

## СРАВНЕНИЕ НЕСРАВНИМОГО: К ВОПРОСУ О СОПОСТАВЛЕНИИ ЧАСТОТНОСТИ

В статье рассматриваются проблемы сопоставительного исследования частотности коррелирующих языковых единиц. Формулируются принципы создания сопоставимых корпусов текстов для обеспечения достоверности данных о частотности коррелирующих единиц в разных языках. Приводятся примеры решения исследовательских задач с помощью сопоставимых корпусов текстов.

В последние десятилетия наблюдается рост значимости количественных методов в лингвистических исследованиях. Он обусловлен, в частности, новыми возможностями, открывшимися благодаря применению компьютерных технологий для обработки больших массивов текстов. Во второй половине XX в. впервые удалось получить картину частотности употребления языковых единиц в письменной и устной речи, провести статистические исследования многих языковых явлений. В современной науке данные о частотности слов в языке имеют широкий спектр применения: они используются в методике преподавания иностранных языков [8; 14], в исследованиях эволюции лексики, определении авторского идиостиля [5], психолингвистических исследованиях процессов переработки лексической информации [4], изучении когнитивной функции языка [11, с. 21–22], в семантических исследованиях [12] и др.

Авторы опираются на данные о частотности и при сопоставлении языков. Например, А. Вежбицкая, основываясь на сравнении частотности лексем английского языка *homeland, truth, fool, stupid, stupidly, idiot, terribly, awfully, horribly* и их коррелятов в русском языке *родина, правда, дурак, глупый, глупо, идиот, ужасный, страшно*, делает вывод о разной значимости данных понятий для соответствующих лингвокультур [3, с. 30–34]. Данные об употребительности лексической единицы *上 shang* ‘верх’ в китайском языке привлекаются для обоснования особого места соответствующего полюса в китайской языковой картине пространства [10, с. 56–58].

Сопоставительные исследования частотности языковых единиц, на наш взгляд, могут пролить свет и на ряд других вопросов, касающихся эволюции языков, тенденций их развития и т.п. Однако, несмотря на несомненную перспективность, системные компаративные исследования частотности не получили пока широкого распространения. Одна из основных причин видится в недостаточной разработанности методологических основ сравнения такого рода. Речь здесь идет, прежде всего, о создании исследовательского корпуса, который бы обеспечил достоверность количественных данных. В статье излагается опыт построения и применения такого корпуса.

Сегодня корпусная лингвистика достигла таких масштабов развития, что, казалось бы, встает вопрос не о создании нового, а о выборе для исследования того или иного корпуса из числа существующих в большом количестве корпусов для самых разных языков. Наиболее известные из них – национальные корпуса текстов, например, НКРЯ для русского языка, DWDS для немецкого и т.д. Проблема заключается в том, что такие корпуса использовать для целей сопоставления можно только частично и то при соблюдении некоторых условий в силу большой разницы между ними в объеме и характере языкового материала, который вошел в эти лингвистические базы данных.

Для целей сопоставления, как правило, привлекаются параллельные корпуса текстов, в которых представлены тексты на языке оригинала и их переводы, выравненные по предложениям, абзацам и т.д. Параллельные корпуса позволяют решить целый ряд задач по выявлению спектра переводных эквивалентов, сравнению их семантики и особенностей функционирования и др. [см. также 7], однако на базе параллельных корпусов текстов получить достоверные данные об употребительности коррелирующих языковых единиц в сравниваемых языках не представляется возможным, поскольку частотность сопоставляемых единиц предопределяется текстом оригинала. Отсюда встает задача создания нового типа корпусов – сопоставимых<sup>1</sup>.

Сопоставимость корпусов, с нашей точки зрения, должна обеспечиваться следующими основными показателями: сравнимым количеством словоупотреблений, одинаковой жанровой и временной соотнесенностью текстов на сравниваемых языках. Важным фактором, обеспечивающим репрезентативность сопоставимых корпусов, является разнообразие авторства. Оптимальным источником для наполнения их содержания представляется газетный материал. Сегодня, по меньшей мере, центральные газеты, издающиеся в разных странах и на разных языках, имеют одинаковую рубрикацию (политика, экономика, спорт, культура и т.д.), они с большой точностью могут быть соотнесены по времени, а статьи пишутся большим количеством авторов.

Что касается количества словоупотреблений, т.е. объема корпуса, то оно определяется исследовательскими задачами и напрямую зависит от характера языковой единицы (лексема, синтаксическая конструкция, предложение и т.д.), частотность которой необходимо определить. Чем частотнее языковая единица, тем меньший по объему корпус необходим для получения данных о частотности. Например, для установления частотности предлогов привлекаются корпуса размером от 100 000 до 5 000 000 словоупотреблений [9, с. 18; 15, S. 2–4]. Если исследованию подвергаются менее частотные явления, например, метафоры, то в качестве источника языкового материала привлекаются многомиллионные корпуса [2].

Основываясь на изложенных принципах, мы создали сопоставимые корпуса текстов русского и немецкого языков для исследования частотности употребления предлогов с пространственным значением. В данные корпуса вошли тексты немецкой центральной газеты «Die Tageszeitung» (сокращенное название – «taz») ‘Ежедневная газета’ и тексты таких известных газет

---

<sup>1</sup> О типах корпусов текстов см. [1, с. 115–117].

России, как «Московский комсомолец», «Литературная газета», «Сегодня» и др. Объем корпусов составил 3 153 772 словоупотребления для немецкой части и 3 272 332 словоупотребления для русской. Экспериментально доказана репрезентативность обоих корпусов для исследования высокочастотных языковых единиц, в частности, предлогов (подробнее об эксперименте см. в [13, с. 75–77]).

Сопоставимые корпусы текстов предоставили возможность решить принципиально новую задачу: получить картину употребления предлогов с пространственным значением в русском и немецком языках в реальной письменной коммуникации. Частотному анализу подверглись все несоставные локативные предлоги с пространственным значением в сравниваемых языках. Их количество для немецкого языка составило 40 единиц, для русского – 47.

Анализ частотности отобранных предлогов проводился в два этапа. На первом этапе формировался частотный список на основании количества употреблений анализируемых предлогов во всей совокупности значений, т.е. не только локативного, но и темпорального, модального, цели и т.д. (табл. 1). На втором этапе составлялся частотный список употреблений предлогов только в пространственном значении, при этом высчитывался процент употребления в пространственном значении от общего количества употреблений (табл. 2).

Т а б л и ц а 1

Частотный список предлогов с пространственным значением  
в немецком и русском языках (общее количество употреблений)

№	1	2	№	1	2
1	в	117 087	1	in	83 790
2	на	49 381	2	von	34 023
3	по	22 437	3	an	25 644
4	к	16 043	4	auf	24 143
5	о	15 465	5	zu	23 830
6	из	15 112	6	aus	13 299
7	за	14 062	7	nach	12 343
8	от	12 824	8	bei	10 447
9	у	10 130	9	vor	10 054
10	до	7 018	10	um	9 505
11	с	6 522	11	über	8 884
12	при	4 405	12	durch	5 730
13	под	3 465	13	bis	5 306
14	между	2 414	14	gegen	4 345
15	через	2 071	15	neben	4 077
16	перед	1 621	16	zwischen	3 328
17	над	1 479	17	hinter	1 325
18	против	1 406	18	unter	1 106
19	около	1 003	19	gegenüber	904
20	из-за	795	20	ab	796
21	вокруг	513	21	innerhalb	413
22	внутри	252	22	außer	354

Продолжение табл. 1

23	вне	241	23	entlang	251
24	свыше	183	24	außerhalb	207
25	из-под	156	25	jenseits	204
26	мимо	146	26	inmitten	108
27	впереди	146	27	abseits	68
28	вслед	131	28	unweit	49
29	сквозь	114	29	unterhalb	38
30	возле	108	30	oberhalb	26
31	вдоль	75	31	links	20
32	навстречу	69	32	diesseits	19
33	меж	48	33	fernab	17
34	напротив	43	34	beiderseits	14
35	вблизи	38	35	rechts	13
36	посреди	35	36	längs	11
37	близ	31	37	rechterhand	5
38	поверх	27	38	weitab	2
39	позади	6	39	beidseits	1
40	подле	4	40	linkerhands	0
41	сбоку	2			
42	сзади	2			
43	посередине	2			
44	снаружи	1			
45	вглубь	1			
46	спереди	0			
47	сверху	0			
		<b>307153</b>			<b>284699</b>

Легенда: 1 – предлог, 2 – общее количество употреблений в корпусе.

Т а б л и ц а 2

Частотный список предлогов с пространственным значением  
в немецком и русском языках

№	1	2	3	№	1	2	3
1	в	34 875	29,8	1	in	46 419	55,4
2	на	11 169	22,6	2	auf	8 356	34,6
3	из	3 460	22,9	3	an	7 800	30,4
4	по	1 996	8,9	4	aus	6 348	47,7
5	за	1 268	9	5	nach	2 936	23,8
6	с	1 189	18,2	6	von	2 737	8
7	у	1 122	11,1	7	zu	2 302	9,7
8	к	906	5,6	8	durch	1 525	26,6
9	под	569	16,4	9	über	1 323	14,9
10	от	494	3,9	10	vor	1 044	10,4
11	до	376	5,4	11	bei	910	8,7
12	через	355	17,1	12	hinter	807	60,9
13	над	247	16,7	13	zwischen	764	23
14	перед	165	10,2	14	um	584	6,1
15	внутри	112	44,4	15	bis	545	10,3

16	возле	107	99,1	16	unter	492	44,5
17	между	95	3,9	17	neben	485	11,9
18	мимо	79	54,1	18	entlang	243	96,8
19	вдоль	75	100	19	gegen	144	3,3
20	вокруг	72	14	20	gegenüber	114	12,6
21	при	57	1,3	21	jenseits	112	54,9
22	сквозь	56	49,1	22	außerhalb	110	53,1
23	из-под	47	30,1	23	innerhalb	103	24,9
24	вне	42	17,4	24	inmitten	78	72,2
25	напротив	41	95,3	25	unweit	48	98
26	около	37	3,7	26	außer	23	6,5
27	вблизи	35	92,1	27	oberhalb	23	88,5
28	о	34	0,2	28	unterhalb	22	57,9
29	из-за	33	4,2	29	links	20	100
30	близ	31	100	30	abseits	18	26,5
31	навстречу	28	40,6	31	ab	17	2,1
32	поверх	19	70,4	32	diesseits	15	78,9
33	посреди	18	51,4	33	beiderseits	13	92,9
34	впереди	15	10,3	34	rechts	13	100
35	меж	14	29,2	35	längs	8	72,7
36	вслед	12	9,2	36	fernab	2	11,8
37	против	7	0,5	37	rechterhand	1	20
38	подле	4	100				
39	сзади	2	100				
40	позади	2	33,3				
41	сбоку	2	100				
42	посередине	2	100				
43	снаружи	1	100				
44	вглубь	1	100				
		<b>59271</b>				<b>86504</b>	

Легенда: 1 – предлог, 2 – количество употреблений в пространственном значении, 3 – доля употреблений в пространственном значении от общего количества употреблений.

Из полученных данных следует несколько важных выводов. Во-первых, это вывод о роли предлогов в языках разных типологических классов. Предположение о том, что степень синтетичности языка и употребительность предлогов находятся в зависимости друг от друга, по меньшей мере, на основании данных этого сопоставимого корпуса не находит подтверждения. Следуя данному предположению, роль предлогов в языках с меньшим количеством падежей возрастает, поскольку с помощью предлогов замещаются функции падежных флексий. В этом случае можно было бы ожидать, что частотность предлогов в немецкоязычной части корпуса должна быть выше, чем в корпусе русского языка. Однако установлено, что при наличии большего количества падежей в русском языке предлоги отличаются большей частотностью по общему количеству употреблений (307 153) и меньшему – по коли-

честву употреблений в пространственном значении (59 271). В немецком языке эти показатели составляют 284 699 и 86 504 употреблений соответственно.

Обращает на себя внимание несоответствие ранга частотности у большинства локативных предлогов, как по всей совокупности употреблений, так и по количеству употреблений в пространственном значении. Например, предлог *за* в общем списке частотности занимает 7-ю позицию, а его коррелят в немецком языке – предлог *hinter* – 17-ю; в списке употреблений в пространственном значении эти предлоги занимают 5-ю и 12-ю позиции соответственно. Данный факт свидетельствует о различиях в объеме семантики сравниваемых языковых единиц и о специфике пространственного значения предлогов, обозначающих одно и то же отношение.

У коррелирующих предлогов существенно отличается доля употреблений в пространственном значении в общем количестве употреблений. Так, если у наиболее частотного предлога русского языка *в* она составляет приблизительно треть (29,8 %), то для его коррелята – предлога *in* в немецком языке – более половины (55,4 %). Из этих данных можно сделать вывод о различиях в значимости пространственного значения сопоставляемых предлогов как многозначных единиц, о разной степени их вовлеченности в процессы грамматикализации и др.

Таким образом, сопоставление данных о количественных показателях употребительности пространственных предлогов в русском и немецком языках стало эффективным методом выявления специфических свойств семантики этих языковых единиц. Сущность этих свойств устанавливалась в ходе семантического анализа.

## ЛИТЕРАТУРА

1. Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – М. : Эдиториал УРСС, 2001. – 358 с.
2. Баранов, А. Н. Проблема репрезентативности корпуса данных (на примере политической метафористики) [Электронный ресурс] / А. Н. Баранов // Труды международного семинара Диалог`2001 по компьютерной лингвистике и ее приложениям. – М., 2001. – Режим доступа : <http://www.dialog-21.ru/materials/archive.asp?id=6812&y=2001-&vol=6078> – Дата доступа : 15.09.2010.
3. Вежбицкая, А. Понимание культур через посредство ключевых слов / А. Вежбицкая / пер. с англ. А. Д. Шмелева. – М. : Яз. слав. культуры, 2001. – 288 с.
4. Власов, М. С. Взаимодействие эмоциональных и лингвистических факторов в процессе переработки лексической информации (на материале имен существительных русского языка) / М. С. Власов, О. А. Сычев // Вестн. Том. гос. ун-та. – Филология. – № 52. – С. 18–52. [Электронный ресурс]. – Режим доступа : [http://journals.tsu.ru/philology/&journal\\_page=archive&id=1674&article\\_id=37690](http://journals.tsu.ru/philology/&journal_page=archive&id=1674&article_id=37690). – Дата доступа : 04.03.2020.

5. Глинкина, Л. А. Частотность как значимый регистр лексикографии и фразеологии / Л. А. Глинкина // Проблемы истории, филологии, культуры. [Электронный ресурс]. – Режим доступа : – <https://www.elibrary.ru/item.asp?id=17072384> – Дата доступа : 06.02.2019.
6. Грахольская, М. И. Семантика английских глаголов разных частотных рангов [Электронный ресурс] / М. И. Грахольская. – Режим доступа : <https://scipress.ru/philology/articles/semantika-anglijskikh-glagolov-raznykh-chastotnykh-rangov-kvantitativnyj-analiz.html> – Дата доступа : 02.04.2018.
7. Добровольский, Д. О. Корпус параллельных текстов в исследовании культурно-специфической лексики [Электронный ресурс] / Д. О. Добровольский. – Режим доступа : <http://ruscorpora.ru/sbornik2008/20.pdf>. – Дата доступа : 03.01.2011.
8. Жилина, И. А., Частотные словари как инструмент формирования иноязычной компетенции / И. А. Жилина, Л. В. Смолина // Современные проблемы гуманитарных и общественных наук. – № 5. – 2018. – С. 70–74. [Электронный ресурс]. – Режим доступа : <https://www.elibrary.ru/item.asp?id=36534589>. – Дата доступа : 12.01.2016.
9. Комарова, З. И. Ядерные служебные слова в русском подязыке информатики: квантитативно-квалитативное исследование / З. И. Комарова, С. В. Краев. – Екатеринбург : Урал. литератур. агентство, 2008. – 303 с.
10. Лю, П. Частотные характеристики языковых средств ориентации в пространстве (на материале китайского, русского и английского языков) / П. Лю // Материалы ежегод. науч. конф. студ. и магистрантов ун-та, 18–19 апр. 2019 г. : в 4 ч. Ч. 3 / отв. ред. Н. П. Баранова. – Минск : МГЛУ, 2019 – С. 56–58.
11. Норман, Б. Ю. Когнитивный синтаксис русского языка : учеб. пособие / Б. Ю. Норман. – М. : Флинта, Наука, 2013. – 254 с.
12. Тарасевич, Л. А. Семантика пространственных предлогов в зеркале количественных показателей (на материале немецкого и русского языков) / Л. А. Тарасевич // Вестн. БГУ. Сер. 4, № 2, 2014. – С. 40–47.
13. Тарасевич, Л. А. Семантика и функционирование предлогов с пространственным значением (на материале немецкого и русского языков) / Л. А. Тарасевич. – Минск : МГЛУ, 2014. – 272 с.
14. Харина, М. В. Применение частотных словарей в процессе развития англоязычной лексической компетенции студентов ИТ-направлений [Электронный ресурс] / М. В. Харина. – Режим доступа : <https://doi.org/10.21686/1818-4243-2018-5-65-73>. – Дата доступа : 24.08.2020.
15. DiMeola, C. Die Grammatikalisierung deutscher Präpositionen / C. DiMeola. – Tübingen: Stauffenburg-Verl. – 2000. – 285 S.

The article is devoted to the problems arising in comparative studies of the frequency of correlating language units. The author formulates the principles of creating comparable text corpora to ensure the reliability of data on the frequency of correlating units in different languages, provides examples of effective use of comparable text corpora to solve research problems.