

ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ ЛЕКСИКИ ЯЗЫКОВ СТРАН ПРИБАЛТИКИ

Д. С. Воевудский

Воронежский государственный университет

Поступила в редакцию 2 октября 2013 г.

Аннотация: *статья посвящена параметрическому анализу лексики государственных языков стран Прибалтики – латышского, литовского и эстонского, их сравнению; целью исследования является выделение их лексико-семантических ядер. Исследование основано на теоретической концепции параметрического анализа, предложенного В. Т. Титовым.*

Ключевые слова: *квантитативная лексикология, длина слов, синтагматическая активность слов, синонимический ряд, многозначность, латышский язык, литовский язык, эстонский язык.*

Abstract: *the article deals with the parametric analysis of the vocabulary of the Baltic states languages – Latvian, Lithuanian and Estonian – and their comparison. The aim of the investigation is to single out the lexico-semantic nuclei of these languages. The work is based on the theoretical conception of parametric analysis suggested by V. T. Titov.*

Key words: *quantitative lexicology, length of words, syntagmatic activity of words, synonymic set, polysemy, the Latvian language, the Lithuanian language, the Estonian language.*

Исследование балтийских языков началось с XVII в., когда появились первые словари и опыты грамматического описания отдельных языков: грамматика Д. Клейна и словарь К. Сирвидаса (для литовского языка), грамматика Г. Адольфи и словари Х. Фюрекера и Я. Лангия (для латышского языка). Новый этап начинается с середины XIX в., когда труды Р. К. Раска, Ф. Боппа, А. Ф. Пота и др. вводят балтийские языки в русло сравнительно-исторического языкознания и индоевропеистики: в это время появляются труды по литовскому языку (А. Шлейхер), латышскому (А. Биленштейн). В последующие десятилетия сравнительно-историческое изучение балтийских языков стало господствующим в балтийском языкознании. Большое значение для изучения истории литовского, латышского, вымерших балтийских языков, их лексического состава имели труды К. Буги, Р. Траутмана («Балто-славянский словарь»), Ю. Герулиса, Э. Френкеля («Литовский этимологический словарь») и др. Новый этап в развитии балтистики связан с созданием фундаментальных трудов по лексикологии и диалектологии, в частности диалектологических атласов, по описательной грамматике и истории балтийских языков, по топонимике и ономастике [1].

Территориально к языкам Прибалтики относятся и эстонский язык, хотя он принадлежит к южной группе прибалтийско-финской ветви финно-угорской семьи языков. Первые грамматики и словари эстонского языка, как и балтийских языков, появились в

XVII в. (Х. Шталь, И. Гутслафф, Г. Гёзекен). С возникновением сравнительно-исторического языкознания изучением германских и балтийских контактов прибалтийско-финских языков занимались В. Томсен и И. Ю. Миккола [2].

В настоящее время латышский и литовский языки являются единственными ныне «живыми» среди балтийских языков. Это государственные языки Латвии и Литвы, на них говорят около 1,8 млн человек и более 3 млн являются носителями. Эстонский язык насчитывает более 1,1 млн носителей. Предлагаемое исследование посвящено квантитативному анализу лексики этих государственных языков стран Прибалтики.

Статистические методы исследования словарей национальных языков являются действенными инструментами познания лексико-семантической системы языка. Параметрический анализ лексики, предложенный В. Т. Титовым в его монографиях «Общая квантитативная лексикология романских языков» и «Частная квантитативная лексикология романских языков» [3; 4], дает возможность познать лексико-семантическую систему языка, анализируя ее по четырем параметрам – функциональному (длина слов), синтагматическому (количество фразеосочетаний и иллюстративных примеров с данным словом в словарной статье), парадигматическому (максимальный размер синонимического ряда, в который входит данное слово) и эпидигматическому (количество значений). Операции над этими множествами позволяют выделить два ядра: большое (представленное единицами, входящими в не менее чем

одно из ядер) и малое (представленное единицами, входящими во все ядра). Выделение малого ядра – главный результат, который можно получить с помощью параметрического анализа лексики (далее ПАЛ).

Материалом для исследования послужили латышско-русский словарь объемом около 52 000 слов [5], литовско-русский словарь объемом около 19 000 слов [6] и эстонско-русский словарь объемом около 42 000 слов [7], полученные в результате превращения электронных версий словарей-источников в базы данных*.

Функциональная стратификация лексики языков стран Прибалтики

Показателем функциональной активности слова является его длина в буквах. Известно, что частотность слов обратно пропорциональна их длине: чем короче слово, тем чаще оно употребляется, и наоборот [8]. Поскольку мы изучаем лексико-семантический

уровень языка, то из дальнейшего рассмотрения исключим местоимения, числительные, наречия и служебные слова.

Кроме того, более достоверным показателем являются данные о длине слов в звуках, поскольку именно звуковая форма является первичной реальностью языка. После обработки имеющихся данных по правилам чтения эстонского [9, с. 14–22], латышского [10, с. 7–15] и литовского [11, с. 6–8] языков было подсчитано общее количество букв и звуков и вычислен коэффициент совершенства орфографии (КоСоГ), который оказался примерно равным 0,95 (для эстонского), 0,98 (для латышского) и 0,99 (для литовского) языков. Эти значения показывают, что расхождение в длинах слов, выраженных разными способами, составляет от 1 до 5 %.

Распределение лексики языков стран Прибалтики представлено в табл. 1.

Т а б л и ц а 1

Распределение полнзначной лексики языков стран Прибалтики по длине слов в звуках

Звуков	Латышский	Литовский	Эстонский	Латышский	Литовский	Эстонский
	Количество слов данной длины			Доля слов данной длины		
1	0	0	3	0,00	0,00	0,01
2	10	1	65	0,02	0,01	0,16
3	146	8	852	0,29	0,04	2,13
4	1174	203	2104	2,32	1,08	5,25
5	2464	831	2493	4,86	4,41	6,22
6	4654	1657	3692	9,18	8,80	9,22
7	6766	2990	4919	13,35	15,88	12,28
8	7985	3437	5764	15,75	18,25	14,39
9	7697	3226	5429	15,18	17,13	13,55
10	6884	2605	4089	13,58	13,84	10,21
11	5175	1742	3409	10,21	9,25	8,51
12	3284	1067	2516	6,48	5,67	6,28
13	2024	587	1765	3,99	3,12	4,41
14	1159	282	1187	2,29	1,50	2,96
15	649	126	719	1,28	0,67	1,79
16	321	39	480	0,63	0,21	1,20
17	148	18	269	0,29	0,10	0,67
18	83	7	134	0,16	0,04	0,33
19	42	4	88	0,08	0,02	0,22
20	21	0	39	0,04	0,00	0,10
21	8	1	19	0,02	0,01	0,05
22	1	0	13	0,00	0,00	0,03
23	0	0	5	0,00	0,00	0,01
24	0	0	5	0,00	0,00	0,01
25	0	0	3	0,00	0,00	0,01
26	0	0	1	0,00	0,00	0,00
27	0	0	1	0,00	0,00	0,00
28	0	0	1	0,00	0,00	0,00
29	0	0	1	0,00	0,00	0,00
Всего:	50 695	18 831	40 065			

* Сведения по литовскому языку взяты из дипломной работы Ю. И. Макаровой (2008 г.), научный руководитель – профессор А. А. Кретов.

Таблица 2

Синтагматическая активность лексики языков стран Прибалтики

ФС	Лат	Лит	Эст	ФС	Лат	Лит	Эст
82	1	0	0	29	3	0	0
78	1	0	0	28	6	0	0
71	1	0	0	27	5	0	0
70	1	0	0	26	6	0	0
68	1	0	0	25	10	0	0
66	1	0	0	24	9	0	0
65	1	0	0	23	11	0	0
62	1	0	0	22	14	0	0
60	2	0	0	21	21	0	0
58	1	0	0	20	18	0	3
57	1	0	0	19	21	0	1
56	1	0	0	18	17	0	2
55	1	0	0	17	26	0	1
54	1	0	0	16	22	0	1
51	3	0	0	15	33	0	2
50	1	0	0	14	42	0	14
49	2	0	0	13	39	1	7
44	2	0	0	12	62	1	2
43	0	0	1	11	60	2	9
42	2	0	0	10	117	0	10
41	3	0	0	9	145	1	9
40	2	0	0	8	181	4	13
39	1	0	0	7	271	3	18
38	2	0	0	6	402	8	36
36	1	0	0	5	666	12	78
35	4	0	0	4	1151	19	96
34	3	0	0	3	2009	72	211
33	0	0	1	2	4535	291	492
32	6	0	0	1	12 202	1294	1689
31	3	0	0	0	28 535	17 124	37 368
30	6	0	1				

В функциональные ядра будут включены слова, длиной до 5 звуков включительно в латышском и литовском языках, т.е. 3794 и 1043 слова соответственно, и длиной до 4 звуков – в эстонском языке, т.е. 3024 слова.

Наибольшее количество слов во всех трех языках приходится на 8 звуков. Средняя длина слова в эстонском языке составляет 8,79 букв, латышском – 8,95 букв, в литовском – 8,76 букв.

Обращает на себя внимание тот факт, что самым длинным словом оказалось слово из эстонского языка – *toiduainetetööstuseministeerium* ‘министерство пищевой промышленности’ (29 звуков), максимальная длина слова в латышском языке составляет 22 звука – *rentgenkinematogrāfija* ‘рентгенокинематография’, в литовском языке – слово длиной в 21 звук ‘контрреволюционер’. Эти слова представляют собой терминологическую лексику и, по сути, являются лексикализованными словосочетаниями, использующимися как одно слово.

Следует отметить, что эстонский язык вообще отличается большим количеством сложных слов, которые могут состоять из двух, трех и более компонентов [9, с. 6].

Самыми короткими оказались: в эстонском языке – слово *a* ‘год’; слова-омонимы *õö* 1 ‘ночь’ и *õö* 2 ‘ночной’; в латышском – 10 двузвучных слов, из которых 5 слов-омонимов: *ass* 1 ‘быстрый’, *ass* 2 ‘острый’, *ass* 3 ‘ось’, *ass* 4 ‘сажень’, *ass* 5 ‘ас’; экзистенциальный глагол *ir* ‘быть’, прилагательное *īss* ‘короткий’, существительное *pa* ‘па’ (танц.), а также два слова, обозначающие насекомых-паразитов – *ods* ‘комар’ и *uts* ‘вошь’; в литовском языке – двузвучное существительное *us* ‘подчиненный’. На наш взгляд, слова *ass* 3 ‘ось’, *ass* 5 ‘ас’ и *pa* ‘па’ (танц.) можно отнести к общекультурным заимствованиям.

Синтагматическая стратификация лексики языков стран Прибалтики

Оценка синтагматического веса лексики производится путем подсчета фразеосочетаний и иллюстративных примеров в заромбовой части словарной статьи. В латышско-русском словаре таких слов оказалось 22 160, и объединяют они 52 828 фразеологизмов и иллюстративных примеров, в литовско-русском – 1708 синтагматически активных слов, которые объединяют 2386 фразеологизмов, в эстонско-русском – 2697 слов с 5401 фразеологизмом (табл. 2).

Самое большое количество присловных фразеосочетаний оказалось сразу в двух языках у одного и того же существительного ‘рука’, называющего главный рабочий орган человека, – в латышском *roka* (82) и в эстонском *käsi* (43); в литовском – у существительного *galva* ‘голова’ (13). Таким образом, наи-

более синтагматически активной оказалась соматическая лексика, которая носит явно антропоцентрический характер и связана с активными действиями по преобразованию окружающей действительности (рука) или с осмыслением человеком своего места в мире (голова). Довольно заметный разрыв в количестве фразеосочетаний между латышским и эстонским, с одной стороны, и литовским, с другой стороны, может быть объяснен разницей в объемах словарей.

В синтагматическое ядро латышского языка будут включены все слова, имеющие при себе три и более фразеосочетаний, т.е. 5423 слова, в синтагматическое ядро литовского и эстонского языков – все слова, имеющие присловные фразеосочетания и иллюстративные примеры, т.е. 1708 и 2697 слов.

Парадигматическая стратификация лексики языков стран Прибалтики

Т а б л и ц а 4

После анализа синонимических рядов лексики языков стран Прибалтики были получены следующие результаты (табл. 3).

Т а б л и ц а 3

Показатели парадигматической активности лексики языков стран Прибалтики

	Латышский	Литовский	Эстонский
Синонимов	Синонимических рядов		
11	0	0	1
10	0	0	1
9	0	0	2
8	0	0	1
7	1	0	6
6	2	1	29
5	27	8	97
4	52	17	486
3	167	145	1192
2	2820	1130	4374
1	52 937	16 483	41 122

В латышском языке наибольшее количество синонимов (7) зафиксировано в ряду со значением 'блеснуть, сверкнуть': *paspidet, nomirgot, ieplaiksniēties, iespīguļoties, nozībsnit, pavizet, uzmirzēt*.

В литовском языке наибольшее количество синонимов (6) зафиксировано в ряду со значением 'развитие': *plītimas, ugdymas, vystymas, lavinimas, plėtra, raida*.

В эстонском языке наибольшее количество синонимов (11) зафиксировано в ряду с доминантой 'проворный': *elav, hakkaja, kārme, kergejalgne, krābe, nobe, tragi, väle, varmas, viigas, vilka*.

Эпидигматическая стратификация лексики языков стран Прибалтики

Учет эпидигматического потенциала слов прибалтийских языков, основанный на данных о многозначности, позволил получить следующие результаты (табл. 4).

Как видно из таблицы, при выделении эпидигматических ядер в латышском и литовском языках следует использовать все многозначные слова (6278 и 1596 соответственно), в эстонском языке – слова с 3 и более значениями, т.е. 3812 слов.

Максимальный вес по данному параметру в латышском и литовском языках имеют глаголы со сходной семантикой – перемещение в пространстве:

– в латышском: *kāpta* 1) подниматься, 2) взбираться, 3) повышаться, 4) расти, 5) спускаться, 6) сходить,

Показатели эпидигматической активности лексики языков стран Прибалтики

Значений	Латышский	Литовский	Эстонский
15	1	0	0
11	0	0	3
10	5	0	6
9	6	0	12
8	7	0	35
7	22	2	79
6	37	7	213
5	126	12	402
4	313	53	940
3	1065	234	2122
2	4696	1288	5437
1	44 417	17 235	30 816

7) пойти, 8) слезать (например, с воза), 9) переступить (через что-л.), 10) лезть, 11) перелезть (через что-л.), 12) выходить (из чего-л.), 13) сходить (с чего-л.), 14) вылезать (из чего-л.), 15) идти;

– в литовском: *kėlti* 1) поднимать; 2) повышать; 3) будить; 4) открывать, закрывать; 5) переправлять, перемещать, выселять; 6) устраивать; 7) возбуждать; и *padėti* 1) положить, класть; 2) поместить, помещать; 3) поставить, ставить; 4) снести; 5) подложить, подкладывать; 6) деть, девать; 7) помочь, помогать, содействовать.

В эстонском языке верхушку эпидигматического ядра представляют два глагола и одно прилагательное с явно «социальной» окраской:

teotama 1) порочить, 2) бесчестить, 3) бесславить, 4) чернить, 5) грязнить, 6) пачкать, мараить, 7) пятнать, 8) срамить, 9) позорить, 10) хаять, 11) хулить;

uhkeldama 1) задаваться, 2) важничать (разг.), 3) шеголять (разг.), 4) форситься, 5) задирать или поднимать нос, 6) ходить гоголем, 7) держать фасон (разг.), 8) кичиться, 9) чваниться, 10) пыжиться, 11) зазнаваться;

vilets 1) бедный, 2) скудный, 3) убогий, 4) жалкий, 5) плохой, 6) дрянной, 7) плачевный, 8) мизерный, 9) нищенский, 10) невидный, 11) скверный.

Параметрическая стратификация лексики языков стран Прибалтики

Итак, в результате анализа лексики языков стран Прибалтики по четырем параметрам были получены четыре ядра по каждому из них. Результаты представлены в табл. 5.

Затем в результате слияния частнопараметрических ядер (далее ЧПЯ) получаем четыре множества: 1) малое параметрическое ядро словаря каждого из языков (лексика, вошедшая во все четыре ядра);

Т а б л и ц а 5

Параметрическая стратификация лексики языков стран Прибалтики

	4 параметра	3 параметра	2 параметра	Нерелевантная лексика
Латышский	105	846	3400	12 229
Литовский	86	59	611	5297
Эстонский	63	434	1943	10 361

2) большое параметрическое ядро (лексика, вошедшая в три ЧПЯ); 3) периферия большого параметрического ядра (лексика, вошедшая в два ЧПЯ); 4) «нерелевантная лексика» (лексика, вошедшая в одно ЧПЯ).

Слово с максимальным значением параметрического веса будет считаться доминантой лексико-семантической системы. В латышском языке – это существительное *mala* ‘край, берег, сторона’; в литовском – глагол *deti* ‘класть, помещать, ставить’; в эстонском – существительное *maa* ‘земля’. Доминанта эстонского языка приобретает особое значение с учетом того, что «эстонцы называли себя *maarahvas* – букв. ‘народ земли’, а свой язык — *maakeel* – букв. ‘язык земли» [12, с. 115]. Слово, занимающее вторую позицию, – вице-доминанта. Для латышского языка – это прилагательное *ass* ‘острый’; для литовского – существительное *dalis* ‘участь, доля, судьба’; для эстонского – прилагательное *tihed* ‘плотный’.

Кроме доминант и вице-доминант, в первой «десятке» малых ядер словарей языков стран Прибалтики оказались:

– для латышского: *tapa* ‘втулка, затычка, штырь’; *sīks* ‘мелкий’; *sīvs* ‘острый’; *kāpta* ‘подниматься’; *sens* ‘старинный’; *kalps* ‘слуга’; *kalt* ‘ковать’; *sist* ‘бить’;

– для литовского: *tureti* ‘иметь, держать, содержать’; *srove* ‘течение, поток’; *verstis* ‘валить’, *vykti* ‘направляться, следовать’; *juosta* ‘пояс, полоса’; *vardas* ‘имя, название’; *tauta* ‘народ, нация’; *tverti* ‘хватать, огораживать’;

– для эстонского: *kuju* ‘образ, вид’, *viis* ‘образ’; *tüli* ‘ссора’; *nägu* ‘лицо’; *hale* ‘жалкий’; *alus* ‘основание’; *himu* ‘охота’; *kipp* ‘банка’.

Опираясь на результаты, полученные объективными методами статистического подсчета, можно переходить на качественный уровень исследования лексико-семантической систем языков.

Воронежский государственный университет

Воевудский Д. С., методист научно-методического центра компьютерной лингвистики

E-mail: dimavoev@mail.ru

Тел.: 8-906-581-69-31

Voronezh State University

Voevudskiy D. S., Methodologist of the Scientific Centre of Computational Linguistics

E-mail: dimavoev@mail.ru

Tel.: 8-906-581-69-31

ЛИТЕРАТУРА

1. Языки мира. Балтийские языки / Рос. акад. наук, Ин-т языкознания; редкол.: В. Н. Топоров [и др.]. – М.: Academia, 2006. – 221 с.
2. Языки народов СССР: в 5 т. / гл. ред. В. В. Виноградов. – Т. 3: Финно-угорские и самодийские языки. – М.: Наука, 1966. – 464 с.
3. Титов В. Т. Общая квантитативная лексикология романских языков / В. Т. Титов. – Воронеж: Изд-во Воронеж. гос. ун-та, 2002. – 240 с.
4. Титов В. Т. Частная квантитативная лексикология романских языков: монография / В. Т. Титов. – Воронеж: Изд-во Воронеж. гос. ун-та, 2004. – 552 с.
5. Латышско-русский словарь: в 2 т.: около 53 000 слов / М. Бейтиня, А. Дарбиня, Ф. Марцинкевич и др. – Рига: Лиесма: Авотс, 1979–1981. – Т. 1: А–М. – 699 с. – Рига. – Т. 2: N–Ž. – 755 с.
6. Либерис А. Литовско-русский словарь (для школ): около 22 000 слов / А. Либерис, В. Косухин. – Вильнюс: Гос. Изд-во политех. и науч. лит-ры Литовской ССР, 1956. – 392 с.
7. Тамм Й. Эстонско-русский словарь / Й. Тамм. – Таллин: Валгус, 1974. – 768 с.
8. Zipf G. K. The Psycho-Biology of Language: an introduction to dynamic philology / Zipf G. K. – Cambridge: Mass. MIT Press, 1965. – 336 p.
9. Валмет А. Учебник эстонского языка / А. Валмет, Э. Ууспылд, Э. Туру. – Таллин: Валгус, 1981. – 504 с.
10. Векслер Б. Х. Латышский язык: самоучитель / Б. Х. Векслер, В. А. Юрик. – Рига: Звайгзне, 1978. – 465 с.
11. Александрявичус Ю. Литовский язык: учебник / Ю. Александрявичус. – Вильнюс: Мокслас, 1984. – 526 с.
12. Языки мира. Уральские языки. М.: Наука, 1993. – 398 с.