

Жолобова Анна Олеговна

НОВЫЕ КОРПУСА ИСПАНСКОГО ЯЗЫКА CORPES XXI И CDH - НОВЫЕ ВОЗМОЖНОСТИ

В статье дается характеристика новым корпусам испанского языка, над которыми работает Испанская королевская академия: современному корпусу CORPES XXI и историческому CDH. Приводятся данные географического, жанрово-тематического, хронологического и статистического характера. Анализируются новые технические возможности системы поиска данных корпусов в сопоставлении с предыдущим корпусом CREA, в котором отсутствует морфологическая разметка.

Адрес статьи: www.gramota.net/materials/2/2015/2-1/20.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2015. № 2 (44): в 2-х ч. Ч. I. С. 85-87. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2015/2-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

УДК 811.134.2

Филологические науки

В статье дается характеристика новым корпусам испанского языка, над которыми работает Испанская королевская академия: современному корпусу *CORPES XXI* и историческому *CDH*. Приводятся данные географического, жанрово-тематического, хронологического и статистического характера. Анализируются новые технические возможности системы поиска данных корпусов в сопоставлении с предыдущим корпусом *CREA*, в котором отсутствует морфологическая разметка.

Ключевые слова и фразы: национальный корпус испанского языка; исторический корпус; современный корпус; географические данные; жанрово-тематические данные; хронологические данные; статистические данные; морфологическая разметка.

Жолобова Анна Олеговна, к. филол. н., доцент
Казанский (Приволжский) федеральный университет
anijola@yahoo.es

НОВЫЕ КОРПУСА ИСПАНСКОГО ЯЗЫКА *CORPES XXI* И *CDH* – НОВЫЕ ВОЗМОЖНОСТИ[©]

Работа выполнена в рамках гранта Российского гуманитарного научного фонда: проект № 14-04-0061.

Отличительной особенностью электронного корпуса языка по сравнению с простым собранием текстов в электронном виде является его разметка или аннотация, которая может быть морфологической, семантической, синтаксической и т.п. (см., например, статью Г. В. Колпаковой [2]). Отсутствие необходимой морфологической разметки и лемматизации в первом национальном корпусе испанского языка *CREA / CORDE* не раз становилось предметом критики (см. подробнее в: [1]). Признав недочеты корпуса, Испанская королевская академия приступила к воплощению амбициозного проекта по созданию нового национального корпуса испанского языка, включающего современный корпус [4] и исторический [3].

В 2007 году было принято решение о необходимости создания корпуса испанского языка XXI века *CORPES XXI*, который должен отличаться представительностью и состоять из текстов различного типа с ежегодным объемом пополнения в 25 млн словоформ [4]. На первом этапе работы, который планируют завершить в декабре 2014 года, корпус составят устные и письменные тексты с 2001 г. по 2012 г. общим объемом 300 млн словоформ. На июль 2014 года версия бета 0.7 насчитывает 174 156 994 словоформ. Устные тексты составляют лишь 10%, письменные – 90%, причем последние распределены по типу источника следующим образом: книги – 40%, пресса – 40%, Интернет – 7,5%, разное – 2,5% [Ibidem].

Рассмотрим поисковые возможности *CORPES XXI*. Во-первых, в отличие от корпуса *CREA* поиск можно осуществлять как по лемме (основной форме слова), так и по словоформе. Также отличительной особенностью является то, что при поиске сразу нескольких лексем или словоформ некоторые команды теперь вводятся не вручную в поисковую строку, а выбираются из выпадающего списка в меню *Conector* «связка». Так, команда *Y* «и» позволит найти контекст с двумя заданными лексемами: *campo Y casa* «поле И дом»; команда *O* «или» позволит найти контекст с одной из заданных лексем: *campo O casa* «поле ИЛИ дом»; команда *NO* «не» исключает из поиска вторую лексему: *campo NO casa* «поле НЕ дом». Сохранилась возможность использовать в поисковой строке команды «⁻²» (с помощью вопросительного знака можно заменить один другой знак в запросе, что позволит найти в корпусе все возможные варианты с подстановкой данного знака) и «^{-*}» (позволяет заменить сразу несколько знаков) [6]. Однако, благодаря более продвинутому программному обеспечению, система поиска в *CORPES XXI* в отличие от *CREA* успешно справляется со сложными запросами типа *sal** [1].

Кроме того, появилась возможность выбрать интересующий класс слова (*clase de palabra*): *adjetivo* «прилагательное», *adverbio* «наречие», *afijo* «аффикс», *amalgama* «сочетание», *artículo* «артикль», *conjunción* «союз», *cuantificador* «квантификатор», *demostrativo* «указательная форма», *desconocido* «неизвестная форма», *interjección* «междометие», *interrogativo* «вопросительная форма», *numeral* «числительное», *posesivo* «притяжательная форма», *preposición* «предлог», *pronombre personal* «личное местоимение», *puntuación* «пунктуация», *relativo* «относительное местоимение», *sustantivo* «существительное», *verbo* «глагол». В зависимости от выбранной грамматической категории можно задать другие связанные с ней грамматические признаки. Так, например, указав категорию глагола, возможно задать определенное число, лицо, время (сюда включены и безличные формы глагола), и залог.

Выбор в меню поиска графы *Subcorpus* «подкорпус» открывает доступ к дополнительным параметрам поиска: *título* «название», *autor* «автор», *fecha de clasificación* «дата классификации» (годы издания), *origen* «происхождение» (географическая характеристика), *medio* «канал» (устная или письменная речь), *bloque* «блок», *sopORTE* «формат», *tema* «тема», *tipología* «типология» (тип текста).

Интересно отметить, что в рассматриваемом корпусе по сравнению с *CREA* изменилась географическая дис-трибуция текстов: 70% всего материала составляют латиноамериканские тексты, а 30% – испанские. Кроме того,

представлены Филиппины и Экваториальная Гвинея. При выборе из выпадающего списка критерия *América* «Америка» появляется меню *Zonas lingüísticas* «лингвистические зоны», которое включает следующие регионы:

1. *Andina* «андская» (Боливия, Эквадор, Перу);
2. *Antillas* «антильская» (Куба, Пуэрто Рико, Доминиканская республика);
3. *Caribe continental* «континентальные страны Карибского бассейна» (Колумбия, Венесуэла);
4. *Chilena* «чилийская» (Чили);
5. *México y Centroamérica* «Мексика и Центральная Америка» (Коста Рика, Сальвадор, Гватемала, Гондурас, Мексика, Никарагуа, Панама);
6. *Río de la Plata* «Рио де ла Плата» (Аргентина, Парагвай, Уругвай).
7. *Estados Unidos* «США».

При выборе определенного региона появляется меню *Países* «страны», в котором мы можем задать конкретную страну, принадлежащую данной лингвистической зоне.

Типологию текстов мы можем отразить в следующей Таблице, где первый блок – художественная литература, второй – нехудожественная литература:

№	Формат	Тип текста	Жанр // Тема
1	Книга	Художественный	Сценарий / Роман / Рассказ / Драматургия
2	Книга	Академический / Биография, мемуары / Популярный / Административно-юридический / Учебник	Современность, досуг и быт Искусство, культура и развлечения Социальные науки, убеждения и мышление Наука и техника Политика, экономика и право Здоровье
	Пресса	Академический / Письмо в редакцию / Критика / Хроника / Популярный / Передовая статья / Интервью / Новость / Мнение / Репортаж / Разное	
	Интернет	Академический / Блог / Интервью / Сообщение по электронной почте / Сообщение в социальных сетях / Интернет-страница	
	Разное	Академический / Официальное письмо-рассылка / Частное письмо / Административные документы / Руководство по использованию / Другое / Рекламные памфлеты / Брошюры	

Также в поиске мы можем задать параметры словесного окружения (*proximidad*): лемму, словоформу, класс слова, расстояние (до 5 слов), интервал, положение слова относительно искомого (справа, слева), причем есть возможность выбрать те же дополнительные критерии поиска, как и в случае основной единицы поиска.

Что касается представления результатов выполненного запроса, то отметим следующие особенности:

1. упорядочить полученные результаты конкорданса можно согласно различным параметрам (год издания, страна, автор и т.д.);
2. наведение курсора на высвеченную синим цветом лексему активирует дополнительное окошко, в котором указываются грамматические характеристики лексемы и более широкий контекст его употребления;
3. наведение курсора на первую графу таблицы конкорданса (год издания) активирует дополнительное окошко, в котором указываются выходные данные текста-источника;
4. контексты употребления можно вывести на печать (все или по выбору).

Кроме конкорданса, результаты можно получить в виде статических данных по таким позициям, как распределение по зонам, по странам, по году, по тематике и по типу текста. По каждой позиции статистика представлена в двух формах: табличной и в виде круговой диаграммы. Кроме того, указывается абсолютная частота, количество документов и нормализованная частота появления определенной лексемы или словоформы. Также система позволяет получить фактическую и статистическую информацию по сочетаемости интересующей единицы в корпусе.

Исторический корпус *Corpus del Nuevo diccionario histórico del español* (CDH) «корпус Нового исторического словаря испанского языка» общим объемом более 355 млн словоформ состоит из трех больших блоков: основной корпус (более 53 млн словоформ), корпус 12 в.-1975 г. (более 199 млн словоформ), корпус 1975-2000 гг. (более 103 млн словоформ) [3]. Распределение текстов по географическим зонам выглядит следующим образом: Испания – 73,5%; Мексика и Центральная Америка – 6,9%; Рио де ла Плата – 5,5%; андская зона – 4,3%; континентальные страны Карибского бассейна – 3,9%; Чили – 2,5%; антильская зона – 2,4%; США – 0,5% [5].

Принципы работы в *CDH* такие же, как и в *CORPES XXI*. Однако имеются и некоторые особенности. Так, например, можно выбрать дополнительный тип кодификации корпуса: *leyenda* «подпись» (подпись на изображении или фотографии, принадлежащих автору произведения), *cambio de mano* «» (текст, принадлежащий другому автору, но не являющийся цитатой), *corrección* «правка» (правильная форма исходного текста с ошибками), *desarrollo* «расшифровка» (развернутая форма аббревиатуры), *idioma extranjero* «иностранный язык», *glosa* «глосса» (пометы на полях текста), *nota* «заметка» (авторские комментарии и сноски), *cita* «цитата» (цитата другого автора или фрагмент из другого текста того же автора), *sic* «дословное цитирование».

Итак, проанализировав новый национальный корпус испанского языка с различных позиций, мы можем с уверенностью сказать, что испанская корпусная лингвистика вышла на совершенно новый уровень развития. Главными достоинствами нового корпуса по сравнению с предыдущим корпусом *CREA / CORDE* является его лемматизация, морфологическая аннотация, представительность и удобный интерфейс, что создает новые и гораздо более широкие возможности для проведения лингвистических исследований в области испанского языка и его национальных вариантов.

Список литературы

1. **Жолобова А. О.** Национальный корпус испанского языка: *CREA* и *CORDE* // Филологические науки. Вопросы теории и практики. 2014. № 9 (39). Ч. 1. С. 56-58.
2. **Колпакова Г. В.** Методы анализа корпусной лингвистики // Филологические науки. Вопросы теории и практики. 2012. № 4 (15). С. 75-77.
3. **CDH** [Электронный ресурс]. URL: <http://www.rae.es/recursos/banco-de-datos/cdh> (дата обращения: 10.11.2014).
4. **CORPES XXI** [Электронный ресурс]. URL: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi> (дата обращения: 10.11.2014).
5. **Corpus del Nuevo Diccionario Histórico del Español** [Электронный ресурс]. URL: <http://web.frl.es/CNDHE/org/publico/pages/ayuda/ayuda.view> (дата обращения: 14.11.2014).
6. **Manual de uso: Corpus del español del siglo XXI** [Электронный ресурс]. URL: <http://web.frl.es/CORPES/org/publico/pages/ayuda/manual.view> (дата обращения: 10.11.2014).

NEW CORPUSES OF THE SPANISH LANGUAGE *CORPES XXI* AND *CDH* – NEW POSSIBILITIES

Zholobova Anna Olegovna, Ph. D. in Philology, Associate Professor
Kazan (Volga Region) Federal University
anijola@yahoo.es

The article describes new corpora of the Spanish language, with which the Royal Spanish Academy works: the modern corpus *CORPES XXI* and historical one *CDH*. The data of geographical, genre-thematic, chronological and statistical nature are given. The author analyzes the new technical possibilities of corpora data search engine in comparison with the previous corpus *CREA*, which lacks the morphological marking.

Key words and phrases: National Corpus of the Spanish language; historical corpus; modern corpus; geographical data; genre-thematic data; chronological data; statistical data; morphological markings.

УДК 8;81:42

Филологические науки

Данная статья посвящена анализу интертекстуальных связей романа П. Дашковой «Чувство реальности», которые проявляются во включении в него фрагментов других текстов или отсылке к ним. В романе имеются отсылки к прецедентным текстам и именам из русской (Н. А. Некрасов «Кому на Руси жить хорошо», С. А. Есенин и др.) и зарубежной литературы (М. Шелли «Франкентейн, или Современный Прометей», Ш. Бронте «Джейн Эйр», Э. Хэмингуэй «По ком звонит колокол» и др.), а также из античных мифов, фольклорных произведений и кинематографа (кинофильмы «Неуловимые мстители», «Белое солнце пустыни»).

Ключевые слова и фразы: интертекстуальность; аллюзия; дискурс; прецедентное сравнение.

Зубакова Наталия Валерьевна

Липецкий государственный педагогический университет
Natyse4ka87@mail.ru

ИНТЕРТЕКСТУАЛЬНЫЕ СВЯЗИ РОМАНА П. ДАШКОВОЙ «ЧУВСТВО РЕАЛЬНОСТИ»[©]

Проблема интертекстуальности находится в центре внимания филологов уже достаточно продолжительное время. Началось исследование этого явления еще в работах М. М. Бахтина, который впервые сделал предметом рассмотрения проблему «чужого слова» и его художественных функций. В своих работах Бахтин писал, что, помимо данной действительности, художник имеет дело также с предшествующей и современной ему литературой, с которой он находится в постоянном «диалоге» [2, с. 393].

В последние годы интертекстуальность изучается на материале самых разных текстов (публицистических, научных, эпистолярных, рекламных и некоторых других), но самые яркие образцы интертекстуальности дает художественная литература, для которой опора на прецедентные тексты (интертекстуальность) является обязательным признаком (обязательной текстовой категорией).

Интертекстуальность – одна из неотъемлемых черт литературы XX-XXI вв., в том числе и массовой. Причем в массовой литературе функционирование прецедентных текстов имеет ряд особенностей, связанных с изначальной установкой на потребление текстов массовым читателем. По словам В. Н. Серкиной, прецедентные тексты, которые являются маркерами стилистической манеры автора, могут отвечать и за степень развития различных составляющих коммуникативной компетенции [6, с. 175]. Однако говорить об интертексте применительно к массовой литературе мы можем лишь условно, так как сохраняются только его внешние атрибуты – межтекстовые ссылки, упоминание в одном тексте названия, автора, героев, сюжетных событий другого. Внутреннее же содержание интертекста в массовой литературе практически отсутствует. Рассмотрим употребление прецедентных текстов на примере романа П. Дашковой «Чувство реальности».