

**КВАНТИТАТИВНЫЕ ХАРАКТЕРИСТИКИ ЛЕКСИКИ
БОЛГАРСКОГО ЯЗЫКА ПО ДВУЯЗЫЧНОМУ СЛОВАРЮ**

И. А. Меркулова

Воронежский государственный университет

Поступила в редакцию 28.04.2014 г.

Аннотация. Статья посвящена выявлению наиболее значимых элементов лексико-семантической системы болгарского языка по 4 параметрам: частотность, многозначность, синтагматическая и синонимическая активность. Анализ основывается на болгарско-русском словаре С.Б. Бернштейна. В результате слияния полученных множеств слов определяется лексико-семантическое ядро болгарского языка.

Ключевые слова: лексика, словарь, болгарский язык, параметрический анализ, лексико-семантическое ядро.

Annotation. The paper discusses the most important elements' of lexico-semantic system according to 4 parameters such as frequency, polysemy, syntagmatic and synonymic index. The analysis is based on Bulgarian-Russian dictionary by S.B. Bernstein. As a result of given words' incorporation the lexico-semantic nucleus is derived.

Keywords: vocabulary, dictionary, Bulgarian language, parametric analysis, lexico-semantic nuclear.

Для характеристики лексики любого языка важную роль играют параметры, характеризующие системные отношения соположения (синтагматика), противопоставления (парадигматика), выводимости (эпидигматика) в их динамике (функциональность). Методика параметрического анализа [1; 2] позволяет выявить эти параметры и получить ядро лексической системы языка. На страницах журнала «Вестник ВГУ. Серия: системный анализ и информационные технологии» уже публиковались результаты параметрических исследований некоторых языков [3–6].

Объект нашего исследования – лексика одного из южнославянских языков, болгарского, в том виде, в котором она представлена в двуязычном словаре большого объема [7].

Для получения количественных характеристик лексики языка необходимо решить ряд сопутствующих задач: 1) создать электронную версию словаря; 2) преобразовать полученную версию в базу данных; 3) получить количественную информацию по каждому из параметров исследуемого языка. После этого путем слияния полученных четырех множеств можно получить ядро лексической системы данного языка на базе словаря-источника.

Первым параметром характеристики лексики является длина слова. Существует статистическая закономерность, согласно которой длинные слова употребляются гораздо реже, чем короткие. Именно поэтому наиболее короткими являются служебные элементы: предлоги, союзы, частицы.

Статистические данные по длине болгарских слов представлены в табл. 1.

Таблица 1
Распределение болгарских слов
по функциональному параметру

Букв	Слов	Накопл.	Ф-вес
2	17	17	0,9996
3	404	421	0,9904
4	1367	1788	0,9592
5	3248	5036	0,8851
6	5312	10348	0,7639
7	6721	17069	0,6106
8	7374	24443	0,4424
9	6356	30799	0,2974
10	4858	35657	0,1866
11	3206	38863	0,1135
12	2033	40896	0,0671
13	1249	42145	0,0386
14	784	42929	0,0207
15	374	43303	0,0122
16	216	43519	0,0073
17	140	43659	0,0041
18	82	43741	0,0022
19	39	43780	0,0013
20	23	43803	0,0008
21	12	43815	0,0005
22	10	43825	0,0003
23	8	43833	0,0001
24	3	43836	0,00002
25	1	43837	0,00000

Как видим, наибольшее количество слов проанализированного корпуса имеют длину в 7 букв. Функциональное ядро составляют 1788 слов. Самыми короткими среди полноточных являются слова длиной 2 буквы. Это *ад* 'ад', *ар* 'ар', *ас* 'туз', *еж* 'еж', *ек* 'эхо, отзвук', *ил* 'ил', *ню* 'нюх, обоняние', *ос* 'ось', *ум* 'ум', *юг* 'юг', *юз* 'мера веса', *яз* 'плотина, запруда', *як* 'крепкий, здоровый, сильный', *ям* 'есть, кушать', *яд* 'гнев, злость', *йй* 'пружина (устар.)', *яр* 'яр'.

Естественно, что длина слова в буквах и звуках не всегда совпадает. Например, *ю* и *я* дают два звука, согласный *дж* произносится слитно и др.

Корректировке полученных данных помогает вычисление Коэффициента совершен-

ства орфографии. Для этого необходимо взять контрольное множество (500 или 1000 слов) и получили суммарную длину этих слов в звуках (*ДЗсум*) и буквах (*ДБсум*). Коэффициент совершенства орфографии вычисляется по формуле:

$$K_{\text{осо}} = \frac{ДЗсум}{ДБсум}.$$

В результате сопоставления полученных длин в звуках и буквах получаем коэффициент совершенства орфографии болгарского языка, равный 0,952. Это значение делает несущественным расхождение величин длины слова в звуках и буквах.

Вторым параметром характеристики лексикологии является ее синтагматическая активность. В качестве косвенного показателя синтагматического веса слова принимается его идиоматическая активность, информацию о которой можно извлечь из зарамбовой части двуязычного словаря. В табл. 2 отражены результаты анализа словаря по синтагматическому параметру.

Таблица 2
Поранговая динамика синтагматического
веса болгарских слов

ФС	Слов	Накопл.	С-вес
24	1	1	0,99998
19	1	2	0,99995
15	4	6	0,99986
13	0	6	0,99986
12	1	7	0,9998
11	5	12	0,9997
10	5	17	0,9996
9	6	23	0,9995
8	4	27	0,9994
7	9	36	0,9992
6	16	52	0,9988
5	22	74	0,9983
4	49	123	0,9972
3	84	207	0,9953
2	314	521	0,9883
1	1457	1978	0,9554
0	42441	44390	0,0000

Наибольшее количество сочетаний обнаружилось у слова *око* 'глаз' (24 фразеологизма): *зеница на окото* зрачок глаза; *насълзёни очи* наполненные слезами глаза; *око в очите* на някого в чьих-л. глазах; *между чётири очи* с глазу на глаз; *право очите* прямо в глаза; *с просто око* невооружённым глазом; *на око* на глазок; *лошо око* дурной глаз; *набито око* намётанный глаз; *зорко око* хорошее зрение; *сляпо око* анат. висок; *яица на очи* глазунья; *бия на (в) очи* бросаться в глаза; *хвърлям се в очи* бросаться в глаза; *хвърлям прах в очите* пускать пыль в глаза; *затварям си очите* закрывать глаза (на что-л.); *замазвам очите (на някого)* замазывать глаза (кому-л.); *не ми го хващат очите* он мне не нравится; *правя мили очи* угождать, подлизываться; *гледам с добро око (някого)* относиться хорошо (к кому-л.); *имам дебели очи* быть нахальным; *склопвам очи* умирать; *пуля очи* таращить глаза; *черно ми е пред очите* у меня потемнело в глазах.

Далее следует существительное *време* 'время' (19 фразеологизмов): *работно време* рабочее время; *утре по това време* завтра в это время; *в същото време* в то же время; *за известно време* на время; *с течение на времето* с течением времени; *време да (си) вървя* пора идти (ехать); *нямам време* у меня нет времени, мне некогда; *прекарвам времето* проводить время; *топло (горещо, студено) време* теплая (жаркая, холодная) погода; *лошо (хубаво) време* плохая (хорошая) погода; *на време* вовремя; *от време на време* время от времени; *прекарвамо времето* время проведения; *уречено време* условленное время; *лятно време* летнее время; *четирите годишни времена* четыре времени года; *по времето на* во времена, в эпоху; *време за сеитба* благоприятное время для сева; *време за път* пора в путь.

По 15 фразеологизмов имеют лексемы *име* 'имя', *вързка* 'завязка, связка', *правя* 'делать' и *лице* 'лицо, особа'.

В синтагматическое ядро можно включить все лексемы, имеющие вес, отличный от нуля. Это множество составляют 1978 слов.

Третьим параметром характеристики лексики является эпидигматический. В идеале эпидигматические связи слов должны учиты-

вать как внутрисловные, так и межсловные деривационные связи, но лексикографическая реальность накладывает на это пожелание жесткие ограничения. Отношения выводимости по двуязычному словарю мы можем оценить только на примере внутрисловной деривации. Таким косвенным показателем эпидигматической активности слов является их многозначность. Чем больше значений у слова, тем богаче его эпидигматические связи и, следовательно, тем надёжнее оно укоренено в лексико-семантической системе языка.

Характеристика болгарской лексики по эпидигматическому параметру представлена табл. 3.

Таблица 3

Распределение болгарских слов по эпидигматическому параметру

Значений	Слов	Накопл.	Э-вес
15	2	2	0,99995
14	1	3	0,99993
13	1	4	0,99991
12	3	7	0,99984
10	7	14	0,99968
9	15	29	0,99935
8	20	49	0,99890
7	24	73	0,99836
6	102	175	0,99606
5	165	340	0,99234
4	208	548	0,98765
3	1120	1668	0,96242
2	6249	7917	0,82161
1	36464	44381	0,00000

Максимальный вес имеют 2 слова с количеством значений 15. Это слово *излизам*: 1) выходить; 2) уходить; 3) появляться; 4) высываться; 5) переставать заниматься; 6) выходить из печати; 7) оказываться; 8) быть родом, происходить; 9) выступать, высказываться; 10) становиться; 11) *перен.* исходить; 12) *перен.* разг. подходить, приходить; 13) получаться; 14) *перен.* стоять, обходиться; 15) выходить, наступать и *падам*: 1) падать; 2) падать, выпадать; 3) выпадать, вываливаться; 4) падать, ниспадать; 5) падать, спускаться; 6) спускаться, наступать; 7) падать,

понижаться; 8) падать, приходиться; 9) попадать; 10) погибать; 11) *перен.* падать, ложиться; 12) *перен.* отпадать, терять силу; 13) *перен.* впадать, приходиться; 14) *разг.* быть, являться; 15) *разг.* быть, случаться. Далее следуют лексемы **минавам** (14 значений), и **пускам** (13 значений).

Эпидигматическое ядро составляют 1668 слов, с количеством значений от 3 и больше.

И наконец, последний параметр – парадигматический. О количестве синонимов двуязычный словарь информации не дает, поэтому мы воспользуемся методом позиционного анализа синонимии подробно описанного в [8].

Распределение слов по числу синонимов представлено в табл. 4.

Таблица 4

Парадигматическая характеристика болгарской лексики

Синон.	Рядов	Накопл.	П-вес
10	2	20	0,99963
8	4	52	0,99904
7	6	94	0,99826
6	10	154	0,99716
5	33	319	0,99411
4	177	1027	0,98104
3	664	3019	0,94426
2	2938	8895	0,83578
1	45270	54165	0,00000

Как видим, в ядро попадут слова с максимальным количеством синонимов в интервале от 4 до 10. Размер ядра, соответственно, 1027 слов. Ниже приведены примеры доминантных рядов.

Болтун (10 синонимов): бърбривец; бърбарко; бърболан; лаладжия; плямпало; лапацало; устатник; дрънкало; плескало; хлапацало. **Бормотать** (10 синонимов): гъкам; избръщолевям; измъмрям; ломотя; избълбалвам; пробъбрям; избърбарвам; каканижа; дърдоря; мотолева. **Броситься бежать** (8 синонимов): драена; пръпна; пръхна; тирнасе; гъртя; урнасе; юрнасе; засиям.

Слова, вошедшие во все четыре ядра, составляют так называемое малое ядро; слова, набравшие максимальные вес по трем пара-

метрам – большое ядро, по двум – периферию большого ядра. Слова, попавшие только в одно из ядер интереса для дальнейших исследований, возможно, не представляют.

Для болгарского языка были получены следующие данные: в малое параметрическое ядро вошли 37 слов (таблица 5), в большое – 117, в большую периферию – 1126 слов.

На основании полученных данных можно выделить доминанту лексико-семантической системы болгарского языка – это лексема ред (ряд, вереница), имеющая максимальный параметрический вес (3,98327). Интересно, что полученная доминанта совпадает с доминантой, выделенной в результате исследования болгарско-русского словаря меньшего объема [9], а вице-доминантой является лексема *вадя* со значением “вынимать, доставать”.

СПИСОК ЛИТЕРАТУРЫ

1. *Титов В.Т.* Общая квантитативная лексикология романских языков / В.Т. Титов. – Воронеж: Изд-во ВГУ, 2002.
2. *Титов В.Т.* Частная квантитативная лексикология романских языков / В.Т. Титов. – Воронеж: Изд-во ВГУ, 2004.
3. *Воржева И.В.* Частная квантитативная лексикология тофаларского языка / И.В. Воржева // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2006. – № 2. – С. 130–134.
4. *Терентьева И.А.* Частная квантитативная лексикология польского языка / И.А. Терентьева // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2006. – № 2. – С. 152–155.
5. *Селезнев Г.Д.* Квантитативный анализ лексики верхнелужицкого языка / Г.Д. Селезнев, А.А. Кретов // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2009. – № 1. – С. 73–76.
6. *Багина Д.В.* Параметрическая стратификация исландской лексики / Д.В. Багина, О.М. Воевудская // Вестник Воронежского государственного университета. Серия: Системный

Малое лексико-семантическое ядро болгарской лексики

ЛЕММА	ЗНАЧЕНИЕ	Д	П	С	Ф	Сумма
ред	ряд, вереница	0,99984	0,99716	0,99732	0,98895	3,98327
вадя	вынимать, доставать	0,99836	0,99716	0,99536	0,98895	3,97983
стил	стиль	0,98765	0,98104	0,99732	0,98895	3,95496
убит	убитый	0,98765	0,98104	0,99536	0,98895	3,95300
вид	вид, внешность	0,96242	0,99606	0,99536	0,98895	3,94279
век	век, столетие	0,96242	0,99411	0,99536	0,98895	3,94084
вря	кипеть	0,96242	0,99411	0,99536	0,98895	3,94084
пазя	стеречь, охранять	0,96242	0,99411	0,99536	0,98895	3,94084
вкус	вкус	0,96242	0,99411	0,99536	0,98895	3,94084
сляп	слепой, незрячий	0,96242	0,99411	0,99536	0,98895	3,94084
върл	злой, жестокий, лютый	0,96242	0,98104	0,99838	0,98895	3,93079
лице	лицо	0,96242	0,98104	0,99536	0,98895	3,92777
губя	терять, ронять	0,99234	0,94426	0,99536	0,98895	3,92091
край	конец	0,99234	0,94426	0,99536	0,98895	3,92091
ход	ход, движение, ходьба	0,99234	0,94426	0,99536	0,98895	3,92091
бера	брать, собирать	0,98765	0,94426	0,99838	0,95564	3,91924
кал	грязь	0,98765	0,94426	0,99732	0,98895	3,91818
лют	крепкий, острый	0,98765	0,94426	0,99536	0,98895	3,91622
жив	живой	0,96242	0,94426	0,99926	0,98895	3,89489
бърз	быстрый	0,96242	0,94426	0,99910	0,98895	3,89473
род	род, поколение	0,96242	0,94426	0,99536	0,98895	3,89099
шев	шов	0,96242	0,94426	0,99536	0,98895	3,89099
врѣх	верх, верхняя часть	0,96242	0,94426	0,99536	0,98895	3,89099
въже	верёвка	0,96242	0,94426	0,99536	0,98895	3,89099
грѣм	гром	0,96242	0,94426	0,99536	0,98895	3,89099
част	часть, доля	0,96242	0,94426	0,99536	0,95564	3,89099
бял	белый	0,96242	0,94426	0,99536	0,98895	3,89099
зло	зло	0,96242	0,94426	0,99536	0,98895	3,89099
глас	голос	0,96242	0,94426	0,99838	0,95564	3,86070
меря	мерить, измерять	0,96242	0,94426	0,99536	0,95564	3,85768
удар	удар	0,96242	0,94426	0,99536	0,95564	3,85763
пиша	писать	0,96242	0,94426	0,99536	0,95564	3,85763
мир	спокойствие, тишина	0,99935	0,98104	0,99536	0,98896	3,85083
заря	заря	0,96242	0,99411	0,99926	0,95564	3,83087
кѣт	угол	0,98765	0,94426	0,99536	0,99950	3,80235
игла	игла, иголка	0,99836	0,99411	0,99883	0,95564	3,74658
чист	чистый	0,96242	0,94426	0,99536	0,95564	3,65732

анализ и информационные технологии. – 2011. – № 1. – С. 184–188.

7. *Бернштейн С.Б.* Болгарско-русский словарь: Около 58 000 слов / С.Б. Бернштейн. – 3-е изд., стереотип. – М.: Рус.яз., 1986.

8. *Кретов А.А.* Позиционный анализ парадигматики романско-русских словарей / А.А. Кретов, В.Т.Титов // Проблемы компьютерной лингвистики: сб. научных трудов. Воронеж: ИПЦ ВГУ, 2008. – Вып. 3. – С. 266–282.

9. *Меркулова И.А.* Лексико-семантическая доминанта как типологический признак (на материале славянских языков) / И.А. Меркулова // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2009. – № 2. – С. 22–25.

Меркулова Инна Александровна – доцент, Воронежский государственный университет, факультет романо-германской филологии, кафедра теоретической и прикладной лингвистики. E-mail: igell@andex.ru

Merkulova Inna – Associate Professor, Voronezh State University, Romance and Germanic Philology Faculty, Department of the Theoretical and Computational Linguistics
E-mail: igell@andex.ru