

О. Н. Ляшевская

КОРПУСНЫЕ ИНСТРУМЕНТЫ В ГРАММАТИЧЕСКИХ ИССЛЕДОВАНИЯХ РУССКОГО ЯЗЫКА



ИЗДАТЕЛЬСКИЙ ДОМ ЯСК
РУКОПИСНЫЕ ПАМЯТНИКИ ДРЕВНЕЙ РУСИ
МОСКВА 2016

ББК 81.1
УДК 80/81
Л 99

Издание осуществлено при финансовой поддержке
Фонда фундаментальных лингвистических исследований
проект № В-28-2014

Утверждено к печати Ученым советом
Института русского языка имени В. В. Виноградова РАН

Рецензенты:

д. ф.-м. н. М. Р. Пентус, к. филол. н. И. В. Азарова

Ляшевская О. Н.

Л 29 Корпусные инструменты в грамматических исследованиях русского языка. — М.: Издательский Дом ЯСК: Рукописные памятники Древней Руси, 2016. — 520 с.

ISBN 978-5-9907947-8-8

Русская корпусная лингвистика представлена в книге двумя направлениями. Первая часть содержит описание подходов и методов аннотации Национального корпуса русского языка (<http://ruscorpora.ru>), включая разметку лексико-грамматической, лексико-семантической, семантико-синтаксической и словообразовательной информации. Кроме того, описываются процедуры оценки инструментов автоматической разметки текстов (морфологических и синтаксических парсеров) и идеология создания двух частотных корпусных словарей, общего и лексико-грамматического. Во второй части представлены результаты исследований грамматики и лексики русского языка с применением количественных корпусных методов: изучение грамматических, конструкционных и семантических профилей языковых единиц, в том числе глаголов и глагольных приставок, имен существительных и пространственных конструкций.

УДК 80/81
ББК 81.1

*В оформлении переплета использована картина
Пита Мондриана «Серое дерево», 1911*

ISBN 978-5-9907947-8-8

© Ляшевская О. Н., 2016
© Издательский Дом ЯСК, 2016

Электронная версия данного издания является собственностью издательства, и ее распространение без согласия издательства запрещается.

СОДЕРЖАНИЕ

Предисловие	7
-------------------	---

Часть 1. Развитие корпусных инструментов и технологий

1.1. Национальный корпус русского языка и его аннотация	13
1.2. Словоизменение	19
1.2.1. Морфологический стандарт корпуса	19
1.2.2. Пополнение грамматического словаря по корпусным данным	40
1.2.3. Соревнования морфологических анализаторов	49
1.3. Лексико-семантические классы	64
1.3.1. Принципы лексико-семантической разметки	64
1.3.2. Разрешение лексико-семантической неоднозначности с помощью векторов контекстных маркеров	88
1.4. Интерфейс морфосинтаксиса и семантики	112
1.4.1. Аннотация лексических конструкций в системе ФреймБанк	112
Приложение	164
1.4.2. Распознавание семантических ролей на основе ФреймБанка	176
1.4.3. Автоматическая синтаксическая аннотация корпуса и соревнования парсеров зависимостей	193
1.5. Словообразование	210
1.6. Частотные словари на базе корпуса	224
1.6.1. Частотный словарь современного русского языка	225
1.6.2. Частотный лексико-грамматический словарь	246

Часть 2. Квантитативные подходы к исследованию на корпусных данных

2.1. Векторное представление корпусных данных и профили контекстного «поведения» языковых единиц	257
2.2. Грамматические профили	279
2.2.1. Грамматическая специализация глаголов в формах времени и наклонения	279

2.2.2. К описанию дистрибуции форм единственного и множественного числа имен существительных	319
2.3. Конструкционные профили	338
2.3.1. Конструкционные профили приставочных видовых пар	338
2.3.2. Инкорпорация и экскорпорация в глагольном управлении: участник «часть тела»	358
2.3.3. Инструментальная и генитивная конструкция формы имен существительных	373
2.4. Семантические профили: классы глаголов и выбор видовых приставок.	382
2.5. Радиальный профиль значения: пространственная конструкция с предлогом <i>поверх</i>	407
Заключение	430

Приложения

Приложение 1	435
Приложение 2	457
Приложение 3	468
Приложение 4	474
Библиография	480
Принятые сокращения	514
Abstract	517