

80. Шалина И. В. Уральское городское просторечие: лингвокультурные типажи / И. В. Шалина // Известия Урал. ун-та. — 2009. — № 4 (70). — С. 144—150. (Серия 2. Гуманитарные науки. Филология. Языкознание).
81. Шалина И. В. Уральское городское просторечие: культурные сценарии: монография / И. В. Шалина. — Екатеринбург: Изд-во Урал. ун-та, 2009. — 444 с.
82. Шалина И. В. Уральское городское просторечие как лингвокультурный феномен: автореф. дисс. ... доктора филол. н.: 10.02.01 — русский язык / Ирина Владимировна Шалина. — Екатеринбург, 2010. — 43 с.
83. Шумарова Н. П. Про розвиток мовної ситуації в Києві / Н. П. Шумарова // Язык и культура: IV междунар. конф.: материалы. — Киев: Collegium, 1996. — Ч. 2. — С. 183—197.
84. Яковлева Е. А. Особенности русского речевого поведения горожанина в полиэтнической среде / Е. А. Яковлева // Проблемы коммуникации и номинации в концепции общегуманитарного знания. — Челябинск: Изд-во Челябинского гос. ун-та, 1999. — С. 188—196.
85. Geddes P. Cities in evolution / Patrick Geddes. — London: Williams & Norgate, 1915. — 117 p.
86. Gruchmanowa M. Mowa mieszkańców Poznania / Gruchmanowa M., Witaczek-Samborska M., Żak-Święcicka M. — Poznań: Wyd. Poznańskie, 1986. — 176 s.
87. Mumford L. The city in history. Its Origins, its Transformations, and its Prospects / Lewis Mumford. — London: Harcourt, 1961. — 568 p.
88. Saunders P. Social Theory and the Urban Question / Peter Robert Saunders. — London: Hutchinson, 1981. — 234 p.
89. Seiffert-Nauka I. Dawny dialekt miejski Lwowa / I. Seiffert-Nauka. — Wrocław: Wyd. Univ. Wrocławskiego, 1993. — Cz. 1: Gramatyka. — 172 s.
90. Wedel E. «Одессит» — Bemerkungen zum heutigen Russisch von Odessa / Erwin Wedel // Ars transferendi Sprache, Übersetzung, Interkulturalität: Festschrift für Nikolai Salnikow zum 65: Geburtstag. — Sonderdruck, 1997. — S. 247—259.
91. Wiczorkiewicz B. Gwara warszawska dawniej i dziś / Bronisław Wiczorkiewicz. — Warszawa, 1974.

**Г. М. Степанов**

#### **МІСТО В КОНТЕКСТІ НАУКОВОЇ ЛІНГВІСТИЧНОЇ ПАРАДИГМИ**

У статті подано аналіз урбанолінгвістичних досліджень нової та новітньої доби, виконаних у межах діалектично взаємозумовлених наукових парадигм: порівняльно-історичної, системно-структурної й антропоцентричної. Визначено місце урбанолінгвістики серед інших напрямів урбанології та місце російської урбанолінгвістики у світових дослідженнях мови міста.

**Ключові слова:** російське міське койне, урбанолінгвістика, наукова парадигма, порівняльно-історична парадигма, системно-структурна парадигма, антропоцентрична парадигма.

**Ye. N. Stepanov**

#### **CITY IN THE CONTEXT OF LINGUISTIC RESEARCH PARADIGMS**

The article analyzes the urbanolinguistic studies of modern and contemporary periods, carried out under the dialectically interrelated scientific paradigms: comparative and historical, systemic and structural, anthropocentric. The place of urbanolinguistics among other areas of urbanology and the place of Russian urbanolinguistics among the world researches of city dwellers' speech is defined.

**Keywords:** Russian urban, urbanolinguistics, scientific paradigm, comparative and historical paradigm, systemic and structural paradigm, anthropocentric paradigm.

УДК 81'26'271.12'42'322.2

**С. В. ПЛОТНИЦКАЯ**

#### **КОРПУСНАЯ ЛИНГВИСТИКА: НАЦИОНАЛЬНЫЙ КОРПУС ЯЗЫКА**

Статья посвящена вопросам возникновения корпусной лингвистики, её участия в процессе развития языка нации и изучения различных аспектов функционирования языковой системы.

**Ключевые слова:** корпусная лингвистика, национальный корпус текстов.

С оптимизацией гносеологической функции языка связана корпусная лингвистика, разрабатывающая общие принципы построения лингвистических корпусов с использованием современных компьютерных технологий. В настоящее время в научном словаре лингвистов появляются очень близкие понятия: «электронные библиотеки», «массив текстов», «коллекция текстов», «электронный архив», «полнотекстовая база данных». Среди них можно выделить лингвистические, или языковые, корпусы. По мнению А. Н. Баранова, корпус текстов — это вид корпуса данных, единицами которого являются тексты, включающие полные фрагменты макроструктуры

текстов проблемной области. Корпус текстов характеризуется четырьмя основными параметрами: 1) он должен быть достаточно большого объёма; 2) корпус должен быть структурированным или размеченным; 3) тексты, составляющие определённый корпус, должны быть в электронном варианте; 4) в понятие “электронный корпус” входит, как правило, специальное программное обеспечение для работы с этим корпусом.

Корпусная лингвистика основана на использовании корпуса, то есть большого объёма живого языкового материала, который можно извлечь из различных источников и ввести в компьютер. Она исследует как речь, так и язык в новом ракурсе, открывая огромный исследовательский словарь. Все это требует особого осмысления, ибо в рамках корпусной лингвистики изучаются вопросы распределения лингвистических явлений в разных языках и объективным путём получают новые лингвистические данные. Преимущества этого направления в том, что оно избегает субъективизма, неизбежного в традиционном языкознании, и опирается на объективные сведения.

Корпус — это информационно-справочная система, основанная на собрании текстов на определённом языке в электронной форме, т. е. это большой массив текстов, специально обработанный для целей лингвистического исследования. Национальный корпус представляет данный язык на определённом этапе его существования и во всём многообразии жанров, стилей, территориальных и социальных вариантов. Национальный корпус создаётся лингвистами для научных исследований. Большинство крупных языков мира уже имеет свои национальные корпуса, различающиеся по полноте и уровню научной обработки текстов. Общеизвестным образцом является, в частности, Британский национальный корпус (BNC): на него ориентированы многие другие современные корпуса. Британский национальный корпус (БНК) включает 100 млн примеров словоупотреблений письменного и устного языка последнего периода XX века. Письменная часть БНК (90 %) включает выдержки из региональных и национальных газет и журналов. Разговорная речь (10 %) состоит из орфографической транскрипции неформальных разговоров и разговорной речи — от правительственных встреч до радио-шоу.

Работа над составлением BNC велась с 1991 по 1994 г. Не было добавлено новых текстов, но корпус был обновлён во 2-м издании: BNC World (2001), — и в 3-м издании: — BNC XML Edition (2007). Британский национальный корпус включает только современный английский, т. е. язык-эталон. Однако слова небританского английского иногда появляются в корпусе. Этот проект был осуществлён BNC Консорциумом, включающим Oxford University Computing Services, British Library's Research и др. Создание корпуса началось со стадии планирования и разработки основных принципов, включающих критерии отбора текстов (как письменных, так и устных). Если было получено разрешение использовать необходимый текст, его переводили в электронную версию, затем итоговый текст переводили в закодированный формат проектного стандарта, где проверялась точность и соответствие внутреннему содержанию. Затем текст переходил в UCREL (Университетский центр по компьютерному исследованию корпусной лингвистики) в Ланкастере, где автоматически происходила морфологическая разметка класса слов — tagging: каждое слово получало морфологическую характеристику (род, число, падеж — у существительных; у глаголов — залог, время). Затем текст возвращался в Компьютерный центр Оксфорда для документирования и введения в Корпус. Каждая стадия процесса записывалась в базу данных Оксфордского Компьютерного Центра. Британцы ввели паспорт текста, т. е. внешние и внутренние параметры. Внешние параметры: автор (возраст, пол), год издания, место, название. Внутренние параметры: морфологическая разметка (tagging). Британский национальный корпус стал образцом для национальных корпусов других языков.

Среди корпусов славянских языков выделяется Чешский национальный корпус, созданный в Карловом университете Праги. 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, представленных в данном языке: художественные тексты разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т. д. — и что все эти тексты входят в корпус пропорционально их доле в языке соответствующего периода. Следует иметь в виду, что хорошая представительность достигается только при значительном объёме корпуса, а именно — десятки и сотни миллионов словоупотреблений. Планируемый составителями объём Национального корпуса русского языка — 200 млн слов.

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 трудоёмкой и не позволяла обрабатывать большие массивы материала. Теперь ограничений по объёму анализируемого материала и скорости поиска информации в нём, по существу, нет, а это означает, что в распоряжении исследователя оказываются колоссальные массивы текстов самого разного типа.

Корпусная лингвистика занимается созданием компьютерных инструментов, предназначенных для обработки текстов. В качестве единиц выступают тексты или их достаточно значительные фрагменты. Среди корпусов текстов выделяются «исследовательские корпусы» и «иллюстративные корпусы». Первые предназначены преимущественно для изучения различных аспектов функционирования языковой системы; они строятся до проведения какого-либо исследования. Иллюстративные корпусы создаются после проведения научного исследования; их цель не столько выявить новые факты, сколько подтвердить и обосновать уже полученные результаты. Такие корпусы не являются слепком, правильным (с точки зрения статистики) отображением проблемной области. Второе важное разграничение касается динамических (мониторных) и статических корпусов. Первоначально корпусы текстов создавались как статические образования, отражающие определённый временной срез языковой системы. Типичными представителями этого вида корпусов являются авторские корпусы — коллекции текстов писателей, например, Корпус художественных текстов Ф. М. Достоевского (Институт русского языка РАН). Однако значительная часть чисто лингвистических и не только лингвистических задач требует выявления исторических изменений в функционировании языковых явлений, например, изменений значения слов, частоты использования тех или иных синтаксических конструкций и т. д. Для этого была разработана специальная технология построения и эксплуатации динамического корпуса текстов. В имеющейся литературе такие корпусы получили название мониторных. Особенность сборки мониторных корпусов заключается в том, что они не предполагают раз и навсегда заданного набора текстов. В течение заранее фиксированного промежутка времени происходит обновление и дополнение множества текстов корпуса. К числу мониторных относится, например, Бирмингемский корпус, созданный под руководством Дж. Синклера. С точки зрения пользователя, к корпусам текстов предъявляются требования репрезентативности, полноты и экономичности.

Некоторые положения корпусной лингвистики были известны достаточно давно, например, дистрибутивная методика, составление конкордансов и пр. Однако как цельное лингвистическое направление она сформировалась сравнительно недавно. Язык, особенно его лексика, постоянно развивается. Слова приобретают новые значения, старые значения исчезают. Кроме того, появляются и совершенно новые слова. В подавляющем большинстве случаев новые слова — это термины, отражающие новые концепты науки, техники, социальных отношений, политики и экономики. Количество терминов в каждом языке исчисляется миллионами, включая словосочетания. Современная языковая ситуация, характеризующаяся информатизацией общества и так называемым «информационным взрывом», резким расширением каналов языковых коммуникаций, заставляет обращать особое внимание на расширение лексических значений и роль терминов и терминологии в современных языках. Слово и проблемы его анализа — основная тема современной лингвистики. Термин — это, прежде всего, слово. Являясь главным носителем семантической информации, слово выступает основным элементом высказывания. Современные теоретические и прикладные исследования исходят из формы и содержания именно слова, поскольку никакая другая языковая единица не обладает таким единством формы и содержания, не играет в высказывании такой важной конструирующей роли, как слово. Инструментом исследования может и должен быть не только словарь как регистратор значения слова в его парадигматике и синтагматике, но и конкорданс, составленный на основе представительной выборки текстов, как регистратор синтагматики и парадигматики слова уже в речи. Несмотря на то, что изменения в значениях только одного слова уже представляют собой самостоятельную проблему, можно представить совокупность слов определённой предметной области в виде некоторой системы значения слов.

Существуют слова с широкой понятийной основой, которые, являясь терминами, одновременно входят в пласт общеупотребительной лексики. В современном языковом общении большую роль играет терминология, которая функционирует как в общеупотребительном пласте, так и в виде достаточно узких специальных обозначений, вошедших, тем не менее, в язык: (*компьютер, файл, база данных* — *computer, file, data base* и т. д.). В рамках корпусной лингвистики появляется возможность по-новому описать такую важную единицу современной лексики, как термин.

1. Захаров В. П. Информационные системы (документальный поиск). — СПб., 2002.
2. Баранов А. Н. Корпусная лингвистика / Введение в прикладную лингвистику. — М., 2001. — С. 112—137.
3. Kennedy G. An Introduction to Corpus Linguistics. — London, 1998.

С. В. Плотницька

### КОРПУСНА ЛІНГВІСТИКА: НАЦІОНАЛЬНИЙ КОРПУС МОВИ

Статтю присвячено проблемі розвитку корпусної лінгвістики, її участі у процесі розвитку мови нації та вивченню різноманітних аспектів функціонування мовної системи.

**Ключові слова:** корпусна лінгвістика, національний корпус текстів.

S. V. Plotnitskaya

### CORPUS LINGUISTICS: NATIONAL LANGUAGE CORPUS

The article is devoted to the question of Corpus Linguistics origin, its partaking in the process of nation language development and linguistic system different aspects study.

**Key words:** corpus linguistics, national corpus of texts.

УДК 811.161.1'42:162.5:[82-96+087.5/.6]

А. Р. ГАБИДУЛЛИНА

### ПАРАДОКС В НАУЧНО-ПОПУЛЯРНОМ ЛІНГВІСТИЧЕСКОМ ТЕКСТЕ

В статті описані способи створення семантичного парадокса в научно-популярному тексті. Охарактеризовані мовні засоби реалізації парадокса. Названі його ознаки.

**Ключевые слова:** парадокс, алогизм, антитеза, оксюморон, катахреза.

Основная задача авторов научно-популярного текста — показать лингвистическое явление с непривычной, неожиданной стороны. Одним из эффективных приёмов популяризации научного знания является парадокс. Цель статьи — показать приёмы его создания в энциклопедии для детей «Языкознание. Русский язык».

С помощью парадоксальных заголовков создатели энциклопедии привлекают внимание адресата к явлениям, которые, на первый взгляд, кажутся противоречащими здравому смыслу: так, в статье *Лень — двигатель прогресса* [7, с. 643] речь идёт о фонетических процессах как результате экономии произносительных усилий.

В научной литературе парадокс понимается как отклонение или противопоставление здравому смыслу (БСЭ), странное, необычное мнение, расходящееся с традиционным (Е. Ihrig), неразрешимое противоречие (Г. А. Брутян), разрушение некоей презумпции, принимаемой как нечто парадоксальное (В. А. Успенский), суждение, в котором сказуемое противоречит подлежащему или определению противоречит определяемому (В. Я. Пропп), суждение, которое расходится с общепринятым представлением, отрицание того, что есть безусловно правильным (Д. П. Горский, Е. А. Селиванова), следствие изъятия логических или событийных звеньев из цепи суждений (Т. Г. Миролюбова), «алогическая связь двух частей предложения или нескольких предложений, в которых объединяются противоречивые понятия, опровергаются общепринятые мнения и штампы» [5]. Таким образом, в логике парадокс рассматривается в широком и узком смысле. В широком смысле это утверждение, которое резко расходится с общепринятыми мнениями, правилами, традициями. В узком смысле парадокс — это два противоположных утверждения, для каждого из которых имеются убедительные аргументы.

В отличие от логики, в языке научно-популярной литературы парадокс может означать «просто какую-то органическую несогласованность, несоответствие, <...> но такую несогласованность, которая не может быть устранена» [1, с. 198]. Речевой парадокс тяготеет к сублимированным формам языкового комизма. В нём «осуществляется транспозиция признаков, причём какой-либо признак становится антитезой общепринятого суждения» [3].

В лингвистической литературе характеризуются, как правило, три вида парадокса: логический, семантический и прагматический. При этом отмечается, что в речи резкой границы между первыми двумя не существует, «поскольку логические парадоксы иногда используют семантику языковых знаков, а семантические являются столкновением логики правды/неправды с семантикой предложения. Языковые парадоксы, как кажется, основываются на функциональных особенностях языка и не связаны с научной картиной мира, а только с опытом говорящего»