## СПИСОК ЛИТЕРАТУРЫ

1. *Азарова Н. М.* Конвергенция философского и поэтического текстов XX–XXI вв.: автореф. дис. … д-ра филол. наук. М., 2010. URL: http://plr.iling-ran.ru/ru/node/96 (дата обращения: 16.08.2017).
2. *Бонч-Осмоловская А., Орехов Б.* Некоторые применения корпусных методов к наивной поэзии. URL: http://www.ruthenia.ru/leibov_50/article_b-osm_orexov.html (дата обращения: 16.08.2017).
3. *Ляшевская О. Н.* Новый частотный словарь русской лексики / О. Н. Ляшевская, С. А. Шаров // Словари на основе Национального корпуса русского языка. URL: http://dict.ruslang.ru/freq.php (дата обращения: 16.08.2017).

## REFERENCES

1. *Azarova N. M.* Konvergentsiya filosofskogo i poeticheskogo tekstov XX–XXI vv.: avtoref. dis. … d-ra filol. nauk. M., 2010. URL: http://plr.iling-ran.ru/ru/node/96 (data obrascheniya: 16.08.2017).
2. *Bonch-Osmolovskaya A., Orehov B.* Nekotoryie primeneniya korpusnyih metodov k naivnoy poezii. URL: http://www.ruthenia.ru/leibov_50/article_b-osm_orexov.html (data obrascheniya: 16.08.2017).
3. *Lyashevskaya O. N.* Novyiy chastotnyiy slovar russkoy leksiki / O. N. Lyashevskaya, S. A. Sharov // Slovari na osnove Natsionalnogo korpusa russkogo yazyika. URL: http://dict.ruslang.ru/freq.php (data obrascheniya: 16.08.2017).



***В. Э. Рогачева***


# КОРПУСНЫЙ МЕТОД УСТАНОВЛЕНИЯ ПЕРЕВОДА ТЕРМИНОЛОГИЧЕСКИХ ЕДИНИЦ


*В статье рассматривается проблема выравнивания текста в сопоставимом двуязычном корпусе и установления переводных эквивалентов терминологических единиц, извлеченных из сопоставимого корпуса. Из русского и английского подкорпусов с помощью программы — корпусного менеджера извлекается список потенциальных терминов, затем исследуется лексический состав извлеченных терминологических единиц и предлагаются методы установления переводных эквивалентов.*



**Ключевые слова**: терминология, извлечение терминов, корпус текстов, сопоставимый корпус текстов, перевод, двуязычная лексикография, корпусная лексикография.



***V. Rogacheva***


## CORPUS-BASED METHOD OF TERMINOLOGY TRANSLATION


*The paper concerns the term systems of the domain «Terminology and knowledge Engineering», extracted from a bilingual comparable corpus, sampled of domain-oriented Russian and English texts. The paper presents an attempt to prescribe translation equivalents to terminology units, extracted from the corpus.*



**Keywords**: terminology, term extraction, text corpus, comparable corpus, translation, bilingual lexicography, corpus-based lexicography.


Статья посвящена методам анализа и сопоставления данных, извлеченных из двуязычного текстового корпуса единой предметной области; решению проблемы установления переводных эквивалентов между ключевыми единицами информации двуязычных сопоставимых корпусов — терминами. Целью исследования корпус-

ных данных является обеспечение двуязычной коммуникации между экспертами конкретной предметной области, создателями и потребителями ее материальных и нематериальных объектов.

Под корпусом текстов понимается структурированный массив текстов на естественном языке в машиночитаемом формате, доступный для обработки программными средствами, достаточно обширный и репрезентативный для проведения исследований [4]. Двуязычный текстовый корпус, охватывающий единую предметную область, содержит примеры употребления характерных для предметной области терминов в рамках специальной коммуникации. Он признан проверенным программным средством, применяемым в исследовании терминологии, управлении операциями с двуязычной терминологией и составлении двуязычных терминологических словарей [10].

В настоящее время существует проблема извлечения из сопоставимых двуязычных корпусов информации для дальнейшего исследования и обработки. Для обеспечения репрезентативности и сопоставимости двуязычного корпуса важен равный объем текста на каждом рабочем языке и выравнивание. Среди двуязычных корпусов различаются параллельные и сопоставимые корпусы. Если в параллельном корпусе сопоставляются тексты-оригиналы и тексты-переводы, выровненные по предложениям, то в сопоставимом корпусе представлены тексты единой предметной области, объема и структуры, но не являющиеся переводами друг друга [1; 8]. В сопоставимом корпусе возможно терминологическое выравнивание текстов обоих подкорпусов, предполагающее извлечение из текстов общих терминологических единиц и словосочетаний, и решение проблемы их перевода [1].

В качестве материала исследования использовались коллекции текстов единой предметной области «Терминология и извлечение знаний» на русском и английском языках. Подкорпус на английском языке включает в себя материалы конференции «Terminology and Knowledge Engineering» за 2010 и 2014 годы. Подкорпус на русском языке состоит из материалов конференции «Терминология и знание» за 2008–2012 годы*.

Исследование проводится с помощью програмы-конкондансера AntConc [2]. Ее инструментарий позволяет подсчитать общее количество токенов в корпусе, из которых могут быть выбраны потенциальные термины. Токены могут быть отсортированы в соответствии с их частотой или в алфавитном порядке. Кроме того, с помощью программы можно получить данные об объеме каждого подкорпуса в токенах и словоформах. Под токеном понимается набор символов, отделенных с двух сторон пробелами или знаками препинания. Общие токены вмещают в себя все токены исследуемого текста без учета их внутритекстовой повторяемости [7].

Общее количество токенов в английском подкорпусе — 188 649. В русском подкорпусе — 136 168.

Инструмент AntConc Wordlist позволяет подсчитывать и объединять словоформы, встречающиеся в корпусе, в единую словарную единицу. Объем частотного словаря для английского подкорпуса — 10 339 разных токенов, для русского — 25 795.

Многие единицы частотных словарей не являются терминологичными. Нетермины составляют отрицательный словарь и подлежат удалению из словарей для сокращения объема списков слов. В отрицательный словарь входят детерминаторы, вспомогательные глаголы, предлоги, союзы, местоимения, имена собственные и их производные. Кроме того, необходимо свести множество встречающихся в корпусе грамматических форм одной лексической единицы к основной форме, то есть провести лемматизацию. Для упрощения лемматизации и сопоставления терминологических единиц можно объединять однокоренные термины-дериваты в единые комплексные

словообразовательные единицы. Все слова-дериваты могут быть объединены в словообразовательное гнездо как наиболее крупную словообразовательную единицу. В рамках гнезда могут быть выделены составляющие его комплексные словообразовательные единицы: словообразовательные парадигмы (парадигмы существительных, парадигмы прилагательных) и словообразовательные цепочки [6]. Сопоставление словообразовательных гнезд на двух языках позволит ускорить поиск эквивалентов для всех дериватов, обнаружить отсутствие терминов в терминосистеме одного из языков, сделать выводы о лексическом составе лемматизированного списка терминов.

В итоговом лемматизированном списке потенциальных терминов зафиксировано 8376 единиц на русском и 5223 единицы на английском языке. Причины неравномерных величин словарей заключаются в развитой морфологической системе русского языка и способности русских слов образовывать обширные словообразовательные гнезда. К примеру, две словообразовательные цепочки «*семантика — семантический — семантизировать — семантизироваться — семантизация*» и «*semantics — semantic — semanticize — semanticization*» образованы по одному принципу. Однако не все компоненты этих цепочек представлены в корпусе. Например, русский глагол *семантизировать* в корпусе не представлен, следовательно, не является частотным для научного дискурса выбранной предметной области. Зато представлен возвратный глагол *семантизироваться*, в то время как в цепочке английских слов аналога ему нет, так как в английском языке идея возвратности не выражается специальными флексиями. В английском подкорпусе мы также не наблюдаем лексических единиц, аналогичных русским «*семантизировать — семантизироваться*». Кроме того, в русском подкорпусе можно отметить множество производных единиц, мотивированных словом *семантика*, таких как *психосемантика*, *семантико-когнитивный*, *семантикопонятийный*, *семантикоономасиологический* и *семантикопрагматический*. В английском подкорпусе аналогов данным наименованиям и признакам нет. Производных слов, мотивированных единицей *semantic*, в английском подкорпусе всего три: *lexico-semantic* и *lexical-semantic*, являющиеся синонимами, а также *non-semantic*.

Из анализа словарей можно сделать вывод о том, что образование новых терминов путем слияния основ более характерно для научных текстов на русском языке. Учитывая большую частоту слова *semantic* в английском языке (189), можно предположить, что оно часто встречается в терминологических словосочетаниях и что для английского языка более характерно выражение идеи «слияние двух предметных областей» путем синтаксической связи. За счет основ и флексий, привносящих новое значение в производные слова, словообразовательные гнезда русскоязычного словаря обширнее, что влияет на объем частотного словаря.

Говоря о лексических единицах, означающих область знания, можно связать их высокий словообразовательный потенциал и способность к формированию обширных словообразовательных гнезд с частотой их употребления и, следовательно, с разработанностью данной области знания в научном дискурсе конкретного языка. Выделение в русском подкорпусе нескольких прилагательных, производных от основы *семантический* (*семантико-когнитивный*, *семантикоономасиологический* и пр.), и отсутствие их аналогов в английском подкорпусе свидетельствуют о внимании авторов научных текстов на русском языке именно к аспектам связи семантики и других разделов филологических наук, склонности изучать семантику в ее взаимосвязи с другими науками.

Выводы о преобладании двухосновных единиц в русском языке справедливы также для других частотных слов корпуса с высоким словообразовательным потенциалом. Как и пара эквивалентов *семантика* —

*semantics*, слова *термин* и *term* являются одними из самых частотных в подкорпусах своего языка. Оба слова мотивируют многочисленные производные, которые формируют обширные гнезда. Частота слова *термин* (1994) в русском подкорпусе незначительно выше его эквивалента в английском подкорпусе (1809). Но словообразовательное гнездо производящей основы *термин* значительно больше гнезда слова *term* за счет множества двухосновных производных.

Так как одной из целей исследования является изучение потенциала сопоставимых корпусов при установлении переводных эквивалентов, интересно сравнение наиболее частотных слов открытого класса в словниках на русском и английском языках. Принимая во внимание тот факт, что пятьдесят наиболее частотных слов открытого класса в корпусе скорее всего являются терминами [9], рассмотрим возможность установления переводных эквивалентов между первыми 50-ю словами частотных словарей каждого из подкорпусов.

В пределах выборки из 100 наиболее частотных слов, принадлежащих обоим языкам, удалось установить переводные эквиваленты практически всем словарным единицам, несмотря на разный ранг этих эквивалентов в объединенном частотном словаре и разные величины частоты (табл. 1). Например, если ранг у слов *term* и *термин* одинаков, то прилагательное *семантический* в русском подкорпусе имеет значительно более высокую частоту (1193) и ранг (2), чем его аналог *semantic*. Его частота равна 189, чему соответствует ранг 38.

В выборке часто встречаются элементы единых словообразовательных гнезд. И если конкретной русской словоформе не был найден английский эквивалент, она может быть соотнесена с другими словоформами, образованными от той же производящей основы, что и предполагаемый английский эквивалент.

Учитывая то, что смысл термина раскрывается в контексте, а английское существительное, в том числе и с терминологическим значением, склонно выступать модификатором ядра именного словосочетания, можно сопоставлять терминологические сочетания с лексическими единицами выбранного гнезда в позиции ядер.

Обратимся к тем словам выборки, эквивалентов которым не было найдено. Примечательно, что практически всем русским словам были найдены эквиваленты, за исключением отдельных единиц. В соответствии с существующими методами классификации терминов, например по частям речи и по сфере распространения [3; 5], оставшиеся слова можно определить как:

– общенаучные термины-глаголы (*определить, привести, соответствовать, отнести*);

– общенаучные термины-существительные (*наука, признак, название, мир, смысл, вопрос, выражение, проблема, объект, компонент, человек*);

– межотраслевые термины-существительные (*лексика, ботаника*).

Присутствие данных слов в корпусе объясняется их широкой семантикой. Выделенные общенаучные термины могут одновременно принадлежать не только терминосистемам разных наук, но и разным языкам для специальных целей (ЯСЦ) [5], что объясняет их высокую частоту.

В выборке можно отметить термины, принадлежащие терминосистемам, не являющимся объектом данного исследования (*ботаника*). Это объясняется тем, что большинство текстов в корпусе посвящено исследованию терминологии, а терминосистемы других наук являются материалом исследования авторов текстов. Термины из терминосистем других наук приводятся в текстах как примеры, иллюстрирующие исследования авторов.

В большинстве ЯСЦ термины-существительные преобладают над терминами-глаголами, за исключением некоторых текстов по прикладным наукам, например инженерным, военным, геологическим [3].

Представленные в словаре частотные термины-глаголы на русском языке не являются узкоспециальными. Их частота объясняется тем, что в текстах они служат для организации синтаксической связи между терминами-объектами ввиду отсутствия в русском языке глаголов-связок, принадлежащих к словам закрытого класса.

Справедливы сделанные выводы и для английских слов, которым не подобраны эквиваленты в рамках выборки. В эту категорию вошли:

*Таблица 1*

**Фрагмент объединенного частотного словаря извлеченных из корпуса лексических единиц**

| *Лемма* | *Частота* | *Лемма* | *Частота* |
|---|---|---|---|
| термин | 1994 | term | 1809 |
| язык | 1211 | concept | 1163 |
| семантический | 1193 | terminology | 1147 |
| слово | 825 | use | 995 |
| словарь | 726 | language | 847 |
| понятие | 648 | data | 814 |
| терминологический | 638 | terminological | 553 |
| знание | 588 | information | 538 |
| научный | 553 | ontology | 520 |
| единица | 514 | model | 519 |
| терминология | 500 | domain | 508 |
| текст | 478 | different | 497 |
| явиться | 478 | relation | 473 |
| значение | 454 | word | 467 |
| определить | 428 | translation | 465 |
| языковой | 407 | user | 450 |
| мир | 405 | corpus | 447 |
| специальный | 388 | base | 442 |
| когнитивный | 379 | work | 413 |
| профессиональный | 341 | system | 412 |
| представить | 335 | knowledge | 397 |
| картина | 280 | text | 383 |
| концепт | 274 | process | 360 |
| исследование | 272 | type | 357 |
| процесс | 271 | provide | 356 |
| данный | 265 | document | 349 |
| система | 264 | definition | 340 |
| современный | 258 | method | 327 |
| использовать | 248 | linguistic | 326 |
| разный | 248 | example | 320 |
| вид | 246 | result | 320 |
| общий | 245 | structure | 320 |
| следовать | 239 | need | 317 |
| тип | 239 | level | 313 |
| определение | 238 | approach | 310 |
| часть | 236 | resource | 294 |
| отношение | 234 | entry | 292 |
| структура | 234 | category | 290 |
| категория | 232 | specific | 289 |

– общенаучные термины-глаголы (*provide, need, define, search, make, give, order, see, create*);

– общенаучные термины-существительные (*document, need, standard, classification, group, class, order, collection, value, method, approach, project, source*);

– межотраслевые и узкоспециальные термины-существительные (*ontology, database, user, corpus, extraction, frequency, translator, web, context, search*);

– общенаучные и узкоспециальные термины-прилагательные (*conceptual, high*).

В словаре английских терминов также присутствуют общенаучные существительные и глаголы, которые организуют в текстах синтаксические связи между существительными. Единственное отличие состава частотных словарей на двух языках заключается в том, что в английском словаре узкоспециальные термины имеют большую частоту и более высокий ранг, чем в русском. Эти термины принадлежат информационным и компьютерным наукам, номинируют объекты и процессы, применимые в исследованиях по прикладной лингвистике, что может говорить о большем внимании авторов англоязычных текстов к компьютерным методам исследования в лингвистике и терминологии.

Если расширить выборку, то эквиваленты выделенным единицам словников обнаруживаются, только соотношение их частоты и ранга в словнике будет увеличено. Например, соотношение частот эквивалентов *ontology* и *онтология* в объединенном русско-английском словнике — 520 к 7, а рангов — 9 к 1745.

На основе лексического состава разных корпусов можно сделать вывод о неравномерном развитии терминосистем разных языков, описывающих единую предметную область. Тем не менее разработка процедур поиска эквивалентов в двуязычном сопоставимом корпусе показала, что выравнивание в сопоставимом корпусе по терминам возможно. Классификация терминов по области применения позволяет делать дальнейшие выводы о терминологическом значении терминологических словосочетаний, влиять на их дефинирование. Извлеченные из сопоставимого корпуса термины могут быть использованы для составления и сопоставления терминосистем и дальнейшего лексикографирования.

## ПРИМЕЧАНИЕ

* Благодарю С. Д. Шелова и Т. Горностай за предоставленные материалы.

## СПИСОК ЛИТЕРАТУРЫ

1. *Беляева Л. Н.* Лингвистические технологии в современном сетевом пространстве: *language worker* в индустрии локализации: монография. СПб.: Книжный дом, 2016. 134 с.

2. *Беляева Л. Н., Пиотровская К. Р.* Сетевой инструментарий филолога: учебное пособие. СПб.: Книжный дом, 2014. 48 с.

3. *Герд А. С.* Введение в изучение языка для специальных целей: учебное пособие. СПб.: СПбГУ РИО. Филологический факультет, 2011. 60 с.

4. *Захаров В. П.* Корпусная лингвистика: учебник для студентов гуманитарных вузов. Иркутск: ИГЛУ, 2001. 161 с.

5. *Лейчик В. М.* Терминоведение: предмет, методы, структура. 3-е изд. М.: Изд-во ЛКИ, 2007. 256 с.

6. *Тихонов А. Н.* Лексическое гнездо в современном русском языке: монография. Елец: ЕГУ им. И. А. Бунина, 2006. 270 с.

7. *Яцко В. А.* Предметная область компьютерной лингвистики // Вестник ИГЛУ. 2014. № 2 (27). С. 24–35.

8. *Delpech E., Daille B.* Dealing with Lexicon Acquired from Comparable Corpora: Validation and Exchange // Proceedings of 9th Conference on Terminology and Knowledge Engineering (TKE). Fiontar: Dublin City University, 2010. P. 229–223.

9. *Fung P.* Finding Terminology Translations from Non-parallel Corpora [Электронный ресурс]. — URL: http://www.cs.columbia.edu/nlp/papers/1997/fung_mckeown_97.pdf
10. *Khurshid A., Rogers M.* Terminology management: a corpus-based approach. — URL: http://www.mt-archive.info/90/Aslib-1992-Ahmad.pdf

**REFERENCES**

1. *Belyaeva L. N.* Lingvisticheskie tehnologii v sovremennom setevom prostranstve: language worker v industrii lokalizatsii: monografiya. SPb.: Knizhnyiy dom, 2016. 134 s.
2. *Belyaeva L. N., Piotrovskaya K. R.* Setevoy instrumentariy filologa: uchebnoe posobie. SPb.: Knizhnyiy dom, 2014. 48 s.
3. *Gerd A. S.* Vvedenie v izuchenie yazyika dlya spetsialnyih tseley: uchebnoe posobie. SPb.: SPbGU RIO. Filologicheskiy fakultet, 2011. 60 s.
4. *Zaharov V. P.* Korpusnaya lingvistika: uchebnik dlya studentov gumanitarnyih vuzov. Irkutsk: IGLU, 2001. 161 s.
5. *Leychik V. M.* Terminovedenie: predmet, metodyi, struktura. 3-e izd. M.: Izd-vo LKI, 2007. 256 s.
6. *Tihonov A. N.* Leksicheskoe gnezdo v sovremennom russkom yazyike: monografiya. Elets: EGU im. I. A. Bunina, 2006. 270 s.
7. *Yatsko V. A.* Predmetnaya oblast kompyuternoy lingvistiki // Vestnik IGLU. 2014. N 2 (27). S. 24–35.
8. *Delpech E., Daille B.* Dealing with Lexicon Acquired from Comparable Corpora: Validation and Exchange // Proceedings of 9th Conference on Terminology and Knowledge Engineering (TKE). Fiontar: Dublin City University, 2010. P. 229–223.
9. *Fung P.* Finding Terminology Translations from Non-parallel Corpora [Elektronnyiy resurs]. — URL: http://www.cs.columbia.edu/nlp/papers/1997/fung_mckeown_97.pdf
10. *Khurshid A., Rogers M.* Terminology management: a corpus-based approach. — URL: http://www.mt-archive.info/90/Aslib-1992-Ahmad.pdf

***М. В. Данильчук***

## ОСОБЕННОСТИ ПЕРЕВОДА НАЗВАНИЙ БРЕНДОВ И РЕКЛАМНЫХ СЛОГАНОВ

*В статье рассматриваются особенности перевода главных коммуникативных элементов бренда — названия и слогана. Слоган как структурный элемент рекламного текста представляет собой наибольшую трудность для переводчика, так как подлежит не только прагматической адаптации, но должен сохранить свою лаконичность, фонетический рисунок, ритм и другие характеристики. Для объективной оценки восприятия перевода названий брендов и слоганов потенциальными потребителями возможно использование метода краудсорсинга.*

**Ключевые слова:** перевод, рекламный текст, нейминг, слоган, краудсорсинг.

***M. Danilchuk***

## SPECIAL ASPECTS OF BRAND NAMES AND ADVERTISING SLOGANS TRANSLATION

*The paper presents the special aspects of translation of the main communicative elements for a brand — its name and slogan. Slogan as a structural element of advertising text presents the major problem for a translator because it needs not only to be transcreated but also should save its brevity, phonetic sounding, rhythm and other properties. An objective evaluation of the brand name and slogan translation perception by the potential customer can be done by method of crowdsourcing.*

**Keywords:** translation, advertising text, naming, slogan, crowdsourcing.