УДК 81-13
DOI 10.17726/phillT.2024.1.2

CC BY

Количественный анализ ольфакторной лексики на примере русского, английского и немецкого языков[1]

Букреева Людмила Александровна,
студентка второго курса магистратуры,
Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

liudmilabukreeva@gmail.com

Велис Лолита Андреевна,
студентка второго курса магистратуры,
Пятигорский государственный университет
Пятигорск, Россия

lolla-99@mail.ru

Аннотация. В данной исследовательской работе проведен анализ коллокаций, связанных с понятиями «запах», «аромат», «вонь» и «смрад» в русском и английском языках, с использованием количественных методов и автоматической обработки языка на базе Национального корпуса русского языка (НКРЯ), корпуса английского языка (COCA) и Мангеймского корпуса для немецкого языка. Полученные статистические показатели позволяют выявить особенности употребления прилагательных, глаголов и существительных, отражающих отношение к ольфакторному опыту в английском, русском и немецком языках. Результаты дают возможность сравнить описания запахов в разных культурах и выявить тенденции в оценке ольфакторных впечатлений. Закономерности в сочетаемости ольфакторной лексики также указывают на тенденцию ключевых слов к приобретению положительной или отрицательной эмоциональной окраски за счет коллокатов.

Ключевые слова: количественные методы; автоматическая обработка языка; корпусная лингвистика; семантика запахов.

[1] Исследование выполнено за счет гранта Российского научного фонда № 24-28-00540, https://rscf.ru/project/24-28-00540. The research was funded by Russian Science Foundation No. 24-28-00540, https://rscf.ru/en/project/24-28-00540.

Quantitative analysis of olfactory vocabulary based on the example of Russian, English and German languages

Bukreeva Lyudmila A.,
2d year master student, St. Petersburg State University
St. Petersburg, Russia

liudmilabukreeva@gmail.com

Velis Lolita A.,
2d year master student, Pyatigorsk State University
Pyatigorsk, Russia

lolla-99@mail.ru

Abstract. In this research work, an analysis of collocations associated with the concepts of "smell", "aroma", "stench" and "stench" in the Russian and English languages was carried out using quantitative methods and automatic language processing on the basis of the National Corpus of the Russian Language (NCRL), corpus English (COCA) and the Mannheim Corpus for German. The obtained statistical indicators make it possible to identify the peculiarities of the use of adjectives, verbs and nouns that reflect the attitude to olfactory experience in English, Russian and German. The results allow us to compare descriptions of odors in different cultures and identify trends in the assessment of olfactory impressions. Patterns in the compatibility of olfactory vocabulary also indicate the tendency of keywords to acquire a positive or negative emotional connotation due to collocates.

Keywords: quantitative methods; automatic language processing; corpus linguistics; semantics of smells.

Введение

Развитие технологий, позволяющих проводить автоматическую обработку языка, в сочетании с корпусной лингвистикой сегодня значительно облегчает проведение исследований с применением количественного (квантитативного) метода.

Количественному методу в лингвистике противопоставляется качественный, который считается менее точным и более субъективным. Проводя сравнительный анализ двух данных методов,

А. В. Кащеева указывает на то, что количественный включает в себя «дисперсионный анализ и нормальное распределение, кластерный и факторный анализ, регрессионно-корреляционный анализ», в то время как качественный метод сосредотачивается на «исследовании изображений, письменных текстов или записей устных текстов, в ходе которых исследователь пытается выявить единый общий смысл. В лингвистике примерами качественных методов могут служить конверсационный и этнографический анализ» [1, с. 155].

Большая вариативность инструментов количественного метода и его распространенность в различных теоретических направлениях приводят к тому, что, как отмечает С. С. Верхозин, данный метод сложно соотнести с конкретными отраслями языкознания. Автор указывает на то, что применение количественных методов и в математической, и в статистической лингвистике порождают вопросы к положению такого направления как количественная лингвистика [2, с. 147]. В рамках математической лингвистики количественные методы применяются при разработке формального аппарата для описания строения языков (искусственных и естественных), в статистической лингвистике применяются вероятностатистические методы, позволяющие выделять количественные закономерности в тексте.

При проведении исследований с использованием статистических методов выделяют несколько типов переменных. Количественные переменные в данном случае являются лингвистическими единицами (фонемы, морфемы, слова, словосочетания) [3, с. 772]. Поскольку современные технологии позволяют автоматизировать процесс выполнения количественного анализа, то статистическую лингвистику часто определяют как раздел компьютерной лингвистики.

Возможность быстрой и эффективной обработки больших массивов данных за относительно короткий промежуток времени достигается за счет развития программ, предназначенных для обработки естественного языка. С целью выполнения языковой обработки применяется автоматический анализ, в котором принято различать следующие этапы: 1) графематический; 2) морфологический; 3) синтаксический; 4) семантический [4, с. 301], иногда имеет смысл выделять еще один этап – 5) прагматический [5, p. 8]. Каждый из этапов может сталкиваться с определенными неодно-

значностями, возникающими по причине сложности кодирования всех языковых особенностей.

Таким образом, применение возможностей современных технологий совместно с огромными объемами корпусов открывает большое поле для лингвистических исследований. Сегодня языковые корпусы позволяют проводить лексический анализ при помощи поиска коллокаций, под которыми в корпусной лингвистике понимаются статистически устойчивые словосочетания [6, с. 57]. Под «устойчивостью» в данном случае не подразумеваются фразеологические или идиоматические выражения, которые, однако, также могут встречаться. Важность изучения коллокатов с применением количественных методов обуславливается тем, что числовые показатели способны показать закономерности использования лексики, частотность тех или иных лексически единиц в определенных сочетаниях и контекстах. Данные результаты могут применять для сопоставления языков и выявления культурных особенностей.

Цель, материалы и методы работы

В рамках данной работы предполагается провести анализ выборки коллокаций к следующим леммам: «запах», «аромат», «вонь», «смрад» и их словарным эквивалентам в английском языке «smell», «aroma», «stink», «stench».

Полученные данные были сформированы автоматически с помощью имеющихся инструментов Национального корпуса русского языка (НКРЯ) и Корпуса английского языка (Corpus of Contemporary American English – COCA).

В Национальном корпусе русского языка для составления выборки применялся инструмент «портрет слова», из которого отбирались 15 наиболее частотных примеров. Коллокации подбирались по трем признакам: имя существительное с расстоянием, равным 4; глагол и прилагательное с расстоянием –4 и 4. Выборка сортировалась по параметру t-score, данный параметр «вычисляет частоту совместной встречаемости ключевого слова и его коллокатов» [7].

В корпусе COCA использовалась функция поиска коллокаций. Для прилагательных и глаголов было выставлено расстояние от –4 до 4. При поиске существительного для предотвращения неправильного автоматического семантического анализа (корпус

выдавал примеры, где исходное слово выступало в качестве глагола, а не существительного) было принято решение расширить поисковый запрос предлогом «of» («smell of», «aroma of» и т.д.). Это объясняется тем, что данная конструкция в английском языке (N+of+N) выполняет функцию родительного падежа в русском языке. Расстояние для поиска существительного равнялось 4. Выборка сортировалась по количеству частотных вхождений.

Результаты исследования и их обсуждение

Поскольку при формировании выборки в НКРЯ список оформлялся с помощью t-score, а не по количеству вхождений, то лексемы в русскоязычной выборке не всегда чередуются от большего количества вхождений к меньшему. Тем не менее, интересным кажется тот факт, что корпус COCA, обладающий базой данных, равной 1 млрд, имеет приблизительно одинаковое (иногда меньшее) число вхождений коллокатов в сравнении с основным корпусом НКРЯ (374 млн на 2024 год).

При анализе русско- и англоязычных прилагательных видно, что количественный показатель вхождений в русском языке уменьшается в зависимости от ключевого слова «запах» – «аромат» – «вонь» – «смрад», что может обуславливаться несколькими причинами: 1) употребимость ключевого слова («запах», будучи более нейтральным и общеупотребительным словом, используется значительно чаще «аромата», который положительно окрашен и используется ситуативно, отрицательно окрашенные «вонь» и «смрад» используются реже); 2) в русском языке видна тенденция чаще характеризовать с помощью прилагательных нейтральные или положительно окрашенные лексемы («запах», «аромат») и не описывать отрицательно окрашенные.

	НКРЯ	частотность	COCA	частотность	НКРЯ	частотность	COCA	частотность
	запах / smell – прилагательные				аромат / aroma – прилагательные			
1.	свежий	632	sweet	621	тонкий	241	sweet	140
2.	сильный	570	bad	361	нежный	158	fresh	74
3.	неприятный	504	strong	360	приятный	145	rich	73

4.	приятный	479	fresh	275	свежий	137	pungent	72
5.	острый	438	burning	249	пряный	116	wonderful	61
6.	сладкий	405	faint	207	сладкий	103	strong	54
7.	тяжелый	404	musty	173	сильный	90	delicious	52
8.	резкий	394	hot	172	теплый	80	faint	49
9.	густой	340	familiar	158	полный	83	heady	48
10.	теплый	339	warm	156	легкий	77	pleasant	45
11.	знакомый	331	sour	150	душистый	72	spicy	41
12.	крепкий	315	thick	148	густой	70	floral	31
13.	тонкий	283	sharp	147	новый	82	warm	30
14.	легкий	276	wet	144	особый	68	baked	28
15.	кислый	249	pungent	138	весенний	52	earthly	25

Квантитативность английских лексем также показывает тенденцию к снижению «smell» – «aroma» – «stink» – «stench». Однако данные таблицы дают интересные результаты: разрыв между первой и последующими лексемами с точки зрения частотности практически кратен двум при ключевых словах «smell», «aroma» и почти шести для «stink». Т.е. «stink» встречается чаще, чем «stench», только в сочетании с прилагательным «big». Из этого следует, если результаты автоматической выборки английского корпуса не ложны, что лексема «stink» имеет прочную частотную связь лишь с одним прилагательным и практически не используется в других вариациях. Такое поведение лексемы «stink» при сопоставлении с ее русским эквивалентом «вонь» может говорить в пользу того, что не всегда корректен перевод только с учетом подбора эквивалента.

Несмотря на то, что прилагательные к слову «stench» изначально имеют невысокие показатели, их уменьшение происходит не так стремительно, как у «aroma» или «stink», что может свидетельствовать о большей описательной и коннотативной вариативности лексемы «stench» и, напротив, невысокой вариативной возможности для «stink».

В сравнении с русскоязычной выборкой видно, что английские лексемы имеют тенденцию более сильного соединения с конкретным прилагательным, о чем свидетельствует разрыв между первым и последующими словами, чего не наблюдается в русскоязычной выборке. Разрыв в употреблении между нейтральным «smell» и положительно («aroma») или отрицательно («stench», «stink») окрашенными словами свойственен как русскому, так и английско-

му языку, т.е. лексема «smell» также более общеупотребима из-за своей изначальной нейтральности. Однако количество прилагательных для отрицательных слов в английском языке больше, т.е. неприятные запахи англоговорящее население описывает так же часто.

	смрад / stench – прилагательные				вонь / stink – прилагательные			
1.	нестерпимый	23	rotting	55	нестерпимый	44	big	120
2.	удушливый	20	foul	41	ужасный	34	brown	21
3.	невыносимый	20	burning	40	невыносимый	31	sweet	15
4.	тяжелый	20	overwhelming	33	кислый	30	dead	13
5.	густой	19	strong	31	страшный	28	terrible	13
6.	трупный	17	dead	29	бензиновый	19	huge	10
7.	страшный	14	awful	26	едкий	18	sharp	9
8.	отвратительный	12	acrid	24	отвратительный	18	foul	9
9.	тяжкий	12	horrible	19	густой	18	rotting	9
10.	черный	12	sweet	18	тяжелый	17	heavy	8
11.	человеческий	11	thick	18	удушливый	16	rotten	8
12.	полный	10	overpowering	18	трупный	15	sweaty	8
13.	вонючий	9	terrible	17	человеческий	14	familiar	7
14.	ужасный	9	unbearable	17	гнилой	12	chemical	7
15.	великий	9	heavy	15	жуткий	12	sour	7

Выборка глаголов показала, что, как и в случае с существительными, частотные характеристики в русском языке уменьшаются от большего к меньшему: «запах» – «аромат» – «вонь» – «смрад».

В англоязычной выборке происходят следующие изменения: коллокации к слову «smell» являются наиболее частотными и разрыв между словами (начиная от первого и заканчивая четвертым) достаточно значителен. Показатели для «aroma» падают незначительно в отличие от «stink» и «stench», однако мы видим, что по количественному употреблению глагола «stink» превосходит «aroma», что в данном случае свидетельствует о том, что воздействие отрицательного ольфакторного опыта в английском языке встречается так же часто, как и положительного, чего нет в русском языке.

Примечательно, что лидирующим в русском языке является коллокат «быть», в то время как в английской выборке его эквивалент «to be» не встречается. Можно предположить, что это опять же связано с автоматическим анализатором, который по тем или иным причинам: 1) либо распознает данный глагол в английском языке как вспомогательный (что возможно из-за грамматического строя английского языка) и включает его в список «стоп-слов», т. е. не учитывающихся при подсчете; 2) либо конструкции типа «N+to be+ N» редко используются при описании запаха, что кажется маловероятным.

	запах / smell – глагол				аромат / aroma – глагол			
1.	быть	1778	love	400	вдыхать	242	smell	129
2.	вдыхать	567	fill	299	быть	269	fill	103
3.	пахнуть	456	waft	202	источать	102	waft	81
4.	стоять	486	breathe	145	пахнуть	102	inhale	59
5.	чувствовать	451	notice	135	дышать	98	release	37
6.	почувствовать	427	mix	117	напоить	87	flavor	36
7.	иметь	444	hang	116	распространять	78	carry	30
8.	идти	445	rise	111	наполнять	76	mingle	26
9.	пропитать	379	hate	107	пропитать	68	produce	25
10.	слышать	320	remind	91	благоухать	64	cook	25
11.	исходить	302	inhale	91	наполнить	61	mix	23
12.	издавать	300	cook	90	цвести	60	bake	23
13.	распространять	268	burn	82	веять	57	enjoy	22
14.	напоминать	270	linger	76	вдохнуть	56	rise	22
15.	ощущать	233	detect	71	обладать	57	breathe	22

	смрад / stench – глагол				вонь / stink – глагол			
1.	быть	77	smell	99	быть	124	make	215
2.	стоять	41	fill	49	стоять	54	raise	85
3.	дышать	29	rise	35	нести	25	smell	48
4.	идти	25	waft	25	исходить	21	cause	20
5.	пахнуть	20	emanate	22	задыхаться	19	rise	16
6.	исходить	19	hang	20	пахнуть	18	fill	15
7.	задыхаться	18	ignore	20	стать	22	blow	13
8.	наполнять	15	carry	16	идти	20	wash	12
9.	нести	15	burn	16	пропитать	15	throw	11
10.	гореть	14	notice	14	распространять	12	hang	11
11.	гнить	12	over-power	14	чувствовать	13	notice	8

12.	наполнить	12	blow	13	тянуть	12	complain	8
13.	распространять	9	mix	13	дышать	12	burn	7
14.	подниматься	9	perme-ate	12	пойти	14	ignore	6
15.	удушать	8	emit	11	обдать	11	mask	6

Рассматривая выборку существительных, для более точных результатов НКРЯ появилась необходимость помимо метки «существительное» добавлять метку «родительного падежа», т. к. без данного параметра появляется большое количество «зашумлений», т. е. неправильно распознанного коллоката автоматическим анализатором в русском языке. Так, в списки попадали такие слова: «запах», «дух», «вкус», «цвет», «аромат» и др., что в принципе невозможно, т. к. порождает либо тавтологию, либо лишенное семантики словосочетание.

Выборка с меткой родительного падежа дала количественно более низкие результаты. При этом надо обратить внимание на некоторые погрешности анализатора:

1) слово «дух», представленное в начальной форме, при рассмотрении примеров на самом деле в большинстве своем реферировало к «духИ», т. е. парфюмерному изделию;

2) слово «цвет» также в большинстве своем относилось к слову «цветок»;

3) слово «мочь» – на самом деле неправильно подобранная начальная форма к слову «моча».

Для «смрада» не набралось 15 коллокатов. Коллокат «червь» оказался также «зашумлением», т. к. это слово часто встречалось рядом, но как однородный член предложения.

Говоря о выборке, представленной к слову «вонь», стоит отметить, что неожиданно появляющееся «благоухание» действительно встречается в текстах 6 раз, однако в одном и том же зафиксированном виде «в воню благоухания духовного», что на самом деле говорит о высокой ситуативности (религиозный дискурс) и, следовательно, нераспространенности в повседневной речи данного словосочетания.

В английском языке «aroma» и «stink» показали низкую сочетаемость с существительными, в отличие от «stench», у которого большая часть коллокатов относится к семантическому полю тела и смерти.

Помимо прямой референции к источнику запаха в английском языке также появляются такие метафорические выражения, как «smell of success», «stench of failure», «stink of exhaust», «stink of fear».

Со словом «аромат» в русском языке сочетаются различные лексемы. Большая часть из них относится к природе и растениям в обобщенном смысле – «цветок», «трава», «лес», «сад», «дерево» и т.д., реже встречаются отсылки к конкретным цветам – «роза», «фиалка». В английском «aroma» имеет больше коллокаций с напитками и едой. Данный показатель подтверждает отсутствие полных синонимов в языках, т.к. английское «aroma» тематически не соответствует в применении русской лексеме «аромат».

	запах / smell of – существительное				аромат / aroma of – существительное			
1.	**дух** (духи)	589	smoke	258	цветок	253	coffee	68
2.	трава	439	blood	192	**дух** (духи)	174	bread	24
3.	пот	410	coffee	155	трава	121	smoke	18
4.	дым	405	death	146	роза	82	food	16
5.	цветок	376	grass	126	кофе	53	flowers	15
6.	земля	354	sweat	121	лес	48	tobacco	15
7.	гарь	327	oil	109	сад	45	spices	14
8.	кровь	287	food	107	земля	42	oil	13
9.	масло	234	fish	104	**цвет** (цветы)	40	beans	12
10.	тело	216	earth	92	дерево	36	beef	12
11.	табак	208	flesh	91	фиалка	34	bacon	12
12.	мясо	198	meat	91	зелень	30	meat	12
13.	хлеб	198	hair	87	весна	29	garlic	10
14.	рыба	191	success	84	хвоя	28	turkey	10
15.	вода	197	urine	84	сено	28	wood	10

	смрад / stench of – существительное				вонь / stink of – существительное			
1.	труп	14	death	80	грязь	13	sweat	15
2.	тело	13	flesh	50	тело	13	smoke	14
3.	мясо	9	blood	33	мясо	9	fish	12
4.	**червь**	8	decay	31	труп	8	cigarettes	11
5.	разложение	8	urine	26	масло	7	fear	9
6.	дым	7	bodies	21	**благоухание**	6	oil	8
7.	испарение	6	smoke	19	нечистота	6	death	7
8.	грех	6	fish	15	уборная	6	flesh	7
9.	смерть	6	sweat	15	**мочь** (моча)	6	shit	7
10.	земля	6	garbage	14	рот	6	urine	7

11.	гниение	5	failure	12	кожа	6	exhaust	6
12.			manure	12	газ	6	failure	6
13.			rot	10	перегар	5	fuel	6
14.			sewage	10	солярка	5	breath	5
15.			corpses	9	краска	5	body	5

Для сравнения также была произведена попытка получения частотного списка слов на немецком языке. Поскольку авторам на момент исследования не были известны немецкоязычные корпуса, предоставляющие возможность подсчета частотности в интерфейсе самого корпуса, было принято решение провести статистический подсчет на базе Мангеймского корпуса немецкого языка (COSMAS) при помощи языка программирования Python и библиотеки spaCy.

Код для обработки контекста получал на вход языковой материал в виде 10 тысяч примеров со словом «Geruch» («запах»). Чтобы осуществить необходимую синтаксическую разметку текста, было необходимо для начала токенизировать текст, то есть разделить его на более мелкие единицы, в данном случае на слова. Инструменты библиотеки spaCy предоставляют простой и эффективный способ сделать это при помощи предварительно обученной немецкой языковой модели (de_core_news_sm). При употреблении синтаксической разметки происходит извлечение необходимых лексем, составляющих контекстное окружение слова «Geruch».

Прилагательное	Частотность	Глагол	Частотность	Существительное	Частотность
Beißend	194	Liegen	255	Tod	24
Unangenehm	191	Steigen	156	Mensch	19
Übler	180	Verbreiten	121	Buch	15
Stechend	117	Wahrnehmen	111	Frau	14
Streng	111	Kommen	103	Meer	14
Süßlich	97	Hängen	98	Tier	13
Stark	96	Nehmen	98	Fleisch	11
Muffig	95	Verströmen	90	Mann	10
Typisch	89	Ziehen	78	Wasser	10
Penetrant	82	Bemerken	72	Pflanze	10
Intensiv	69	Haben	67	Leiche	10
Modrig	61	Schlagen	58	Verwesung	10

Angenehm	58	Machen	51	Grauen	9
Würzig	53	Dringen	51	Geld	9
Scharf	47	Bleiben	46	Feuer	8

В таблице представлена частота использования прилагательных, глаголов и существительных в сочетании со словом «Geruch» (запах) в немецком языке. Самыми частотными являются прилагательные «beißend» (едкий, резкий, образованное от глагола «beißen» (кусать)), «unangenehm» (неприятный), «übel» (дурной) и «stechend» (острый), использующиеся для описания отрицательных впечатлений от запаха.

Самый частотный глагол, употребляемый со словом «Geruch» (запах) в немецком языке, – «liegen» (лежать). Запах, распространяющийся в воздухе, описывают следующие по частотности глаголы: «steigen» (подниматься) и «verbreiten» (распространяться).

Среди существительных частотными оказались природные источники запахов: «Meer» (море), «Tier» (животное), «Fleisch» (мясо), телесные запахи «Mensch» (человек), «Frau» (женщина). Слово «Tod» (смерть), вероятно, употребляется в контекстах, связанных с запахом разложения, которое тоже присутствует в списке частотных («Verwesung» (разложение)) наряду со словом «Leiche» (труп). Так, существительные в контекстах обнаруживают большое разнообразие источников запахов, от природных явлений до человеческой деятельности. В выборку немецких существительных, как и в русском языке, попало некоторое количество неправильно распознанных коллокатов. Например, в список попадали такие существительные, как «Geschmack» (вкус), «Farbe» (цвет) и «Lärm» (шум).

Таким образом, представленные частотные списки коллокатов могут представить нам, какими способами в немецком языке описываются запахи, включая их интенсивность, качество и происхождение.

Выводы

Применяя количественный метод в исследовании перцептивной лексики, можно говорить как о его достоинствах, так и недостатках. С одной стороны, частотные показатели позволяют сравнить большие объемы материала, выявлять тенденции и за-

висимости в разных языках. Через сопоставление становятся еще отчетливее детали, которые в прочих условиях могли бы остаться незамеченными. С другой же стороны, возникает сложность подбора и сопоставления материала.

Во-первых, тот факт, что для русской и англоязычной выборки используются разные корпуса, уже накладывает отпечаток на качество сопоставления, поскольку неизвестны ни временные, ни количественные, ни жанровые соотношения материала.

Во-вторых, автоматический машинный подсчет, очевидно, допускал ошибки, сопоставляя с ключевым словом лексемы, не имеющие с ним связи, согласно запрашиваемым синтаксическим и семантическим требованиям. Таким образом, полученные списки из пятнадцати слов являются приблизительными.

В случае с англоязычным корпусом возникает еще больше сомнений. В приведенном перечне не только неясно, на каком количестве материала была подготовлена выборка, но и вызывают недоверие количественные показатели, поскольку они часто уступают русскоязычному корпусу. При том, что объем НКРЯ меньше, чем COCA.

Несмотря на наличие погрешностей, можно говорить о том, что благодаря частотному анализу возможно сопоставление ольфакторных картин мира, отраженных в разных языках. Количественный метод позволяет утверждать, что в трех приведенных языках ольфакторный опыт имеет тенденцию преимущественно описываться с положительной или отрицательной коннотацией, т.е. при описании ольфакторных впечатлений реципиент стремится дать запаху оценку, которая достигается за счет прилагательных, экспрессивных глаголов или, что менее эмоционально выражено, существительных.

Более того, при сопоставлении англо- и русскоязычной выборки видно, что для существительных с изначально нейтральной или положительной окраской («запах», «аромат», «smell», «aroma») набор наиболее частотных прилагательных в большинстве своем схож. Лексика, применяемая с отрицательной коннотацией, чаще всего в качестве источника указывает на человека, животных, резко пахнущие растения или нечистоту, что свойственно для всех рассматриваемых в статье языков. Поэтому можно говорить о том, что ольфакторноя картина мира в русском и английском языках с точки зрения отношения к приятным и неприятным запахам от-

личается не слишком сильно, хотя и очевидны различия в употреблении слов, считающихся переводческими эквивалентами («аромат» – «aroma»).

Первичные показатели обработки немецкой выборки также показывают схожесть с англо- и русскоязычной выборкой, однако есть смысл проведения дальнейшего сопоставления по остальной ольфакторной лексике.

Литература

1. Кащеева А. В. Квантитативные и качественные методы исследования в прикладной лингвистике // Социально-экономические явления и процессы. 2013. № 3 (049). С. 155-162.
2. Верхозин С. С. О статусе количественных методов в лингвистике // Вестник Иркутского государственного лингвистического университета. 2013. № 3 (24). С. 145-150.
3. Гречачин В. А. Статистические методы в исследовании текстов // Вестник Башкирского университета. 2018. № 3. С. 772-775.
4. Кочеткова Н. А. Статистические языковые методы. Коллокации и коллигации // Новые информационные технологии в автоматизированных системах. 2013. № 16. С. 301-305.
5. Lee R. S. T. Natural Language Processing: A Textbook with Python Implementation / R. S. T. Lee, United International College: Springer Nature, 2023. 436 с.
6. Павельева Т. Ю. Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2016. № 3-4 (155-156). С. 56-61.
7. Палийчук Д. А. Корпусные технологии в изучении коллокаций (на примере сервисов «antconc» и «sketchengine») // Studia Humanitatis. 2020. № 2.