

Кудинова Татьяна Анатольевна

**ПРОБЛЕМА ГРАММАТИЧЕСКОЙ ОМОНИМИИ ПРИ СТАТИСТИЧЕСКОМ АНАЛИЗЕ КОРПУСА ТЕКСТОВ АНГЛИЙСКОГО ПОДЪЯЗЫКА БИОТЕХНОЛОГИЙ**

Статья посвящена актуальной на сегодняшний день проблеме грамматической омонимии при статистическом анализе малого корпуса текстов английского подъязыка биотехнологий. Автор приходит к выводу, что различение омографов в научной речи способствует более качественной классификации частей речи при составлении корпусов текстов и автоматических баз терминов. Изучение данной проблемы дополняется также рассмотрением основных понятий, свойств и задач корпусной лингвистики и этапов создания электронного корпуса текстов исследуемого подъязыка.

Адрес статьи: [www.gramota.net/materials/2/2017/4-2/35.html](http://www.gramota.net/materials/2/2017/4-2/35.html)

Источник

**Филологические науки. Вопросы теории и практики**

Тамбов: Грамота, 2017. № 4(70): в 2-х ч. Ч. 2. С. 122-124. ISSN 1997-2911.

Адрес журнала: [www.gramota.net/editions/2.html](http://www.gramota.net/editions/2.html)

Содержание данного номера журнала: [www.gramota.net/materials/2/2017/4-2/](http://www.gramota.net/materials/2/2017/4-2/)

**© Издательство "Грамота"**

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: [www.gramota.net](http://www.gramota.net)

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: [phil@gramota.net](mailto:phil@gramota.net)

УДК 81'322.2

*Статья посвящена актуальной на сегодняшний день проблеме грамматической омонимии при статистическом анализе малого корпуса текстов английского подязыка биотехнологий. Автор приходит к выводу, что различение омографов в научной речи способствует более качественной классификации частей речи при составлении корпусов текстов и автоматических баз терминов. Изучение данной проблемы дополняется также рассмотрением основных понятий, свойств и задач корпусной лингвистики и этапов создания электронного корпуса текстов исследуемого подязыка.*

*Ключевые слова и фразы:* корпусная лингвистика; корпус текстов; подязык; грамматическая омонимия; омография; конверсия.

**Кудинова Татьяна Анатольевна**, к. филол. н.

*Орловский государственный аграрный университет имени Н. В. Парахина*

*t.Kudinova77@mail.ru*

### **ПРОБЛЕМА ГРАММАТИЧЕСКОЙ ОМОНИМИИ ПРИ СТАТИСТИЧЕСКОМ АНАЛИЗЕ КОРПУСА ТЕКСТОВ АНГЛИЙСКОГО ПОДЪЯЗЫКА БИОТЕХНОЛОГИЙ**

Изучение языков для специальных целей в последнее время всё чаще делает необходимым использование современных методик изучения лингвистических явлений. Языки для специальных целей или подязыки науки и техники являются отличным языковым материалом для исследования функциональных языковых модификаций. Для их изучения требуется теперь гораздо большее количество контекстов, вариантов, лексико-грамматических парадигм. Вследствие этого специалисты различных отраслей – переводчики, лексикографы, преподаватели постоянно сталкиваются с необходимостью привлечения всё больших массивов текстов. Данная необходимость обусловила появление целой отрасли лингвистики – корпусной лингвистики.

Под термином «корпусная лингвистика» В. П. Захаров предлагает понимать раздел компьютерной лингвистики, занимающейся разработкой методологии создания и использования корпусов текстов с применением компьютерных технологий [6, с. 3].

Термин «корпус» обозначает любой систематический сборник электронных текстов, предназначенных для лингвистического исследования, которые были предварительно обработаны, размечены и систематизированы в электронной форме [2, с. 43].

Важнейшим свойством корпуса текстов считается его репрезентативность, под которой понимается «необходимо-достаточное и пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов» [6, с. 5]. Среди других отличительных признаков корпуса можно выделить выборочность исследования, фиксированный объём (размер), совместимость с вычислительным устройством и базу авторитетного справочника [13, р. 22].

Использование корпусов текстов играет всё более возрастающую роль в обучении языкам, психолингвистике, теоретической лингвистике. Современные корпуса отражают морфологические, синтаксические, семантические зависимости, содержат дискурсивную информацию и т.п. Лингвистическая составляющая работы с корпусом может касаться морфологического анализа, изучения грамматических отношений, значений слова, структуры подачи информации, семантической репрезентации. Всё это делает возможным собрать воедино отдельные элементы корпусов, сформировать ресурсы многократной повторяемости.

Разработка и создание корпусов ведётся по двум самостоятельным направлениям. Различаются они по объёму: большие по размеру – от 100 млн единиц вхождения – и малые корпуса размером до 1 млн единиц вхождения. Важным условием составления корпуса любого объёма является соблюдение всех правил лексикографического отбора: аутентичности, хронологической и жанровой однородности, стилевой отнесенности и т.д.

Материалом нашего исследования послужил малый корпус текстов английского подязыка биотехнологий, отобранный из 103 научных статей [8, с. 163-172]. Объём выборки составил 978 950 словоупотреблений (слов), что обеспечивает практически полный охват лексики подязыка биотехнологий. Материалом для конструкции данного корпуса текстов послужили источники следующих типов:

- 1) оригинальные печатные статьи из таких журналов, как «Biotechniques» («Биотехнологии»), «Biotechnology Progress» («Биотехнологический прогресс»), «Biotechnology Advances» («Биотехнологические достижения»), «Biotechnology and Genetic Engineering Review» («Обзор биотехнологий и генной инженерии») и т.д.;
- 2) научные монографии видных зарубежных ученых, занимающихся актуальными проблемами биотехнологий;
- 3) полнотекстовые версии научных журналов на CD-ROM корпорации *EBSCO Publishing*;
- 4) материалы Интернета.

Нам представляется, что сконструированная выборка текстов отражает реальное соотношение разделов науки и их языковых проекций. Последовательная автоматическая обработка текстов ограниченной тематики объёмом по 3-5 тыс. словоупотреблений позволила получить картину распределения частот в их абсолютном выражении.

В результате были получены:

- 1) алфавитный ранговый список всех словоформ текста;
- 2) частотный список лексем в порядке убывания их абсолютных частот.

При анализе и интерпретации результатов исследуемого корпуса приходилось решать одну техническую и одновременно лингвистическую задачи – устранения омонимии слов. Проблема омонимии разных типов является весьма актуальной проблемой при составлении корпусов текстов и автоматизированных баз терминов. Исследователи отмечают, что омонимия получила широкое распространение в английском языке для специальных целей [3, с. 32; 7, с. 146; 9, с. 158; 10, с. 168]. Для английского языка характерна только частеречная или грамматическая омонимия, возникновение которой не является следствием случайного совпадения: существует связь между омонимами данного типа, так как процесс перехода из одной части речи в другую происходит при условии изменения функционирования и свидетельствует об экономии в языке средств выражения понятий.

По определению З. А. Харитончик, грамматические омонимы – это «омонимические формы в системе одного и того же слова, различающиеся своими грамматическими значениями» [12, с. 72-73]. Частеречная омонимия (омография) состоит в том, что «слова, совпадающие по звучанию, принадлежат к различным частям речи и соответственно имеют разные грамматические значения, разную синтагматическую сочетаемость и участвуют в разных парадигматических связях» [11, с. 37].

Задача определения частей речи и классов слов является теоретически и практически важным для учёта всего арсенала лексических средств, участвующих в построении научного высказывания. Именно с этой целью и целесообразна экспликация на базе существующих грамматических категорий классов слов, более или менее значительных с точки зрения лексической стратификации. По мнению С. Д. Береснева, точное определение частеречной отнесенности – это наиболее распространенная операция при обработке больших массивов текста, «связанная с устранением неопределенности в выборе одного из двух возможных решений при распознавании образа слова» [1, с. 38].

Проблема омографии достаточно остро, по сравнению с другими языками, стоит в английском языке. Так, написание всех грамматических классов слов со строчной буквы значительно усложняет процедуру их различения от омографичных единиц. Сравним: *need* – потребность (сущ.) – *need* – нуждаться (гл.) или *farming* – занятие сельским хозяйством (сущ.) – *farming* – сельскохозяйственный (прил.).

Автоматический анализ корпуса текстов английского подъязыка биотехнологий не дал возможности классифицировать омонимичные единицы сразу в тот или иной разряд слов. Проблема отнесения таких омографичных единиц решалась преимущественно на основании синтаксической функции, выполняемой тем или иным словом в тексте. Для этого слово маркировалось в тексте и определялось процентное соотношение между двумя единицами. Пропорция давала нам абсолютную частоту употребления в общем объёме выборки.

Что касается вспомогательных глаголов (*be, have*) как элементов аналитических временных форм и как личных глаголов, нами не делалось различия. Проблема частеречной отнесенности затрагивает преимущественно служебные части речи или слова заполнения текста. Так, безусловно, омографичными в тексте будут неопределённое местоимение, прилагательное и существительное: *one* – *этот, тот* (неопр. мест.) – *one* – единственный (прил.) – *one* – *единица* (сущ.) – *one* – один (числ.).

Грамматическая омонимия в подъязыке биотехнологий часто наблюдается при попытке частеречной отнесённости существительного и глагола: *clone* – клон (сущ.) – *clone* – клонировать (гл.), глагол в форме прошедшего времени и причастия II: *protected* – защищал (гл.) – *protected* – *защищённый* (прич.), существительного и причастия I: *planting* – посадка (сущ.) – *planting* – сажающий (прич.).

Ниже приводится таблица, которая наглядно демонстрирует явление омонимии в английском подъязыке биотехнологий у наиболее частотных лексических единиц с указанием их абсолютной частоты употребления:

<i>N</i>	<b>F (%)</b>	<i>V</i>	<b>F (%)</b>	<i>PI</i>	<b>F (%)</b>	<i>PII</i>	<b>F (%)</b>
Clone <i>клон</i>	16	Clone <i>клонировать</i>	30	Cloning <i>клоняющийся</i>	43	Cloned <i>клонированный</i>	6
Breed <i>порода</i>	7	Breed <i>размножаться, разводить</i>	14	Breeding <i>размножающийся</i>	54	Bred <i>выведенный</i>	9,4
Plant <i>растение</i>	87	Plant <i>сажать</i>	1,8	Planting <i>сажающий</i>	3	Planted <i>посаженный</i>	3,4
Test <i>тест, опыт</i>	50	Test <i>тестировать</i>	28	Testing <i>тестирующий</i>	29	Tested <i>испытанный</i>	6,6
Cause <i>причина</i>	14,7	Cause <i>вызывать, причинять</i>	62	Causing <i>причиняющий</i>	7,2	Caused <i>вызванный</i>	34
Change <i>изменение</i>	67	Change <i>изменять</i>	17	Changing <i>изменяющий</i>	9,8	Changed <i>изменённый</i>	5,8
Control <i>контроль</i>	78	Control <i>контролировать</i>	38	Controlling <i>контролирующий</i>	7,4	Controlled <i>контролируемый</i>	12

Как видно из таблицы, существительное и глагол являются абсолютными омографами, что обусловлено явлением *конверсии*, при котором «создаётся новая (производная) лексема с полным сохранением формы производящей» [4, с. 3]. В английском языке конверсия представляет собой весьма продуктивный способ

словообразования, необходимым условием которого является омонимия форм исходного и производного слов, что отличает конверсию от других способов словообразования [5, с. 9].

Продуктивность явления конверсии как особого вида омонимии обусловлена особенностями грамматического строя английского языка, а именно отсутствием форм, характеризующих части речи. Данное свойство делает возможным образование от одной части речи нового слова, принадлежащего к другой части речи, без использования словообразовательных элементов. Кроме того, слово, появившееся в результате конверсии, включает в себя некую семантическую область исходного слова-основы, благодаря чему многие конверсированные слова легко восстанавливаемы из контекста.

Таким образом, разработка и методика работы с корпусами в современных условиях обозначает широкие возможности использования корпусов для изучения синтаксиса, функционально-связанных компонентов значения и анализа корпусных баз. Одинаково успешно могут быть исследованы разделы лексики и фразеологии, прагматика и фонология, методологические и структурные методы в грамматике, а также математические основания для статистических подходов к лингвистическому анализу.

Грамматическая омонимия – это лишь одна из особенностей исследуемого корпуса. Различие омографов в речи способствует более качественной частеречной классификации, что немаловажно с точки зрения решения одной из главных задач – создания базовых частотных словарей. Следовательно, различение формы слова и снятие омографии обусловлено, прежде всего, прагматическим подходом к научному тексту. Иноязычному реципиенту для понимания текста необходимо знание не только лексического, но и грамматического значения в каждой словоформе текста.

#### *Список источников*

1. Береснев С. Д. Грамматика декодирования немецкого научно-технического русским реципиентом. Киев: УМК МО, 1991. 142 с.
2. Волоснова Ю. А. Корпусная лингвистика: проблемы и перспективы // Лесной Вестник. 2006. № 7. С. 43-49.
3. Гашков А. В. Автоматический поиск синтаксической омонимии в тексте // Слово, высказывание, текст в когнитивном, прагматическом и культурологическом аспектах: материалы VIII международной научной конференции. Челябинск, 2016. С. 32-34.
4. Голубева А. Ю. Конверсия в словообразовании: узус и окказиональность: автореф. дисс. ... к. филол. н. Воронеж, 2014. 24 с.
5. Губанова И. С. Структурно-квантитативная характеристика омонимии в современном английском языке: автореф. дисс. ... к. филол. н. Н. Новгород, 2010. 18 с.
6. Захаров В. П. Корпусная лингвистика: учеб.-метод. пособие. СПб., 2005. 48 с.
7. Капура Н. В., Дроздова Т. В., Яшина Т. А. К вопросу о состоянии проблемы омонимии в современном английском языке // Наука, образование, общество. 2015. № 4 (6). С. 145-150.
8. Кудинова Т. А. Структурно-семантические особенности многокомпонентных терминов в подязыке биотехнологий (на материале русского и английского языков): дисс. ... к. филол. н. Орёл, 2006. 245 с.
9. Николаева Н. С. Некоторые особенности семантики английских терминов (омонимия) // Филологические науки. Вопросы теории и практики. Тамбов: Грамота, 2015. № 3 (45): в 3-х ч. Ч. 1. С. 157-159.
10. Порохнин А. А. Анализ статистических методов снятия омонимии в текстах на русском языке // Вестник Астраханского государственного технического университета. Серия «Управление, вычислительная техника и информатика». 2013. № 2. С. 168-174.
11. Саргсян Л. В. Категориальная мотивированность звуковой формы слова. Ереван: Изд-во ЕГУ, 2012. 248 с.
12. Харитончик З. А. Лексикология английского языка: учеб. пособие. Минск: Выш. шк., 1992. 229 с.
13. McENERY T., WILSON A. Corpus Linguistics. Edinburgh: Edinburgh University Press, 1996. 240 p.

#### **THE PROBLEM OF GRAMMATICAL HOMONYMY IN STATISTICAL ANALYSIS OF THE ENGLISH BIOTECHNOLOGY CORPUS**

**Kudinova Tat'yana Anatol'evna**, Ph. D. in Philology  
Orel State Agrarian University named after N. V. Parakhin  
*t.Kudinova77@mail.ru*

The article is devoted to the relevant problem of grammatical homonymy in statistical analysis of the English Small Corpus of Biotechnology. The author concludes that homograph differentiation in the scientific speech promotes more qualitative parts of speech classification when compiling text corpuses and electronic term bases. The paper also examines the basic notions, features and purposes of corpus linguistics and the development stages of electronic text corpus of the mentioned sublanguage.

*Key words and phrases:* corpus linguistics; text corpus; sublanguage; grammatical homonymy; homography; conversion.