Реферат на тему: «Национальный корпус» как собрание текстов в электронной форме.

|  |  |
| --- | --- |
|  | Выполнила:студентка I курса факультета ИЯ и МКгруппа 13 Оганнисян Офеля Жирайровна |
|  |  |
|  | Проверила:к.ф.н., доцент кафедры английского языка Е.М. Масленникова  |

Тверь

2019

**Оглавление**

* 1. Что такое национальный корпус……………………………………………...3
	2. Особенности национального корпуса…………………………………….….3
	3. Зачем нужен национальный корпус ……………………………………….…4
1. Корпусная лингвистика……………………………………..…………………6

Список использованной литературы………………………………………….7

***1.1 Что такое национальный корпус***

Корпус - это собрание текстов в электронном виде, но эти тексты предназначены не для того, чтобы их читать, а для того, чтобы их изучать, для чего их особым образом обрабатывают - вносят необходимую информацию, которая называется разметкойили аннотацией. Информация может быть любой, в зависимости от того, что мы хотим изучать: дата написания текста, место написания текста, его объем, авторство, употребление того или иного слова или грамматической конструкции, и так далее.

После того, как корпус обработали, мы можем искать в нем любую необходимую нам информацию. Для обработки информации используются специальные программы, которые называются конкордансеры*.* Они осуществляют поиск по тексту подобно тому, как поисковая система ищет информацию в Сети, и формируют конкорданс*,* то есть перечень всех контекстов, в которых какое-либо слово или слово­сочетание встречается в исследуемом тексте.

Большой корпус, который включает в себя тексты из различных источников и представляет язык исчерпывающим образом, называют национальным корпусом*.*

***1.2 Особенности национального корпуса***

 Национальный корпус имеет две важные особенности. Во-первых, он характеризуется представительностью, или сбалансированным составом текстов. Это означает, что корпус содержит по возможности все типы письменных и устных текстов, представленные в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т.п.), и что все эти тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода. Следует иметь в виду, что хорошая представительность достигается только при значительном объеме корпуса (десятки и сотни миллионов словоупотреблений).

Во-вторых, корпус содержит особую дополнительную информацию о свойствах входящих в него текстов (так называемую разметку, или аннотацию). Разметка - главная характеристика корпуса; она отличает корпус от простых коллекций текстов, в изобилии представленных в современном интернете, в том числе и на русском языке. В настоящее время специалистами создана и пополняется также «[Фундаментальная электронная библиотека](http://feb-web.ru/)» русской классической литературы, ориентированная на академический режим подачи текстов, максимально точное воспроизведение авторитетных печатных изданий. Однако такие библиотеки в необработанном виде для научных исследований языка пригодны очень ограниченно. Не следует забывать также, что библиотеки создаются теми, кому интересно в большей степени содержание текстов, чем их языковые качества. Для составителей Национального корпуса такие факторы, как увлекательность или полезность книги, ее высокие художественные или научные достоинства являются важными, но не первостепенными. Национальный корпус, в отличие от электронной библиотеки, - это не собрание «интересных» или «полезных» текстов; это собрание текстов, интересных или полезных для изучения языка. А такими могут оказаться и роман второстепенного писателя, и запись обычного телефонного разговора, и типовой договор аренды и т.п. - наряду, конечно, с классическими произведениями художественной литературы.

***1.3 Зачем нужен национальный корпус***

Национальный корпус предназначен в первую очередь для обеспечения научных исследований лексики и грамматики языка, а также тонких, но непрерывных процессов языковых изменений, происходящих в языке на протяжении сравнительно небольших периодов - от одного до двух столетий. Другая задача корпуса - предоставление всевозможных справок, относящихся к указанным областям (лексика, грамматика, акцентология, история языка). Современные компьютерные технологии многократно упрощают и ускоряют процедуры лингвистической обработки больших массивов текстов. Раньше исследователь мог лишь просматривать тексты и вручную выписывать из них нужные примеры; эта предварительная (но абсолютно неизбежная) деятельность была очень трудоемкой и не позволяла обрабатывать большие массивы материала. Теперь ограничений на объем анализируемого материала и скорость поиска информации в нем по существу нет, а это означает, что в распоряжении исследователя оказываются колоссальные массивы текстов самого разного типа. Это не замедлило сказаться на развитии наших знаний о языке: возможность массовой - в том числе статистической - обработки текстов, недоступная прежде, позволила обнаружить в структуре и развитии языка такие закономерности, о существовании которых наука раньше или не подозревала, или лишь смутно догадывалась, но не могла строго обосновать. Теперь подлинно научные описания грамматического строя языков, а также авторитетные академические словари - практически все без исключений - должны составляться на основе корпусов этих языков. Учет корпусных данных оказывается крайне желательным (если не строго обязательным) и при многих других более специальных научных исследованиях.

С этой точки зрения основными потребителями национальных корпусов являются, конечно, исследователи-лингвисты самого разного профиля. Однако круг пользователей корпуса вовсе не ограничивается профессиональными исследователями языка. Надежные статистические данные о языке определенной эпохи или определенного автора могут интересовать литературоведов, историков и представителей многих других областей гуманитарного знания. Важное значение национальные корпуса имеют и для преподавания языка в качестве родного или иностранного; всё больше учебников и учебных программ в наше время оказываются ориентированы на корпус. Быстро и эффективно проверить с помощью корпуса особенности употребления незнакомого слова или грамматической формы у авторитетных авторов сможет и иностранец, и школьник, и учитель, и журналист, и редактор, и писатель. Таким образом, национальный корпус обращен ко всем, кто в силу профессии, по необходимости или из простой любознательности ищет ответ на вопросы об устройстве и функционировании языка, то есть фактически к большинству образованных носителей этого языка и ко всем, изучающим его в качестве иностранного.

***2. Корпусная лингвистика***

**Корпусная лингвистика**- это наука о создании и использовании корпусов. Наряду с опросами носителей языка и экспериментами корпусная лингвистика является методом исследования аутентичных языковых данных.

 Для большинства крупных языков мира уже созданы уже так называемые национальные корпуса. Общепризнанным образцом является, в частности, [Британский национальный корпус](http://www.natcorp.ox.ac.uk/) (100 млн. слов), на который ориентированы и многие другие корпуса.

 [Национальный корпус русского языка](http://www.ruscorpora.ru/) – это корпус современного русского языка объемом более 300 млн. слов и информационно-справочная система, основанная на собрании русских текстов в электронной форме.

Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов, преподавателей языка, школьников и студентов, иностранцев, изучающих русский язык.

**Список использованной литературы**

1. Ruscorpora : <http://www.ruscorpora.ru/corpora-intro.html>
2. E-Language: [https://annapal.jimdo.com/](https://annapal.jimdo.com/%D0%BC%D0%BE%D0%B8-%D0%BA%D1%83%D1%80%D1%81%D1%8B/%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5-%D1%82%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D0%B8-%D0%B2-%D0%BB%D0%B8%D0%BD%D0%B3%D0%B2%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B5/%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5-%D0%BA%D0%BE%D1%80%D0%BF%D1%83%D1%81%D0%B0/)
3. EnglishZoom : <https://www.englishzoom.ru/stati-po-metodike-prepodavaniya/chto-takoe-elektronnyj-korpus-tekstov.html>