрубляя реальное богатство связей каждого из узлов нашего графа.

Полученные результаты подтверждают традиционное деление славянских языков на южные, западные и восточные и вместе с тем свидетельствуют, что такое деление существенно упрощает и обедняет реальную сложность картины лексико-семантической дифференциации славянской лексики.

СПИСОК ЛИТЕРАТУРЫ

- Гак 1989 – В.Г. Гак. Сопоставительная типология французского и русского языков. М., 1989.
 Городецкий 1969 – Б.Ю. Городецкий. К проблеме семантической типологии. М., 1969.
 Гудавичюс 1985 – А. Гудавичюс. Сопоставительная семасиология литовского и русского языков. Вильнюс, 1985.
 Дыбо 1996 – А.В. Дыбо. Семантическая реконструкция в алтайской этимологии. Соматические термины (плечевой пояс). М., 1996.
 Казакова 2008 – Т.А. Казакова. Параметрический анализ немецкой лексики: Автореф. дис. ... канд. филол. наук. Воронеж, 2008.
 Кретов 2006 – А.А. Кретов. Основы лексико-семантической прогностики. Воронеж, 2006.
 Кретов, Меркулова 2008 – А.А. Кретов, И.А. Меркулова. Динамика и устойчивость в лексической системе славянских языков // Динамика та стабільність лексичних і словотвірних систем слов'янських мов – тематичний блок. XIV Міжнародний з'їзд славістів 10.09.–16.09.2008, Охрид, Республіка Македонія. 2008.
 Кретов, Титов 2006 – А.А. Кретов, В.Т. Титов. Алгоритм позиционного выявления синонимии // Вестник ВГУ. Серия «Системный анализ и информационные технологии». 2006. № 1.
 Майсак, Рахилина 2007 – Т.А. Майсак, Е.В. Рахилина (ред.). Глаголы движения в воде: лексическая типология. М., 2007.
 Мечковская 2000 – Н.Б. Мечковская. Общее языкознание. Структурная и социальная типология языков: Учеб. пособие. Минск, 2000.
 Москович 1965 – В.А. Москович. Опыт количественной типологии семантического поля // ВЯ. 1965. № 4.
 Рахилина, Прокофьева 2004 – Е. Рахилина, И. Прокофьева. Родственные языки как объект лексической типологии: русский и польские глаголы вращения // ВЯ. 2004. №1.
 Супрун 1983 – А.К. Супрун. Лексическая типология славянских языков. Минск, 1983.
 Супрун 1988 – А.К. Супрун. Принципы сопоставительного изучения лексики // Методы сопоставительного изучения языков. М., 1988.
 Титов 2002 – В.Т. Титов. Общая количественная лексикология романских языков. Воронеж, 2002.
 Титов 2004 – В.Т. Титов. Частная количественная лексикология романских языков. Воронеж, 2004.
 Толстой 1963 – Н.И. Толстой. Из опытов типологического исследования славянского словарного состава. I // ВЯ. 1963. № 1.
 Толстой 1966 – Н.И. Толстой. Из опытов типологического исследования славянского словарного состава. II // ВЯ. 1966. № 5.

- Томмола 2003 – Ханну Томмола. К лексико-семантической типологии славянских языков // J. Papinniemi, J. Lindstedt, B. Hellman (eds.). *Studia Slavica Finlandensia*. Т. XX. Finnish contributions to the 13th International Congress of Slavists. Ljubljana, August 15–21, 2003.
- Шмелев 1973 – Д.Н. Шмелев. Проблемы семантического анализа лексики: на материале русского языка. М., 1973.
- Zipf 1949 – G.R. Zipf. Human behavior and the principle of least effort. Cambridge (Mass.), 1949.

СПИСОК ИСПОЛЬЗОВАННЫХ СЛОВАРЕЙ

1. Белорусско-русский словарь / Под ред. К.К. Крапивы. М., 1962 (ок. 90 000 слов).
2. Э. Благова, Р.М. Цейтлин, С. Геродес, Л. Панцерова, М. Бауэрова. Старославянский словарь (по рукописям X–XI веков). М., 1994 (около 10 000 слов).
3. Верхне-лужицко-русский словарь / Сост. К.К. Трофимович; Под ред. Ф. Михалка, П. Фёлькиеля. М.; Бауцен, 1974 (36 000 слов).
4. Д.И. Ганич, И.С. Олейник. Русско-украинский и украинско-русский словарь. К., 1992 (12 000 слов).
5. С.М. Грабчиков. Белорусско-русский словарь. Для сред. школы / Под ред. А.Е. Баханькова. 2-е изд., перераб. Минск, 1975.
6. В.П. Гудков. Сербско-русский и русско-сербский словарь. 5-е изд., стереотип. М., 2006 (10 000 слов).
7. Д.А. Длуги, Б.Г. Раевский, Н.Р. Буравцева. Чешско-русский Словарь. 2-е изд., стереотип. М., 1973 (около 11 000 слов).
8. А.П. Загнітко, М.О. Вінтонів, І.Г. Данилюк. Великий українсько-російський і російсько-український словник. Донецьк, 2007 (ок. 22 000 слов).
9. Д. Коллар, В. Доротьякова, М. Филкусова, Я. Марушиакова. Карманный словацко-русский словарь. 4-е изд., испр. и доп. М., 1982 (10 000 слов).
10. М.К. Клышка. Слоўнік сінонімаў і блізказначных слоў. Мінск, 2005.
11. М.А. Леонидова. Карманный болгарско-русский словарь. Джебен българско-руски речник. 9-е изд., испр. и доп. М., 1986 (10 000 слов).
12. І.Я. Лепешаў. Слоўнік фразеалагізмаў: У 2 т. Мінск, 2008.
13. Н.С. Мажэйка. Частотны слоўнік беларускай мовы. Мінск, 2006 (64 536 слоў).
14. Македонско-русский словарь / Под общ. ред. Р.П. Усиковой, Е.В. Верижниковой. М., 2003 (свыше 40 000 слов).
15. С.И. Ожегов. Словарь русского языка.
16. М.Ф. Розвадовская. Польско-русский словарь. 7-е изд. М., 1963 (ок. 50 000 слов и выражений).
17. Словарь синонимов украинского языка: В 2 т. Киев, 1999–2000.
18. Тлумачальны слоўнік беларускай мовы / Пад рэд. М.Р. Судніка, М.Н. Крыўко. Мінск, 1996 (больше за 65 000 слоў).
19. Толковый словарь русского языка: В 4 т. / Под ред. Д. Н. Ушакова, М., 1935–1940.
20. Д. Толовски, В.М. Иллич-Свитыч. Македонско-русский словарь: с приложением краткого грамматического справочника. М., 1963.
21. Фразеологічний словник української мови в двох томах / Под ред. В.М. Білоноженко та інш. Київ, 1993.
22. J. Doruľa, J. Kačala, M. Marsinová et al. Krátky slovník slovenského jazyka. Bratislava, 1989 (около 50 000 slov).
23. A. Hiltunen, U.-L. Heino. Suomi venäjä suomi. Porvoo, 1968.
24. K. Polański, J. Sehnert. Polabian-English dictionary. The Hague; Paris, 1967.
25. S. Ramult. Słownik języka pomorskiego czyli kaszubskiego. Warszawa, 1893.
26. J. Sever, O.S. Plotnikova. Rusko-slovenski in slovensko-raski moderni slovar = Russko-slovenskij i slovensko-russkij sovremennij slovar'. Ljubljana, 2006.
27. M. Starosta. Dolnosorbско-němski słownik: šulski słownik dolnosorbiskeje řečy /Niedersorbisch-deutsches Wörterbuch. Leipzig, 1984.