



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»

Н.Р. Ураева

КОРПУСНЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ

*Утверждено Редакционно-издательским советом университета
в качестве учебно-методического пособия*

Магнитогорск
2018

УДК 81.32
ББК 81.1-923

Рецензенты:

кандидат педагогических наук,
доцент кафедры философии, культурологии
и социально-гуманитарных дисциплин,
ГБОУ ВО ЧО «Магнитогорская государственная
консерватория (академия) им. М.И. Глинки»
А.Ю. Трутнев

кандидат педагогических наук, доцент,
доцент кафедры иностранных языков по техническим направлениям,
ФГБОУ ВО «Магнитогорский государственный технический
университет им. Г.И. Носова»
О.А. Лукина

Уразаева Н.Р.

Корпусные технологии в лингвистике [Электронный ресурс] : учебно-методическое пособие / Наиля Раdifовна Уразаева ; ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». – Электрон. текстовые дан. (1,09 Мб). – Магнитогорск : ФГБОУ ВО «МГТУ им. Г.И. Носова», 2018. – 1 электрон. опт. диск (CD-R). – Систем. требования : IBM PC, любой, более 1 GHz ; 512 Мб RAM ; 10 Мб HDD ; MS Windows XP и выше ; Adobe Reader 8.0 и выше ; CD/DVD-ROM дисковод ; мышь. – Загл. с титул. экрана.

ISBN 978-5-9967-1283-0

Учебно-методическое пособие нацелено на изучение онлайн-курса «Введение в корпусную лингвистику» на образовательной платформе «Открытое образование», разработанного Национальным исследовательским университетом «Высшая школа экономики» (автор – старший преподаватель факультета гуманитарных наук Левинзон Анна Иосифовна). Курс знакомит с концепциями корпусной лингвистики, дает возможность освоить основы корпусных технологий, приобрести навыки работы с корпусами, определить место дисциплины и собственно корпусов в ряду информационных технологий.

Учебно-методическое пособие предназначено для студентов направлений подготовки 44.03.01 Педагогическое образование; 45.03.01 Филология; 45.03.02 Лингвистика; 45.03.03 Фундаментальная и прикладная лингвистика; 45.03.04 Интеллектуальные системы в гуманитарной сфере, а также могут быть полезны для магистрантов, аспирантов филологических специальностей и всех интересующихся возможностями применения компьютерных технологий в лингвистике.

Данное учебно-методическое пособие содержит дополнительные задания в основном для изучающих немецкий язык.

УДК 81.32
ББК 81.1-923

ISBN 978-5-9967-1283-0

© Уразаева Н.Р., 2018
© ФГБОУ ВО «Магнитогорский государственный
технический университет им. Г.И. Носова», 2018

Содержание

ВВЕДЕНИЕ.....	4
ЦЕЛИ И ЗАДАЧИ КОРПУСНОЙ ЛИНГВИСТИКИ.....	6
ПЕРЕЧЕНЬ ОСНОВНЫХ РАЗДЕЛОВ.....	7
Тема 1. Интернет и развитие гуманитарного знания.....	8
Тема 2. Что такое «корпус»?.....	13
Тема 3. Возникновение и развитие корпусной лингвистики.....	19
Тема 4. Виды корпусов.....	23
Тема 5. Поиск информации в корпусе.....	25
Тема 6. Разметка корпусов.....	27
Тема 7. Исследование языка с помощью лингвистического корпуса: общие принципы.....	31
Тема 8. Исследование языка с помощью лингвистического корпуса: направления исследования и практическое применение результатов.....	35
Тема 9. Корпусное преподавание.....	37
Тема 10. Корпус и междисциплинарные исследования.....	49
ОЦЕНОЧНЫЕ СРЕДСТВА.....	53
Итоговый тест к видеолекциям.....	53
Тест для самоконтроля.....	60
ГЛОССАРИЙ.....	63
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	69
ПРИЛОЖЕНИЯ.....	70
Приложение 1. Список корпусов в интернете.....	70
Приложение 2. Классификация корпусов (по Захарову).....	72
Приложение 3. Правила пользования немецкими корпусами.....	73
ЗАКЛЮЧЕНИЕ.....	78

ВВЕДЕНИЕ

*Хочешь накормить человека один раз –
дай ему рыбу.*

*Хочешь накормить его на всю жизнь –
научи его рыбачить.*

Конфуций

Стремительное развитие информационных технологий ставит вопрос о новом историческом этапе развития общества, в котором информация и знания играют центральную роль и являются мощным мотивационным средством эволюции.

Важнейшей целью современного образования является формирование целостной системы универсальных знаний, умений, навыков, а также опыта самостоятельной деятельности и личной ответственности обучающихся, т. е. ключевых компетенций, определяющих современное качество содержания образования. Компетентностный подход в условиях высшего образования предполагает, в первую очередь, «умение работать с информацией и готовность к реальным жизненным ситуациям». Как справедливо заметил еще Д.И. Писарев, «кто дорожит жизнью мысли, тот знает очень хорошо, что настоящее образование есть только самообразование и что оно начинается только с той минуты, когда человек, распроставшись навсегда со всеми школами, делается полным хозяином своего времени и своих занятий».

Целенаправленное накопление сведений с возможностью их передачи оказало огромное влияние на развитие общества, культуры и, в частности, образования, радикально изменив всю организацию деятельности общества.

На место сдающей свои позиции печатной продукции в настоящее время пришли информационные технологии, позволяющие человеку получить несоизмеримо больше информации по интересующей его проблеме, обеспечивающие человеку безграничные возможности для получения, переработки и производства информации. Появились новые типы продуктов, новые виды услуг, внедряются модели интерактивного онлайн-взаимодействия. Информатизация образования развивается стремительно. Формируются новые виды и типы электронных образовательных продуктов, которые дополняют, а в некоторых случаях заменяют традиционные бумажные издания.

Следствием высоких темпов обновления научных знаний становится необходимость непрерывного образования на основе умения учиться. Индивидуализация и персонализация образовательного процесса становятся одной из ведущих тенденций. Все участники образовательного процесса имеют возможность учить и учиться в удобное время и в удобном месте.

Целью настоящего учебно-методического пособия является обеспечение эффективности самостоятельной работы студентов с онлайн-курсом «Введение в корпусную лингвистику» на образовательной платформе «Открытое образование» на основе рациональной организации её изучения. Задачи методических указаний по изучению учебной дисциплины:

- активизация самостоятельной работы студентов;
- помощь студентам в ориентации в учебном материале,
- выработка умений и навыков рациональной работы с материалом;
- управление познавательной деятельностью студентов;
- содействие развитию творческого отношения к учебной деятельности.

Функции методических рекомендаций по изучению курса: определение содержания работы студентов по самостоятельному овладению программным материалом курса; описание технологии самостоятельной работы; установление требований к результатам изучения курса.

Автор курса «Введение в корпусную лингвистику» (А.И. Левинзон) ставит перед собой две главные цели:

Во-первых, представить аудитории корпусную лингвистику – науку, развивающуюся в нашей стране всего полтора десятилетия, но давно использующуюся в обучении языкам, теории и практике перевода и литературоведении во всем мире. Корпусная лингвистика будет рассмотрена, с одной стороны, как отдельная область знания, с другой – как набор методов, которые могут быть применены в самых разных научных и практических областях.

Во-вторых, предлагается целый набор удачных практик, способных повысить эффективность работы с учениками и текстами. В процессе освоения курса, таким образом, студенты получают ряд практических навыков от навыка сбора информации в корпусах до умения составлять упражнения с использованием корпусных данных.

Формат

Курс состоит из коротких видеолекций продолжительностью 10-15 минут, внутри которых встроены нецениваемые вопросы и практические задания. Раз в неделю слушателям будет предложен оцениваемый тест из 15 вопросов или эссе в формате взаимного оценивания. Итоговый экзамен представляет собой тест.

Автор учебно-методического пособия выражает искреннюю признательность **старшему преподавателю факультета гуманитарных наук НИУ ВШЭ Левинзон Анне Иосифовне** за разработанный на высоком профессиональном уровне онлайн-курс «Введение в корпусную лингвистику», вызвавший огромный интерес и предоставивший материал для творчества в работе и ценный инструмент для дальнейших научных исследований.

ЦЕЛИ И ЗАДАЧИ КОРПУСНОЙ ЛИНГВИСТИКИ

Цель курса – познакомить студентов с концепциями корпусной лингвистики, дать им возможность освоить основы корпусных технологий, приобрести навыки работы с корпусами, определить место дисциплины и собственно корпусов в ряду информационных технологий.

Задачи курса:

- ознакомление студентов с новой парадигмой в лингвистических исследованиях;
- ознакомление студентов с историей корпусных исследований;
- изучение языковых и программных средств корпусной лингвистики;
- формирование навыков работы с программными средствами и информационными ресурсами корпусной лингвистики;
- ознакомление студентов с конкретными лингвистическими исследованиями, основанными на корпусных данных.

Курс вносит вклад в формирование следующих компетенций:

- способность применять полученные знания в области теории и истории основного изучаемого языка (языков) и литературы (литератур), теории коммуникации, филологического анализа и интерпретации текста;
- умение готовить учебно-методические материалы для проведения занятий и внеклассных мероприятий на основе существующих методик;
- владение методикой предпереводческого анализа текста, способствующей точному восприятию исходного высказывания.

Направления подготовки

44.03.01 Педагогическое образование

45.03.01 Филология

45.03.02 Лингвистика

45.03.03 Фундаментальная и прикладная лингвистика

45.03.04 Интеллектуальные системы в гуманитарной сфере

Результаты обучения

Слушатели курса научатся пользоваться существующими корпусами русского и английского языков, освоят базовые методы корпусного преподавания, смогут экономить время и усилия, используя корпуса в повседневной работе.

Для освоения курса слушателю желательно иметь базовые знания в области лингвистики и преподавания языков.

Для изучения курса понадобится 11 недель, от 4 до 5 часов в неделю.

Количество часов – 2 зачётных единицы.

ПЕРЕЧЕНЬ ОСНОВНЫХ РАЗДЕЛОВ

Тема 1. Интернет и развитие гуманитарного знания.

Тема 2. Что такое «корпус»?

Тема 3. Возникновение и развитие корпусной лингвистики.

Тема 4. Виды корпусов.

Тема 5. Поиск информации в корпусе.

Тема 6. Разметка корпусов.

Тема 7. Исследование языка с помощью лингвистического корпуса: общие принципы.

Тема 8. Исследование языка с помощью лингвистического корпуса: направления исследования и практическое применение результатов.

Тема 9. Корпусное преподавание.

Тема 10. Корпус и междисциплинарные исследования.

Тема 1. Интернет и развитие гуманитарного знания

План

- 1) Интернет-технологии и развитие гуманитарного знания: введение
- 2) [Digital humanities](#): теория и практика
- 3) Компьютерные технологии в переводе
- 4) [CAT](#): за и против

Просмотрите видеолекции на платформе «Открытое образование».

Задания для самостоятельной работы

Ответьте на следующие вопросы или обсудите их в группе.

1. Найдите в словарях слово «корпус». Какие значения оно имеет? Какова этимология этого слова? Как правильно оно произносится в различных грамматических формах?
2. Кто автор этой картины? Какое отношение она имеет к теме лекции?



3. Как развивалась история корпусов? Проследите основные этапы от картотеки к корпусу.
4. Найдите в интернете примеры виртуальных музеев и объясните, какие новые возможности открывают перед музеями достижения в области [digital humanities](#).
5. Познакомьтесь с виртуальным музеем [«Дрезденская галерея»](#). Какую информацию можно получить в виртуальном музее? Какие вы видите возможности в использовании подобных ресурсов на занятиях со школьниками или студентами?
6. У некоторых ученых присутствует тревога по поводу отчуждения человека от мира культуры и опасения по поводу дегуманизации реальности, подмены реального мира виртуальным, подлинного культурного наследия его цифровыми копиями и реконструкциями. Каковы плюсы и минусы [цифровой гуманитаристики](#)?
7. Что понимается под [digital humanities](#)? Какие примеры цифровой гуманитаристики приводятся в лекции? Можете ли вы привести еще примеры использования цифровой гуманитаристики?
8. Ознакомьтесь с проектом [«Живые страницы»](#). Как в нем реализуются принципы цифровой гуманитаристики?
9. Что такое [CAT](#)? Как можно использовать CAT в изучении и преподавании иностранных языков?

Рекомендуемая литература

1. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005.
2. Зубов А. В. Основы искусственного интеллекта для лингвистов: [учеб. пособие для студ. вузов, обуч. по спец. "Теоретическая и прикладная лингвистика"] / Зубов А. В., Зубова И. И. - М. : Логос, 2007.

Ссылки на интернет-ресурсы, упомянутые в лекции:

1. Galaxy Zoo: galaxyzoo.org
2. Zooniverse: zooniverse.org
3. Pynchon Wiki: pynchonwiki.com
4. Виртуальный музей ГУЛАГа: gulagmuseum.org
5. The Beatles Wiki: beatles.wikia.com
6. Древнерусские берестяные грамоты: gramoty.ru

7. Фейсбук прошлого / Facebook of the past: web-corpora.net/Facebook_of_the_Past/facebook.html
8. Обнинский цифровой проект: obninsk-project.net
9. Живые страницы: time2read.ru
10. Грамота.ру: gramota.ru
11. Электронные словари: lingvo.ru
12. Электронные словари: multitrans.com
13. Специальная лексика – словари: lexicool.com
14. Специальная лексика – словари + общение: uva.fi
15. IATE – InterActive Terminology for Europe: iate.europa.eu
16. UNTERM—United Nations Multilingual Terminology Database: unterm.un.org
17. UNESCOTERM: termweb.unesco.org
18. Google Translator Toolkit: translate.google.com

Литература:

1. Lynne Bowker, Gloria Corpas Pastor. Translation Technology: The Oxford Handbook of Computational Linguistics 2nd edition, 2015.
2. М. Копотев. Введение в корпусную лингвистику.
3. Richard Grusin. The Dark Side of Digital Humanities.

Выполните тест для самопроверки

1. Цифровая гуманитаристика – это

- а) раздел лингвистики, объектом изучения которого являются искусственные языки.
- б) реализация возможностей, которые появляются в связи с исследованием культуры при помощи компьютерных технологий.
- в) создание справочных сайтов, предназначенных для профессионалов, работающих в области гуманитарных наук.
- г) принципы создания электронных библиотек, призванных сохранить тексты в оцифрованном виде.

2. В цифровой гуманитаристике НЕ

- а) устанавливаются сложные связи между разными типами данных.
- б) усиливается междисциплинарная составляющая исследования.
- в) выявляются закономерности на обширном и статистически значимом материале.
- г) является обязательным представлением результата исследования в виде печатного текста.

3. В цифровой гуманитаристике ученый часто берет на себя роль куратора, то есть

а) предлагает аудитории понятные решения сложных проблем, актуальных для изучения культуры.

б) дает аудитории инструменты для пользования материалом и предлагает самой делать заключительные выводы.

в) с помощью компьютерных технологий следит за способами общения читателя и электронного текста.

г) представляет всю информацию, которую должна воспринять аудитория, в виде инфографики.

4. Критики цифровой гуманитаристики НЕ

а) считают, что она была изобретена для того, чтобы отделить гуманитарное знание от политики.

б) видят отрицательное влияние этой области знания на коммерциализацию научных центров.

в) признают за традиционными методами гуманитарных наук права на существование в эпоху компьютерных технологий.

г) согласны заменять творческую интерпретацию материала творчеством на уровне создания компьютерных программ.

5. К области цифровой гуманитаристики НЕ относится

а) Text Encoding Initiative.

б) Проект Живые страницы.

в) Проект Galaxy Zoo.

г) Проект Фейсбук прошлого.

Выберите все правильные варианты ответа.

6. Компьютерные технологии в переводе

а) позволяют повысить скорость работы переводчика.

б) связаны с созданием электронных словарей и платформ профессионального общения.

в) активнее используются при переводе художественных, чем деловых текстов.

г) могут эффективно использоваться мелкими компаниями.

7. Официальными электронными базами терминов являются

а) Terminology Forum University of Vaasa

б) United Nations Multilingual Terminology Database

в) Inter-Active Terminology for Europe

г) UNESCOTERM

8. В компьютерную систему помощи переводчику (CAT) могут входить

а) память, которую предстоит заполнить переводчику.

б) инструмент для обнаружения терминов.

- в) инструмент проверки орфографии.
- г) инструмент поиска непереведенных фрагментов.

9. Переводчикам удобно работать с САТ, поскольку система может

- а) отметить фрагменты, смысл которых был неверно понят.
- б) осуществить предварительный перевод.
- в) помочь выделить термины в тексте.
- г) дать рекомендации относительно сроков и оплаты.

10. Переводчики испытывают дискомфорт при работе с САТ, так как

- а) ощущают зависимость от чужого стиля.
- б) не чувствуют своего превосходства над непрофессионалами.
- в) перевод становится исключительно творческим занятием.
- г) боятся снижения оплаты труда.

Впишите слово или сочетание слов.

11. В цифровой гуманитаристике большие массивы текстового материала исследуются с помощью методов математической _____

12. Основателем цифровой гуманитаристики считается священник Роберто Буза, обратившийся к компании IBM за помощью в исследовании текстов _____

13. Методы _____ лингвистики используются не только в цифровой гуманитаристике, но и в компьютерных системах, помогающих переводчикам.

14. В процессе _____ система САТ делит тексты перевода и оригинала на фрагменты, а затем ставит сегменты в соответствие друг другу.

15. Недостаток системы САТ состоит, в частности, в том, что она отмечает как ошибки _____ особенности, которые переводчик включает в текст.

Тема 2. Что такое «корпус»?

План

- 1) Понятие «[лингвистический корпус текстов](#)»
- 2) Главные свойства корпуса: [репрезентативность](#)
- 3) Главные свойства корпуса: [сбалансированность](#) и наличие [разметки](#)
- 4) Интернет как корпус

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Создатели [НКРЯ](#) не включают в его основную часть переводы. Согласны ли вы с этой стратегией? В каких случаях переводы могут дать ценную информацию о состоянии языка?
2. Подумайте, какой корпус вы могли бы собрать для ваших исследований. Какие тексты должны в него войти? Как сделать такой корпус сбалансированным?
3. Пользуясь ресурсом [Linguee](#), найдите как можно больше переводов на русский язык глагола to challenge.
4. По ссылке [webcorpora.ru](#) познакомьтесь с ресурсом [ГИКРЯ](#). Какие цели ставят перед собой его авторы? Каким образом используют они интернет?
5. Пользуясь [НКРЯ](#), сравните частотности существительных «оптимист» и «пессимист» в XIX и XXI веках.
6. Пользуясь цифровой словарной системой немецкого языка [DWDS](#), проанализируйте частотность использования в немецком языке слов „Einwanderung“, „Zuwanderung“, „Migration“ за последнее столетие. Какие выводы можно сделать на основе полученных данных?
7. Ознакомьтесь со следующей информацией. Какие значения имеет слово «корпус» в немецком языке? Сопоставьте с русским языком. В чем заключаются особенности грамматической категории рода данного слова?

Ein **Korpus** ist lateinisch „Körper“. Selbst das deutsche Wort Körper ist eine sprachliche Entlehnung aus dem Lateinischen und fungiert laut Kluge (2002: 530) in der deutschen Sprache seit dem 13. Jh.:

„...mhd. korper, körper, fnhd. auch körpel mit Dissimilierung des zweiten r Entlehnung. Entlehnt aus l. corpus (-poris) n. "Leib" (der spätere Umlaut ist nicht ausreichend erklärt). Ersetzt die älteren Wörter Leib und Leiche.“

In der heutigen deutschen Sprache gehen mehrere Lexeme auf diese gemeinsame Wurzel (Etymon) zurück: Körper, Korps, Korpuskel, korpulent, und natürlich auch Korpus oder Corpus. Laut DUDEN (1996: 885-886 und 2006 [CD-ROM]) ist der Korpus die Bezeichnung des menschlichen Körpers) und das Korpus eine Sammlung bzw. der Körper eines Instruments:

1. Korpus, der; -, -se [lat. corpus, ↑Körper]: 1. (ugs. schrezh.) menschlicher Körper ... 2. (bild. Kunst) Christusfigur am Kruzifix. 3. <o. Pl.> (Fachspr.) (bei Möbeln) das massive, die eigentliche Gestalt ausmachende Teil ohne die Einsatzteile ... 4. (schweiz.) Ladentisch; [Büro]möbel mit Fächern ...;

2. Korpus, Corpus, das; -, Korpora bzw. Corpora [lat. corpus, = Gesamtwerk, Sammlung...]: 1. (Sprachw.) Sammlung einer begrenzten Anzahl von Texten, Äußerungen o. Ä. als Grundlage für sprachwissenschaftliche Untersuchungen. 2. <heute meist: der; o. Pl.> Klangkörper bes. eines Saiteninstrumentes;

3. Korpus, die; (Druckw.) Schriftgrad von 10 Punkt; Garmond.

Im Kontrast zu den angeführten Angaben bringt eine Recherche im größten deutschen Korpus/ Corpus Ergebnisse, die auf den folgenden Gebrauchssusus hindeuten:

- Das Schriftbild mit „C“ kommt in modernen deutschen Texten der letzten 20 Jahre nur in einer lateinischen Verbindung vor: z.B. Corpus delicti, Corpus Christi, Ave verum corpus, Corpus iuris Civilis.

- Bis auf Corpus Christi ist das Lexem immer nur sächlich. In der Verbindung Corpus Christi, bzw. nur elliptisch Corpus ist es ausschließlich männlich⁴.

- Korpus mit K geschrieben ist fast immer maskulin, egal ob es sich um einen menschlichen Körper, das Geschöpf eines Instrumentes oder eine Sammlung handelt (alle folgenden Belege sind aus dem DeReKo, W – Archiv der geschriebenen Sprache):

Die rhythmisch pointierte Sprache der eigenwilligen Geige kommt am eindrucksvollsten bei ihren perkussiven Elementen zur Entfaltung, wenn er die Saiten zupft und den Korpus als Trommel zusammenwirken läßt.

Eine repräsentative und hinreichend große Menge an verfügbaren schriftlichen Quellen, genannt der Korpus, aus einer Sprache wird als Grundlage für die Auswertung verwendet.

Deswegen wird auch in manchen Kinos von Wolfram Weber weiter ein analoger Zelluloid-Projektor stehen. Sogar wenn der gesamte Korpus der Filmgeschichte einst digitalisiert sein sollte.

- Ein seltener Fall ist das neutrale Genus von Korpus im Sinne einer Sammlung, wie das folgende Textbeispiel zeigt:

Kurt Ostbahn. [...] Nach drei überfüllten Abenden en suite hängt der echtste aller Wiener noch zwei Auftritte im "Zelt" an. Der Doppler nennt sich das Ereignis, bei dem das gesamte historische Korpus des Kurti-Schaffens - auf zwei Konzerte verteilt - ausgebreitet werden wird, natürlich ohne auch nur eine Nummer doppelt zu spielen.

Der nächste interessante Punkt im oben genannten Lexikoneintrag (Duden Universalwörterbuch 1996, und in allen späteren Versionen) stellt die Bedeutungserklärung dieses Lexems dar. An der ersten Stelle (dies bedeutet wohl die häufigste Verwendung) steht:

1. (ugs. schrezh.) menschlicher Körper ...

Sieht man das typische Umfeld von Korpus in der geschriebenen deutschen Sprache an, findet man keinerlei Hinweise auf umgangssprachliche oder scherzhafte Elemente: ein gusseiserner Korpus

Wegekrenz ... mit|bezeichnet ... Korpus bezeichnet ...

der|die Saiten und|auf den Korpus

spätbarockes Friedhofskrenz mit Korpus ... 1755

Friedhofskrenz mit Korpus

einen ... hohlen [...] Korpus der

mit hohlem Korpus

ein Kruzifix [mit ...] Korpus

des|Hohler Korpus ... die|der ... Zargen mit Tonabnehmer

mit massivem [...] Korpus

Der heutige Begriff Sprachkorpus bezeichnet eine elektronische Sammlung/Datenbank natürlicher Texte, die (meist) in voller Länge gespeichert wurden⁶. Diese Datenbank ist strukturiert und mit einem Korpusmanager versehen. Diese zwei Eigenschaften heben ein Korpus von einer einfachen Textdatenbank oder von einer elektronischen Bibliothek ab. (Dazu auch Káňa/ Peloušková 2005 und McEnery/Wilson 2001: 30). Als ein Sprachkorpus kann nicht jede (wenn auch wertvolle) elektronische Datenbank bezeichnet werden. Die Mindestanforderungen an die Suchmöglichkeiten in einem Sprachkorpus sind: Abfragen nach Wörtern und Wortteilen, statistische Angaben über die Ergebnisse, Angaben über die Größe und Zusammenstellung des Korpus. Ein Sprachkorpus ist auch kein Wörterbuch, reflektierte User können es jedoch als solches verwenden.

Ein Sprachkorpus ist also eine elektronische Textdatenbank, in der man effektiv nach sprachlichen Phänomenen suchen kann.

Ссылки на интернет-ресурсы, упомянутые в лекции:

- Corpus of Bilingual Russian child speech (BiRCh):
people.brandeis.edu/~smalamud/MILa/birch_ru.html
- Корпус М. В. Ломоносова: lomonosov.pro
- Весь Толстой в один клик: readingtolstoy.ru
- Национальный корпус русского языка: ruscorpora.ru
- MLibrary Digital Collections: quod.lib.umich.edu/lib/collist
- Google Books Ngram Viewer: books.google.com/ngrams
- Linguee: linguee.ru

Литература:

- В. Плунгян. Корпусная лингвистика – лекция на сайте «Постнаука».
- И. Левонтина. О чем речь.
- Д. Сичинава. Полк, дивизия, корпус русского языка.

Выполните тест для самопроверки

Выберите один или несколько верных вариантов ответа.

Будьте внимательны! В некоторых случаях правильными могут оказаться все предложенные варианты.

1. Тексты, составляющие лингвистический корпус,

- а) организованы для удобства чтения.
- б) сконструированы лингвистами.
- в) не включают речь необразованных носителей языка.
- г) размечены лингвистами и программистами.

2. В лингвистический корпус могут входить

- а) тексты электронной коммуникации.
- б) записи устной речи детей.
- в) транскрипции теле- и радиопередач.
- г) произведения художественной литературы.

3. Для того, чтобы электронная коллекция текстов могла считаться корпусом,

- а) в ней не должно быть текстов, посвященных неактуальным для общества проблемам.
- б) она должна отражать состояние языка в определенный период времени.

- в) она должны быть размечена.
- г) в ней должны быть исправлены ошибки носителей языка.

4. Обязательными свойствами хорошего лингвистического корпуса являются

- а) представительность.
- б) присутствие разметки.
- в) снятая грамматическая омонимия.
- г) баланс между разными типами текстов.

5. Создавая репрезентативный корпус, лингвисты должны ответить на вопросы:

- а) Какие тексты считать значимыми для данной культуры?
- б) Отражают ли оригинальные тексты художественной литературы состояние языка эпохи?
- в) Можно ли, изучая язык периода или жанра, опираться на переводные тексты?
- г) Какое количество текстов будет достаточным для демонстрации состояния языка конкретного периода?

Впишите слово или число.

6. Интерпретация корпусных данных невозможна, если исследователю неизвестен _____ корпуса, по которому велся поиск.

7. Согласно материалам лекции, основная часть Национального корпуса русского языка включает _____ слов.

8. Количество найденных результатов обозначается в НКРЯ как количество «_____».

10. Нормализация значений обычно предполагает вычисление отношения количества вхождений к объему корпуса, умноженного на _____.

Выберите единственный правильный вариант ответа.

11. Что означает понятие «накопительная сбалансированность»?

- а) В большом корпусе в некоторых случаях можно пренебречь представлением текстов разных типов в соотношениях, сопоставимых с соотношениями в естественном языке.
- б) С увеличением числа текстов, входящих в корпус, увеличивается необходимость разнообразить состав включаемых в него документов.
- в) Лингвистический корпус может быть относительно небольшим, важно, чтобы в нем были представлены разные подязыки в пропорциях, соответствующих аналогичным пропорциям в естественном языке.

г) Понятия «сбалансированность» и «репрезентативность» никак не связаны; хороший корпус должен обладать обеими характеристиками.

12. Что отличает разметку Обучающего корпуса от аннотации других частей НКРЯ?

- а) В ней присутствует синтаксическая разметка.
- б) Она проведена в соответствии со школьной грамматикой.
- в) Она позволяет вести поиск по речевым действиям.
- г) Она целиком произведена вручную.

13. Чем отличается лингвистический корпус от электронного словаря?

- а) Электронный словарь не предлагает контекстов употребления слова.
- б) В лингвистическом корпусе языковая норма отражена только опосредованно.
- в) Материалы лингвистического корпуса не дают представления о том, насколько слово распространено в языке.
- г) В электронном словаре могут содержаться сведения о семантике слова.

14. Какой вид поиска невозможен для пользователя в поисковых системах Яндекс и Google?

- а) Поиск в пределах одного сайта.
- б) Поиск точных форм.
- в) Поиск словоформ одной лексемы.
- г) Поиск смысловых категорий.

15. На каком ресурсе можно получить информацию о частотности употребления слова в конкретный период времени?

- а) Яндекс Переводчик
- б) Google Books Ngram Viewer верно
- в) ABBYY LINGVO
- г) Google Translator Toolkit

Тема 3. Возникновение и развитие корпусной лингвистики

План

- 1) Корпусные инструменты в докомпьютерную эпоху
- 2) Корпусная лингвистика во второй половине XX века
- 3) Первые корпуса
- 4) Корпуса русского языка

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Какую помощь в исследовании стихосложения может оказать филологу [конкорданс](#)?
2. Каким образом появление словарей американского языка, основанных на живом словоупотреблении, связано с политикой США в XIX веке?
3. Можно ли считать словарь, составленный В. Далем, словарем на основе живого словоупотребления? Обоснуйте свою точку зрения.
4. Прокомментируйте, высказывание Н. Хомского, записанное одним из его интервьюеров: Corpus linguistics doesn't mean anything.
5. Какие перспективы открывает перед вашим собственным исследованием «корпусная идеология»?
6. Ознакомьтесь с данными в приведенной ниже таблице. Какие преимущества имеют корпуса по сравнению с поисковыми системами?

Suchmaschine		Korpus
Geschwindigkeit	schnell	oft langsam
Benutzerfreundlichkeit	einfach, intuitiv	oft kompliziert, ohne Einleitung kaum verständlich
Daten(Text)menge	riesig	oft (sehr) eingeschränkt
Wesen der Sprache	nur geschrieben	oft nur geschrieben, aber auch gesprochen
Sprache der Texte	eingeschränkt eindeutig	eindeutig
Originalsprache d. Texte	kaum feststellbar	eindeutig feststellbar
Zeit der Verfassung	kaum feststellbar	eindeutig feststellbar
Textfunktion	keine Angaben	(eindeutig) zuordenbar
Verfasser/-in(nen)	schwer feststellbar	relativ eindeutig

Структуриertes Suchen im Text (in Texten)2	nicht möglich	möglich
Suche nach Wortteilen	eingeschränkt möglich	möglich
Statistische Angaben	fast keine (nur Anzahl der Ergebnisse)	vielfältig

Ссылки на интернет-ресурсы, упомянутые в лекции:

Список корпусов других языков на сайте НКРЯ: ruscorpora.ru/corpora-other.html

Lancaster-Oslo-Bergen Corpus: purl.ox.ac.uk/ota/0167

Упсальский и Тюбингенский корпуса: www.lingexp.uni-tuebingen.de/sfb441/b1/en/korpora.html

Машинный фонд русского языка: cfri.ruslang.ru

Литература:

Andor József. The Master and His Performance: an Interview with Noam Chomsky. Intercultural Pragmatics 1.1, 2004.

Tony McEnery, Andrew Wilson. Corpus Linguistics: An Introduction. Edinburgh: University Press, 2001.

Charles J. Fillmore. "Corpus linguistics" or "Computer-aided armchair linguistics".

В. Плунгян. Зачем нужен Национальный корпус русского языка. Неформальное введение.

В. Плунгян. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики.

Выполните тест для самопроверки

Впишите слово или сочетание слов.

1. _____ – это представленный в алфавитном порядке список всех слов текста с указанием мест в тексте, в которых это слово встречается.

2. В современных лингвистических корпусах аналогом симфоний служит формат выдачи результатов _____.

3. Словарь, составленный _____ является первым словарем американского английского, созданным на основе реального словоупотребления.

4. Авторы _____ грамматик описывают не норму, а реальное состояние языка.

5. Противники корпусной лингвистики в оценке языковых явлений предлагают опираться не на корпусные данные, а исключительно на интроспекцию и языковую _____.

Выберите все правильные варианты ответа.

6. Какие принципы составления лингвистического корпуса использованы создателями НКРЯ?

- a) Значимые выводы можно делать только на основании материалов корпуса достаточно большого объема.
- b) В национальном корпусе должны быть представлены несколько разных языковых регистров.
- c) Корпус создается для решения конкретной лингвистической задачи.
- d) Пользователь имеет возможность получить информацию о каждом входящем в корпус тексте.

7. В каких корпусах собраны тексты на русском языке?

- a) Корпус Браун
- b) Упсальский корпус
- c) СОСА
- d) Тюбингенский корпус

8. Какие утверждения описывают корпусную лингвистику в России?

- a) С момента своего появления развивается в тесном взаимодействии с иностранными коллегами.
- b) Выработала стандартизированный формат сравнения разных региональных вариантов языка.
- c) Ввела в научный обиход понятие «национальный корпус языка».
- d) Вырабатывает особую идеологию исследования языковых данных.

9. Какие утверждения описывают корпусную лингвистику в Великобритании?

- a) С момента своего появления ставила целью исследование современного языка.
- b) Впервые продемонстрировала методы машинной разметки текстов.
- c) Разрабатывается активными последователями Н. Хомского.
- d) Провозглашает первостепенную значимость интроспекции для изучения грамматических конструкций.

10. Какие области исследования являются особенно привлекательными для корпусных лингвистов?

- a) Изучение регистров языка.
- b) Описание и установление нормы.
- c) Изучение диахронических языковых изменений.
- d) Изучение контекстуальной вариативности лексических значений.

11. Какие корпуса дают материал для изучения американского английского языка?

- a) Корпус Браун

- b) BNC
- c) COCA
- d) LOB Corpus

Выберите единственный правильный вариант ответа.

12. Национальный корпус НЕ

- a) включает поэтических текстов.
- b) отражает диалектные особенности говоров.
- c) отражает состояние языка страны в конкретные временные периоды.
- d) отражает государственную идеологию страны.

13. Sample corpus

- a) должен состоять из образцово грамотных текстов.
- b) отражает состояние языка данной местности в данный период.
- c) не может включать целые тексты и составлен исключительно из фрагментов.
- d) не может быть создан на материале русского языка.

14. На материале сравнимых корпусов обычно изучается

- a) словообразование.
- b) морфология.
- c) синтаксис.
- d) вариативность.

15. Авторы некоторых корпусов отказываются включать в них драматические произведения, поскольку

- a) такие произведения сложно переводить в электронный формат.
- b) речь персонажей является имитацией устной речи и не может дать представления о реальном языке диалога.
- c) объем корпуса значительно увеличивается за счет введения в них драмы.
- d) драматические авторы редко вносят заметный вклад в развитие литературного языка.

Тема 4. Виды корпусов

План

- 1) [Классификация корпусов](#)
- 2) Корпуса устной и письменной речи
- 3) [Параллельные корпуса](#)
- 4) Общие и специфические корпуса

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Воспользовавшись описаниями проекта «[Один речевой день](#)» или других звуковых корпусов, выясните, какие не характерные для обычных корпусов экстралингвистические параметры текста значимы для корпусов устной речи.
2. Для чего в «паспорт» устного текста включают сведения об информанте? Какие его характеристики могут быть важны для корпусного исследования устной речи?
3. Каким образом не совпадающие в разных языках правила оформления прямой речи могут затруднить [выравнивание](#) текстов в параллельном корпусе?
4. Вспомните, какие тексты представляют национальный язык, с точки зрения создателей НКРЯ. В чем вы с ними согласны и в чем не согласны?
5. Какие возможности предоставляет работа с корпусом [МУРКО](#) для исследователей и практиков нефилологических специальностей?

Ссылки на интернет-ресурсы, упомянутые в лекции:

Рассказы о свидениях и другие корпуса звучащей речи: spokencorpora.ru

Звуковой корпус «Один речевой день»: model.org.spbu.ru

ParaSol: A Parallel Corpus of Slavic and other languages: parasolcorpus.org

«Слово о полку Игореве»: Параллельный корпус переводов: nevmenandr.net/slovo

Мультимедийный русский корпус (МУРКО): www.ruscorpora.ru/instruction-murco.html

Верные слова: vsrussian.com

MICASE: Корпус устного академического английского: quod.lib.umich.edu/m/micase

Литература:

А. Кибрик, В. Подлеская. К созданию корпусов устной русской речи: принципы транскрибирования.

Д. Добровольский. Какие задачи решают параллельные корпуса текстов? – лекция на сайте «Постнаука»

Создатели НКРЯ на сайте корпуса пишут: «Следует учитывать, что не всё, что может показаться неподготовленному пользователю ошибкой Корпуса, в действительности ей является». Опишите три гипотетические ситуации, в которых «ошибки», найденные неподготовленным пользователем в тексте, не надо исправлять.

Найдите в устном, параллельном или мультимедийном подкорпусе НКРЯ один пример такой «ошибки».

Тема 5. Поиск информации в корпусе

План

- 1) [Авторитетные корпуса английского языка](#)
- 2) [Национальный корпус русского языка](#): обзор ресурса
- 3) НКРЯ: поисковые возможности
- 4) Проекты на базе НКРЯ

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Пользуясь НКРЯ, составьте список из 20 низкочастотных [коллокаций](#). Каким образом с помощью такого списка можно выявить связь между усвоением коллокаций и частотой их появления в текстах учебника?
2. Создайте в НКРЯ [подкорпус](#), соответствующий вашим исследовательским задачам. Какие параметры текстов вы будете учитывать?
3. Пользуясь НКРЯ, докажите, что Андрей Болконский мог видеть единорогов.
4. Какие способы существуют в НКРЯ для того, чтобы выяснить, когда лексема вошла в язык?
 7. Докажите, что [проект «Русская грамматика»](#) является дескриптивной, а не нормативной грамматикой.
 8. Пользуясь корпусом DeReKo, ответьте на следующие вопросы:
 - В текстах какой страны встречается слово *Korruptions-Untersuchungsausschuss*?
 - К какой тематике относятся тексты, в которых встречается это слово?
 - Установите период, когда оно встречается наиболее часто.
 9. Пользуясь данными корпуса немецкого языка, определите управление выражения: *im Vergleich mit/oder zu*?

Ссылки на интернет-ресурсы, упомянутые в лекции:

Британский национальный корпус BNC: www.natcorp.ox.ac.uk

Корпус современного американского английского COCA: corpus.byu.edu/coca

News on the Web: corpus.byu.edu/now

The Diachronic Corpus of Present-Day Spoken English: www.ucl.ac.uk/english-usage/projects/dcpse

Корпус языка лондонских подростков: clu.uni.no/icame/colt

Корпус английского как lingua franca: univie.ac.at/voice
Русская корпусная грамматика: rusgram.ru
Словари на основе НКРЯ: dict.ruslang.ru
Проект RusVectōrēs: rusvectores.org/ru

Литература:

S. Hoffmann, H.-M Lehmann. Collocational Evidence from the British National Corpus. Language and Computers 30, 2000.

Летучий А.Б. Сравнительные конструкции. Материалы для проекта корпусного описания русской грамматики. На правах рукописи. М., 2011.

rusgram.ru/Сравнительные_конструкции

Описывая работу пользователя в лингвистическом корпусе, многие лингвисты употребляют латинское выражение *saveat emptor*, которое дословно переводится как «покупатель должен остерегаться». Чего именно должен «остерегаться» пользователь лингвистического корпуса? Подтвердите свою мысль двумя примерами из НКРЯ.

Тема 6. Разметка корпусов

План

- 1) [Экстралингвистическая](#) и [лингвистическая](#) разметка
- 2) [Морфологическая](#) разметка
- 3) [Синтаксическая](#) разметка
- 4) [Семантическая](#) разметка

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Почему разметка корпуса должна быть теоретически нейтральной?
2. Познакомившись с [Грамматическим словарем А. Зализняка](#), объясните, каким образом его данные служат основой для автоматической морфологической разметки текстов.
3. Приведите по два примера на каждый тип омонимии, которая не может быть автоматически снята при разметке.
4. Прокомментируйте высказывание авторов «морфологического стандарта НКРЯ»: «Русская морфология с точки зрения корпуса – более «либеральная» и более широкая морфология, чем та, которая представлена в нормативных грамматиках».
5. Подбирая контексты, пользователь НКРЯ отсеивает значительное количество «шума», если сочетает поиск по грамматическим категориям с поиском по семантическим категориям. Приведите пример поискового запроса подобного рода.

Ссылки на интернет-ресурсы, упомянутые в лекции:

Синтаксически размеченный корпус русского языка:
www.ruscorpora.ru/instruction-syntax.html
Грамматический словарь русского языка А. А. Зализняка: www.morfologija.ru
Wordtjes: wordnet.princeton.edu

Литература:

О. Ляшевская, В. Плунгян, Д. Сичинава. О морфологическом стандарте в Национальном корпусе русского языка.

Г. Кустова, О. Ляшевская, Е. Падучева, Е. Рахилина. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы.

Выполните тест для самопроверки

Выберите один или несколько верных вариантов ответа.

Будьте внимательны! В некоторых случаях правильными могут оказаться все предложенные варианты.

1. Утверждение «Корпус – это интерпретированный текст» означает, что

- a) при помощи разметки создатели корпуса передают пользователю свои теоретические установки.
- b) прежде, чем включать тексты в корпус, лингвисты тщательно изучают их и делают выводы об их содержании.
- c) начиная работу в корпусе, пользователь по умолчанию соглашается принять на веру принципы, положенные в основу аннотации.
- d) каждое языковое явление должно быть интерпретировано создателями корпуса как правильное или неправильное и соответствующим образом помечено.

2. Аннотирование – это

- a) обработка текста в целях облегчения его изучения.
- b) необходимый этап создания лингвистического корпуса.
- c) присвоение каждому элементу текста особого тега.
- d) сведения, размещенные на главной странице корпуса.

3. Утверждение «Разметка не несет новой информации» означает, что

- a) разметчики описывают характеристики, присущие элементам текста.
- b) принципы разметки – это внутренняя информация, важная только для создателей корпуса.
- c) теоретические установки, положенные в основу аннотации, должны быть максимально нейтральными.
- d) разметка не выводится на экраны пользователей корпуса.

4. Автоматическая разметка

- a) начинается с деления текста на минимальные фрагменты.
- b) невозможна на уровне семантики.
- c) предполагает отнесение словоформы к начальной форме.
- d) позволяет быстро провести аннотирование по частям речи.

5. Экстралингвистическая разметка

- a) предполагает демонстрацию сведений об авторе.
- b) не затрагивает содержания текста.
- c) позволяет вести поиск в отобранной по социолингвистическим параметрам группе текстов.
- d) в НКРЯ выражается в виде «паспорта текста».

6. Морфологическая разметка в НКРЯ

- a) произведена с помощью компьютерных программ.
- b) началась с выработки командой лингвистов стандартных решений в спорных ситуациях.
- c) единообразна во всех частях НКРЯ.
- d) предполагает присвоение тэгов каждой единице текста.

7. Синтаксическая разметка в НКРЯ НЕ

- a) проведена в соответствии с традиционным выделением членов предложения.
- b) графически представлена в виде дерева зависимостей.
- c) проведена компьютерной программой, затем доработана вручную.
- d) охватывает все части НКРЯ.

8. Семантическая разметка в НКРЯ

- a) помечает семантические признаки грамматических разрядов.
- b) помечает семантические признаки словообразовательных показателей.
- c) не затрагивает собственно семантические признаки.
- d) предполагает присвоение одной лемме одного семантического тега.

9. Утверждение «Мультимодальный корпус требует нескольких уровней разметки» предполагает, что

- a) для корпуса МУРКО в составе НКРЯ были разработаны дополнительные принципы описания текстов, например, создана разметка значений жестов.
- b) элементы корпуса, представляющие визуальные образы, должны быть описаны иначе, чем элементы письменного или аудио текста.
- c) мультимодальные корпуса должны быть частично закрытыми из соображений этики.
- d) разметка по цели высказывания или по дискурсивным элементам не нужна в обычных текстовых корпусах.

10. В корпусе с неснятой грамматической омонимией словоформа «уже» будет отнесена к

- a) одной лемме.
- b) двум леммам.
- c) трем леммам.
- d) четырем леммам.

Впишите слово.

11. В качестве минимального элемента разметки в НКРЯ принята _____, то есть графическая единица, не обладающая, в отличие от лексем, семантическими характеристиками.

12. Производимый обычно вручную процесс снятия омонимии носит название «_____».

13. _____ – это компьютерная программа, производящая синтаксическую разметку.

14. Многие автоматические анализаторы морфологии русского языка основаны на данных «_____ словаря» А. Зализняка.

15. Источником автоматической семантической разметки часто служат компьютерные _____.

Тема 7. Исследование языка с помощью лингвистического корпуса: общие принципы

План

- 1) Методы корпусных исследований
- 2) [Sketch Engine](#)
- 3) Corpus-based исследование
- 4) Corpus-driven исследование
- 5) Корпусные исследования в практике работы преподавателя

Посмотрите видеолекции.

Задания для самостоятельной работы

1. Используя в качестве образца вопросы, приведенные в работе [James Thomas Discovering English with Sketch Engine](#), выделите десять проблем, которые исследователь-русист может поставить и решить с помощью ресурса [Sketch Engine](#).
2. Проанализируйте известный вам учебник русского языка с точки зрения использования в нем данных о частотности языковых явлений.
3. Представьте, что вы составляете корпус для издательства, специализирующегося на выпуске учебников по академическому письму для университетов. Из каких частей будет состоять ваш корпус?
4. Из каких этапов должен состоять отбор ключевой лексики юридических текстов? Какие корпуса можно использовать в качестве референтных?
5. Предложите три направления использования корпусных методов в преподавании иностранного языка студентам университетов.

Ссылки на интернет-ресурсы, упомянутые в лекции:

Sketch Engine: sketchengine.co.uk

Sketch Engine for Language Learning: sketchengine.co.uk/skell

Литература:

James Thomas. *Discovering English with Sketch Engine*. Raleigh, NC: Lulu.com, 2015.

Douglas Biber et al. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

A. Kutter, C. Kantner. Corpus-based Content Analysis: A Method for Investigating News Coverage for War and Intervention.

Jane Demmen, Elena Semino, Zsófia Demjén, Veronika Koller, Andrew Hardie, Paul Rayson and Sheila Payne. A computer-assisted study of the use of Violence metaphors for cancer and end of life by patients, family carers and health professionals.

Paul Baker, Costas Gabrielatos and Tony McEnery. Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998–2009.

A. Piperski, A. Kukhto. Intra-speaker stress variation in Russian: A corpus-driven study of Russian poetry.

D. Biber, S. Conrad. Corpus Linguistics and Grammar Teaching.

Выполните тест для самопроверки

Впишите слово.

1. С помощью _____ методов исследователь выясняет, случайное или закономерное явление он наблюдает.

2. Результаты корпусных исследований обычно представляются в виде _____ значений, то есть количества слов на миллион.

3. Слова, которые встречаются в данном корпусе или тексте чаще, чем они обычно встречаются в языке, называются _____ словами.

4. Выявить такие частотные слова можно только используя два корпуса: тот, на материале которого проводится исследование, и _____ корпус.

5. Современные методики английского языка предлагают ученику запоминать не только отдельные лексемы, но и неслучайные сочетания слов, то есть _____.

Выберите один или несколько верных вариантов ответа.

Будьте внимательны! В некоторых случаях правильными могут оказаться все предложенные варианты.

6. В ходе corpus-based исследования

a) в корпусе подбираются примеры, подтверждающие гипотезу автора.
b) автор формулирует гипотезу только после анализа результатов, полученных в корпусе.

c) автор проверяет и дополняет свою гипотезу.

d) автор действует методом индукции.

7. Corpus-driven подход к изучению языка

- a) предлагает отбросить исключения и сосредоточиться на анализе правил.
- b) выстраивает гипотезу, базируясь на реальном словоупотреблении.
- c) может быть назван дедуктивным.
- d) может отвергать значимость анализа лексем и основываться на изучении словоформ.

8. Материалом corpus-driven исследования часто становится

- a) группа контекстов, демонстрирующих языковую норму.
- b) языковое явление, которое можно назвать ключевым для изучаемой группы текстов
- c) референтный корпус.
- d) языковое явление, которое демонстрируют тексты художественной литературы.

9. Результаты corpus-based и corpus-driven исследований

- a) не позволяют изменить и расширить существующие лингвистические теории.
- b) обычно не согласуются с выводами, сделанными традиционными лингвистами.
- c) подтверждают мысль о невозможности анализа языковых явлений вне конкретного регистра.
- d) всегда основаны на анализе частотности изучаемого явления.

10. Лингвистический корпус может помочь преподавателю языка

- a) определить содержание учебного курса.
- b) установить порядок изучения материала.
- c) определить количество заданий, необходимое для освоения материала.
- d) найти иллюстрации для изучаемых правил.

Выберите единственный правильный вариант ответа.

11. Величина частотности конструкций

- a) может быть оценена с помощью интроспекции.
- b) определяется интуитивно.
- c) выясняется с помощью статистических методов.
- d) не значима в процессе обучения языку.

12. Учебные пособия, базирующиеся на корпусных данных,

- a) не используются в младшей и средней школе.
- b) часто содержат сведения о вариативности.
- c) не могут быть использованы при обучении родному языку.

d) предполагают знакомство ученика с методами работы в лингвистическом корпусе.

13. Кембриджский корпус английского языка

- a) включает тексты людей, плохо пишущих по-английски.
- b) не имеет коммерческой ценности.
- c) используется всеми британскими издательствами.
- d) находится в открытом доступе.

14. Национальный корпус русского языка

- a) будучи общедоступным ресурсом, не может применяться в преподавании.
- b) является ценным источником примеров для составителей учебных материалов.
- c) стал основой для нескольких серий школьных учебников.
- d) может использоваться в педагогической практике только учителями, имеющими серьезную лингвистическую подготовку.

15. Преподаватель иностранного языка, пользующийся корпусными методиками, НЕ

- a) структурирует время, выделяемое на изучение материала, в зависимости от частотности явления.
- b) предлагает ученикам сведения об отличиях устной речи от письменной.
- c) предлагает для запоминания коллокации.
- d) ограничивается анализом нормативных языковых единиц.

Тема 8. Исследование языка с помощью лингвистического корпуса: направления исследования и практическое применение результатов

План

- 1) Изучение лексики
- 2) Изучение коллокаций
- 3) Изучение грамматики
- 4) Пример корпусного исследования

Посмотрите видеолекции.

Задания для самостоятельной работы

1. Сформулируйте свое отношение к эффективности концепции Basic English в преподавании иностранного языка.
2. Какие сложности приходится преодолевать составителям списков частотности?
3. Оправдано ли, на ваш взгляд, определение «революционный», данное издательством Longman грамматике под редакцией D. Biber?
4. Прочитайте инструкцию по пользованию Словарем глагольной сочетаемости непредметных имен русского языка О. Л. Бирюк, В. Ю. Гусева, Е. Ю. Калининой и выясните, каким образом авторы используют понятие «лексическая функция».
5. Дополните пятью собственными вопросами представленный в лекции список исследовательских проблем в области грамматики академического языка.

Ссылки на интернет-ресурсы, упомянутые в лекции:

General Service List: eapfoundation.com/vocab/general/gsl/frequency

New General Service List: newgeneralservicelist.org

Longman Corpus Network: pearsonlongman.com/dictionaries/corpus/index.html

Литература:

John Sinclair. Corpus, Concordance, Collocation. Oxford: Oxford Univ. Press, 1991.
Susan Hunston, Gill Francis. Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2000.

С. Шаров, О. Ляшевская. Введение к частотному словарю современного русского языка.

А. Пиперски. Лексическая сочетаемость – лекция на сайте «Постнаука»

David Allen. Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus.

Richard Xiao, Tony McEnery. Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective.

Bill Louw. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies.

Jason Fuqua. Semantic Prosody: The Phenomenon of "Prosody" in Lexical Patterning.

John M. Swales. Metatalk in American academic talk: The case of point and thing.

Найдите и исправьте в тексте 10 ошибок.

С помощью частотного словаря можно выяснить, какие слова часто используются в языке, а какие употребляются редко. Для того, чтобы этот показатель был более достоверным, частота слова подсчитывается на основе небольшого корпуса текстов. Словарь О. Ляшевской и С. Шарова является первым словарем подобного рода, изданным на русском языке. Основу словаря составил корпус художественной литературы, входящий в НКРЯ. Принятый в словаре показатель общей частоты (ipm) позволяет понять, сколько раз слово встретилось в одном тексте. Эта величина вводится для того, чтобы сделать проще сравнение частоты в разных корпусах. Прежде чем составить словарь, его авторы приписали каждой словоформе начальную форму (в корпусной лингвистике ее принято называть «лексемой»). Затем начальные формы получили свои частиречные характеристики. В частотном словаре, так же, как в НКРЯ, принята особая морфологическая классификация, не совпадающая с принципами Грамматического словаря русского языка А. Зализняка. Русский язык представляется особенно удобным материалом для составления частотных словарей, поскольку в связи с богатым словоизменением в нем почти отсутствует омонимия словоформ. В словарях предшествующего поколения омонимия разрешалась вручную. Очевидно, что и для 100-миллионного корпуса такое решение является оптимальным. В словаре О. Ляшевской и С. Шарова представлен полный список коллокаций для всех частотных существительных русского языка. Словарь может быть использован для проведения разного рода лингвистических исследований, однако, представленной в нем информации пока недостаточно для того, чтобы применить ее в области преподавания.

Тема 9. Корпусное преподавание

План

- 1) Корпусное преподавание: направления
- 2) Корпусное преподавание: принципы
- 3) «Легкий» и «трудный» способы работы с корпусом
- 4) Составление упражнений на базе корпусов

Просмотрите видеолекции.

Задания для самостоятельной работы

1. Предложите три варианта использования [Корпуса русских учебных текстов](#) в процессе преподавания культуры устной и письменной речи студентам университетов.
 2. Изучите доступный вам современный учебник английского языка известного британского издательства. Каким образом в этом учебнике отражены принципы корпусного преподавания?
 3. Каким образом learner corpora могут помочь в обучении переводчиков?
 4. В тексте пьесы М. Булгакова «Иван Васильевич» найдите три фрагмента, в которых речь персонажей указывает на то, что герои прибыли в эпоху Ивана IV из советской России. Выполняя задание, пользуйтесь НКРЯ.
10. Придумайте три корпусных упражнения для освоения интересующей вас дисциплины.

Практические задания

1. Рассмотрите значения слова *Haushalt*, представленные в нижеследующих контекстах. Выделите лексико-семантические варианты, предложите краткое толкование для каждого, укажите номера примеров, относящиеся к каждому лексико-семантическому варианту. Определите, имеет ли место омонимия или только многозначность.

a) Natürlich legt ihr dieser Besitz Pflichten auf, deren Schwere und Größe sich nach dem einfacheren oder komplizierteren Zuschnitt ihres Haushalts richten und nach deren getreuer oder nachlässiger Erfüllung man ihre Gewissenhaftigkeit beurteilen wird (Baudissin, Wolf von u. Baudissin, Eva von: Spemanns goldenes Buch der Sitte. In: Zillig, Werner (Hg.), Gutes Benehmen, Berlin: Directmedia Publ. 2004 [1901], S. 1873).

b) In Moskauer Mietwohnungen ist die Küche noch sehr beliebt als Schlafräum, die eventuellen Kinder der Mädchen, die der fromme Russe zum Haushalt zählt und stillschweigend duldet, werden in Backtrögen untergebracht (Baudissin, Wolf von u. Baudissin, Eva von: Spemanns goldenes Buch der Sitte. In: Zillig, Werner (Hg.), Gutes Benehmen, Berlin: Directmedia Publ. 2004 [1901], S. 1903).

c) Man betrachte einmal den Haushalt Gottes! (Pöhlmann, Robert von: Geschichte der sozialen Frage und des Sozialismus in der antiken Welt - Zweites Buch. In: Geschichte des Altertums, Berlin: Directmedia Publ. 2001 [1901], S. 14287)

d) Auf der Tagesordnung steht zunächst die zweite Berathung eines Gesetzes betreffend die Kontrolle des Reichshaushalts, des Landeshaushalts von Elsaß-Lothringen und des Haushalts der Schutzgebiete (Berliner Tageblatt (Morgen-Ausgabe), 14.02.1902).

e) Welche Frau, die einen Haushalt führt, wird dies nicht von sich sagen können, daß sie sich den ganzen Tag bewegt, treppauf, treppab usw.? (Reile, Bonifaz (Hg.), Kneipp, Sebastian. Das große Kneippbuch, München: Beckstein 1939 [1903], S. 85)

f) Dort ist alles in Grund und Boden verhagelt, was für den ländlichen Haushalt ein großer Übelstand ist, während die Stadtleut vor wie nach zu Markt gehn, als ob gar nichts paßirt ware (Brief von Wilhelm Busch an Letty Keßler vom 27.07.1905. In: ders., Gesammelte Werke, Berlin: Directmedia Publ. 2002 [1905], S. 5656).

g) Die Tätigkeit, die ihr der elterliche Haushalt in dieser Hinsicht bietet, genügt ihr denn auch nicht; sie gibt als Umziehergehilfe Gastspiele in der Nachbarschaft (Ernst, Otto: Appelschnut, Rostock: Hirnstoff 1989 [1907], S. 74).

h) Hier verlautet, Girardi beabsichtige, seinen hiesigen Haushalt aufzulösen, um ganz nach Berlin überzusiedeln (Berliner Tageblatt (Sonntags-Ausgabe), 08.03.1908).

i) Allenfalls ist noch, in Sachen Scheidung ist die DDR wirklich liberal, der Single-Haushalt mit oder ohne Kinder vorstellbar (Ketman, Per u. Wissmach, Andreas: DDR - ein Reisebuch in den Alltag, Reinbek bei Hamburg: Rowohlt 1986, S. 178).

j) Die in Zusammenhang mit Tschetschenien entstehenden Kosten würden im Rahmen des Haushalts finanziert; zusätzliche Mittel seien nicht erforderlich (Archiv der Gegenwart, 2001 [1995]).

k) Aber nur 43 Prozent der Haushalte besitzen Aktien und Anteile an Aktienfonds (Die Zeit, 05.01.2000, Nr. 2).

l) Die Maschine stammte noch aus Hans-Georgs erstem Haushalt, und allein die Tatsache, dass seine Frau sie ihm zurückgelassen hatte, hätte eine Warnung vor ihrer Benutzung sein sollen! (Die Zeit, 24.02.2000, Nr. 9)

m) Die theaterfreudige Jugend kostümiert sich in den unerschöpflichen Requisiten- und Kleiderkammern der Theater und aristokratischen Haushalte, und doch gilt als das modernste Kostüm die Armeekluse (Schlögel, Karl: Petersburg, München Wien: Carl Hanser Verlag 2002, S. 475).

2. Рассмотрите значения слова *schwarz*, представленные в нижеследующих контекстах. Выделите лексико-семантические варианты (ЛСВ), предложите краткое толкование для каждого ЛСВ, укажите номера примеров, относящиеся к каждому ЛСВ. Определите, имеет ли место омонимия или только многозначность. Укажите прямое значение, идиоматизированные словосочетания, окказиональные ЛСВ. Укажите тип мотивированности каждого из не прямых значений (метафорический или метонимический перенос).

a) Er drückt ihr das Buch in die Hand, da könne sie seine Erlebnisse in allen Einzelheiten nachlesen, dahabe sie es schwarz auf weiß (Beyer, Marcel: Spione, Köln: DuMont 2000, S. 171).

b) Der Helm, die Weste und die schwarze Uniform, an der noch Daunen haften (Beyer, Marcel: Spione, Köln: DuMont 2000, S. 210).

c) Die schwarzen Falken mit ihren unregelmäßigen Flügeln, mit ihren Fehlern in den Zacken und verschnittenen Köpfen sind schon lange von den großen Terrassenscheiben gekratzt und durch gekaufte Folienraubvögel ersetzt (Beyer, Marcel: Spione, Köln: DuMont 2000, S. 211).

d) 42 Prozent der Deutschen, so das Ergebnis einer Focus-Umfrage, haben schon schwarzarbeiten lassen (Die Zeit, 17.02.2000, Nr. 8).

e) Denn ernstgemeint wäre das genauso glaubhaft gewesen wie jemand, der es gewagt hat, in der Straßenbahn schwarzzufahren und deshalb plötzlich anfängt, sich für einen genialen Trickbetrüger zu halten (Kopetzky, Steffen: Grand Tour, Frankfurt am Main: Eichborn 2002, S. 520).

f) Bei frischen Brezeln und schwarzen Bieren, bei Hase und Wildschwein und Wachteln verging Lottes Vater (Koneffke, Jan: Paul Schatz im Uhrenkasten, Köln: DuMont Buchverlag 2000, S. 197).

g) Seine Diagnose mag problematisch sein, aber weil er so schwarz sieht, sehe ich - wenn auch nur für ihn – rosig (Die Zeit, 23.03.2000, Nr. 13).

h) Wenn man es nicht sieht, kann man grübeln, bis man schwarz wird - man siehts nicht (Schaeffer, Albrecht: Helianth II, Bonn: Weidle 1995 [1920], S. 493).

i) Tatsächlich stand das Schiff eines italienischen Reeders, das unter maltesischer Flagge fuhr, beispielsweise bei Shell seit 1997 auf der schwarzen Liste. Es transportierte diesmal Öl für die französische Totalfina (Die Zeit, 20.01.2000, Nr. 4).

j) Mit solchen Betrügereien muß jeder rechnen, der sich nicht schämt, auf dem schwarzen Markt zu kaufen (Süddeutsche Zeitung, 1995 [1945]).

k) Marthe Robert hat durch ihre Übersetzungen und durch Diskussionen dafür gesorgt, daß Kafka in Frankreich nicht mehr als der als Dichter verkleidete Philosoph, als unmittelbarer Nachbar von Hegel und Heidegger, als Beweisstück für surrealistische und existentialistische Theorie und für schwarzen Humor gelesen, zitiert und interpretiert wird, als ein Alchimist, "der unheimliche Mischgetränke in einem dunklen Zimmer in Prag zusammenbraute" (Die Zeit, 11.03.1966, Nr. 11).

3. Рассмотрите значения слова *Licht*, представленные в нижеследующих контекстах. Выделите лексико-семантические варианты (ЛСВ), предложите краткое толкование для каждого ЛСВ, укажите номера примеров, относящиеся к

каждому ЛСВ. Определите, имеет ли место омонимия или только многозначность. Укажите прямое значение, идиоматизированные словосочетания, окказиональные ЛСВ. Укажите тип мотивированности каждого из не прямых значений (метафорический или метонимический перенос).

a) So stand die Sonne der Nachtfee gegenüber, und der Saal war voll von ihrem Licht (Bassewitz, Gerdt von: Peterchens Mondfahrt, München: Dt. Taschenbuch-Verl. 1991 [1900], S. 58).

b) Mildes Licht, von Millionen Kerzen, die man nicht sah, floß ihnen entgegen, und an der Hand des Sandmännchens traten sie mit klopfenden Herzen über die Schwelle zur Weihnachtswiese (Bassewitz, Gerdt von: Peterchens Mondfahrt, München: Dt. Taschenbuch-Verl. 1991 [1900], S. 77).

c) Brummend nahm die Magd das Licht und schlörppte durch den langen Gang (Ganghofer, Ludwig: Der Dorfapostel, Stuttgart: Adolf Bonz 1917 [1900], S. 143).

d) Sie haben auch vor diesen den gewaltigen Vorzug, daß sie keine Briefe schreiben und sich deshalb den Kopf nicht zu zerbrechen brauchen, welche Farbe und Form des zu wählenden Briefpapiers am geeignetsten erscheint, ihre persönliche Distinktion ins rechte Licht zu setzen (Der Tag, 08.01.1901).

e) Es ist mithin auch nicht zu verwundern, daß ihre Moskauer Schüler keine Lichter wurden; nur in den Gebrüdern Subow macht sich ein Stück guter Begabung und braven Könnens geltend; sie kopierten meist Bilder aus der holländischen Bibel (Der Tag, 08.01.1901).

f) Ich aber, gleich der Motte im Lichte, mache meinen Selbsterhaltungs-Trieb zu nichte! (Altenberg, Peter: Was der Tag mir zuträgt. In: Deutsche Literatur von Lessing bis Kafka, Berlin: Directmedia Publ. 2000 [1901], S. 693)

g) Die EU-Kommission habe offenbar "grünes Licht" für einen "Bummelstreik" zur Verzögerung der Beitrittsverhandlungen erhalten, schrieb Gazeta Wyborcza (Die Zeit, 13.04.2000, Nr. 16).

h) Die Wahrheit über Parteispenden und schwarze Konten muss auf legale Weise ans Licht kommen (Die Zeit, 06.04.2000, Nr. 15).

i) Wenn das Kind das Licht der Welt erblickt hat, schläft es fast ununterbrochen, und es wacht gewöhnlich nur durch die Einwirkung des Hungers auf (Schädel, E.: Das Sprechenlernen unserer Kinder, Leipzig: Brandstetter 1905, S. 30).

j) Sebastian Kastner bewunderte, was sein Meister Franz Flaucher in dieser schlimmen Nacht getan hatte, wie er den Kutzner hinters Licht führte und Bayern und das Reich vor dem Ärgsten rettete (Feuchtwanger, Lion: Erfolg. In: ders., Gesammelte Werke in Einzelbänden, Bd. 6, Berlin: Aufbau-Verl. 1993 [1930], S. 705).

k) Die Wahrheit wird geliebt und gehaßt, verteidigt und geflohen; seit alter Zeit steht man bei der Wahrheit, steht ihr bei; man strebt, dürstet nach der Wahrheit (veritatem quaerere); man streitet für die Wahrheit; man gibt der Wahrheit die Ehre, man gibt ihr Raum (to give way to truth), man weicht der Wahrheit (se rendre a la force de la vérité); umgekehrt kann man der Wahrheit im Lichte stehen (to hinder the light of truth), der Wahrheit Gewalt tun, die Wahrheit unterdrücken, gegen die Wahrheit (je nachdem) taub oder blind sein.

Ознакомьтесь с примерами составления упражнений на базе корпуса (НКРЯ).

Упражнение 1

Одна из трудностей немецкого языка – это спряжение сильных глаголов. Помимо изменений личных окончаний в настоящем времени (Präsens) некоторые глаголы в определенных формах меняют корневой гласный и имеют чередующиеся согласные. Возьмем в качестве примера для отработки глагол «eintreten» («входить»).

Выбираем в НКРЯ параллельный подкорпус, немецкий язык, в котором будет вестись поиск контекстов. Далее задаем грамматические параметры поиска: глагол «eintreten» в изъявительном наклонении, в настоящем времени (Präsens), в действительном залоге; дополнительные признаки и языки – немецкий.

Выберем для отработки спряжения глагола упражнение на заполнение пропусков — дается предложение с пропуском, студенту необходимо заполнить его, вставив целиком правильную словоформу.

Выбираем из выдачи примеры.

1) Er wird erst munter, als das Eiserne Pferd eintritt. [Erich Maria Remarque. Der schwarze Obelisk (1956)]

Он оживает только, когда входит Железная Лошадь. [Эрих Мария Ремарк. Чёрный обелиск (В. Станевич, 1961)]

2) «Was in Deutschland geschah, geschieht in jedem Land, wenn gewisse Bedingungen eintreten. [Friedrich Dürrenmatt. Der Verdacht (1953)]

То, что произошло в Германии, может произойти в любой стране, если появятся известные предпосылки. [Фридрих Дюрренматт. Подозрение (Н. Савинков, 1990)]

3) Wenn die Johanna eintritt, ein sogenanntes Juwel, so wird mir angst und bange. [Theodor Fontane. Effi Briest (1894-1895)]

Когда в комнате появляется Иоганна, наше так называемое «сокровище», мне становится не по себе. [Теодор Фонтане. Эффи Брист (Ю. Светланов, Г. Эгерман, 1960)]

4) Ich schreite von Gemach zu Gemach, aber einsam – nur eine heilige Margarita steht jetzt schon auf dem Hausaltare, und grüßt mich, wenn ich eintrete, mit dem goldenen Schimmer. [Adalbert Stifter. Die Mappe meines Urgroßvaters (1847)]

В полном одиночестве перехожу я из комнаты в комнату, и только святая Маргарита, уже водруженная на домашний алтарь, приветствует меня золотистым сиянием. [Адальберт Штифтер. Записки моего прадеда (Р. Гальперина, 1971)]

5) – Zu deinem Oheime mußst du nun schon die Reise machen, ehe du in das Amt eintrittst, weil er es so wünscht. [Adalbert Stifter. Der Hagestolz (1844)]

Но раньше, чем ты поступишь на службу, отправляйся к твоему дяде, раз такова его воля. [Адальберт Штифтер. Холостяк (И. Татарина, 1971)]

Упрощаем примеры:

- 1) Er wird erst munter, als das Eiserne Pferd eintritt.
- 2) Was in Deutschland geschah, geschieht in jedem Land, wenn gewisse Bedingungen eintreten.
- 3) Wenn die Johanna eintritt, so wird mir angst und bange.
- 4) Ich schreite von Gemach zu Gemach, aber einsam – nur eine heilige Margarita ... grüßt mich, wenn ich eintrete, mit dem goldenen Schimmer.
- 5) Zu deinem Oheime mußt du nun schon die Reise machen, ehe du in das Amt eintrittst.

Представляем примеры в виде упражнения.

Вставьте глагол «eintreten» в форме Präsens:

- 1) Er wird erst munter, als das Eiserne Pferd _____.
- 2) Was in Deutschland geschah, geschieht in jedem Land, wenn gewisse Bedingungen _____.
- 3) Wenn die Johanna _____, so wird mir angst und bange.
- 4) Ich schreite von Gemach zu Gemach, aber einsam – nur eine heilige Margarita ... grüßt mich, wenn ich _____, mit dem goldenen Schimmer.
- 5) Zu deinem Oheime mußt du nun schon die Reise machen, ehe du in das Amt _____.

Упражнение 2

При изучении немецкого языка обучающиеся часто испытывают трудности с глаголами «(an)bieten», «vorschlagen», «auffordern», которые переводятся как «предлагать», в русском языке значение унифицировано. Типичные ошибки вызываются тем, что немецкому слову приписывается более широкая семантика их русского эквивалента.

Употребление немецких эквивалентов может быть выявлено студентом самостоятельно с использованием параллельного корпуса путем анализа коллокаций. Это так называемый «трудный путь» (DDL), который может быть задан преподавателем. Очевидно, что значения глагола «предлагать» в сочетаниях «предложить чашку кофе», «предложить покинуть зал» и «предложить пойти в кино» разные. Исходя из словосочетаемости, которая носит ярко выраженный семантический характер, можно выделить значения «дать», «пригласить», «посоветовать», «потребовать» и др. Эти значения будут выражаться в немецком языке разными глаголами.

Можно предложить студентам упражнение на нахождение соответствий (так называемые Match collocations). Задание составляется на основе параллельного корпуса (немецкий язык).

В лексико-грамматическом поиске вбиваем глаголы vorschlagen|anbieten|auffordern и затем выбираем контексты:

- 1) Laut Sozialgesetzbuch ist der Schutz zunächst Aufgabe des Jugendamtes, das die Gefährdung einschätzt und Hilfe anbietet. [Диана Энгель. Дети страдают в первую очередь! Совет и помощь при бытовом насилии (2009)]
- 2) Und da sie in Savathi ankamen, ward ihnen gleich im ersten Hause, vor dessen Tür sie bittend stehen blieben, Speise angeboten, und sie nahmen Speise an,

und Siddhartha fragte die Frau, welche ihnen die Speise reichte: [Hermann Hesse. Siddhartha (1922)]

3) Er wurde ein wenig ausgelacht, und andererseits wurden ihm auch Dinge angegeben, die er noch zu verändern und hinein zu machen hätte. Auch bei mir wurden Verbesserungen vorgeschlagen. Als wir beide mit unsern Ausfeilungen fertig waren, wurden in dem Zimmer, in welchem wir gezeichnet hatten, die Geräte wieder an den Platz gerückt, und die Staffelei und unsere Malergerätschaften wurden daraus entfernt. [Adalbert Stifter. Der Nachsommer (1857)]

4) der Vater lehnte an der Tür, die rechte Hand zwischen zwei Knöpfe des geschlossenen Livreerockes gesteckt; die Mutter aber erhielt von einem Herrn einen Sessel angeboten und saß, da sie den Sessel dort ließ, wohin ihn der Herr zufällig gestellt hatte, abseits in einem Winkel. [Franz Kafka. Die Verwandlung (1912)]

5) " Meinen Sie, daß man sie zum Tanzen auffordern kann? [Erich Maria Remarque. Der schwarze Obelisk (1956)]

6) Er fuhr fort: «Hast du auf dem Wege vorgeschlagen, einen förmlichen Zug zu ordnen und ein Lied zu singen?» [Gottfried Keller. Der grüne Heinrich (zweite Fassung) (1879-1880)]

7) K. tat so, wartete dann aber noch, bis Fräulein Bürstner ihn aus ihrem Zimmer nochmals leise aufforderte zu kommen. [Franz Kafka. Der Prozess (1914)]

Упрощаем примеры:

- 1) Hilfe anbietet
- 2) Speise angeboten
- 3) Verbesserungen vorgeschlagen
- 4) einen Sessel angeboten
- 5) zum Tanzen auffordern
- 6) vorgeschlagen, ein Lied zu singen
- 7) aufforderte zu kommen

Представляем примеры в виде упражнения.

Соедините правую и левую части коллокаций. Составьте предложения с полученными словосочетаниями.

anbieten _____	Hilfe
vorschlagen _____	Speise
auffordern _____	Verbesserungen
_____	einen Sessel
_____	zum Tanzen
_____	ein Lied zu singen
_____	zu kommen

Упражнение 3

Многочисленные ошибки учеников и переводчиков, связанные с использованием в речи многозначных слов, свидетельствуют о необходимости серьезной и систематической работы над многозначностью и над значением слова вообще.

Выберем параллельный корпус, немецкий язык.

Для примера остановимся на словах «кран», «звезда», «язык».

Ищем контексты:

1) Das Studium der russischen Sprache führte den jungen Studenten in die russische Emigrantengemeinde in Frankfurt. [Der Bote, No. 1 (2011)] – Изучение русского языка приблизило молодого студента к эмигрантской русской общине во Франкфурте. [Вестник Германской епархии РПЦЗ, № 1 [ABBYY LingvoPRO] (2011)]

2) Die Bauern hörten mit einer unverschämten Offenheit zu, mimten Sich-Erhängen, als er von den erhängten Bauern sprach, hoben die rechte Hand über den Kopf, als zögen sie an einem Strick, verdrehten die Augen, streckten ihre Zungen heraus. [Friedrich Dürrenmatt. Justiz (1985)] – Крестьяне с наглой откровенностью слушали наш разговор, мимически изображали удушение, когда он говорил о повесившихся, задирали правую руку над головой, словно затягивая веревку, закатывали глаза, высовывали язык. [Фридрих Дюрренматт. Правосудие (С. Фридлянд, 1988)]

3) Dicht über dem Wasser funkelte etwas in der Lichtsäule wie ein heller Stern. [Michael Ende. Momo (1973)] – Над самой водой в луче что-то мерцало, как ясная звезда. [Михаэль Энде. Момо (Ю. И. Коринец, 1982)]

4) «Sie liegt mit einem Super-Rechtsanwalt im Bett «, erläuterte ich,» mit dem Star aller juristischen Stare, mit einem ausgebildeten Galgenvogel, mit dem großen Allerweltsadvokaten Stüssi-Leupin.» [Friedrich Dürrenmatt. Justiz (1985)] – Она связалась с одним суперадвокатом, — объяснил я, — с главным светилом из всех юридических светил, с хорошо обученным висельником, с великим адвокатом на все руки по имени Штюсси-Лойпин. [Фридрих Дюрренматт. Правосудие (С. Фридлянд, 1988)]

5) Rasch öffnete die Alte den am Kessel befindlichen Hahn, und glühendes Metall strömte zischend und prasselnd in eine kleine Form, die sie danebengestellt. [Ernst Theodor Amadeus Hoffmann. Der goldne Topf (1814)] – Старуха быстро отворила кран у котла, и расплавленный металл, шипя и треща, потек в маленькую форму, которую она подставила. [Эрнст Теодор Амадей Гофман. Золотой горшок (В. Соловьев, 1880)]

6) Die Ausdünstungen von Wasser, Kohlen und Teer, die scharfen Gerüche gehäufte Kolonialwaren in der Nase, sah er an den Hafenkais ungeheure Dampfdrehkrane die Ruhe, Intelligenz und Riesenkraft dienender Elefanten nachahmen, indem sie Tonnengewichte von Säcken, Ballen, Kisten, Fässern und Ballons aus den Bäuchen ruhender Seeschiffe in Eisenbahnwagen und Schuppen löschten. [Thomas Mann. Der Zauberberg (1924)] – Он привык к испарениям воды, угля, смолы, к пряным ароматам разнообразнейших колониальных товаров, он видел, как в гавани огромные подъемные краны неторопливо, умно и с чудовищной мощностью, словно подражая рабочим слонам, извлекают тонны грузов — мешки, тюки и ящики, бочки и баллоны — из чрева стоящих на якоре морских судов и переносят их в товарные вагоны и на склады. [Томас Манн. Волшебная гора (В. Станевич, 1959)]

Упрощаем контексты:

1) Das Studium der russischen Sprache führte den jungen Studenten in die russische Emigrantengemeinde in Frankfurt.

2) Die Bauern hörten mit einer unverschämten Offenheit zu, verdrehten die Augen, streckten ihre Zungen heraus.

3) Dicht über dem Wasser funkelte etwas in der Lichtsäule wie ein heller Stern.

4) Sie liegt mit einem Super-Rechtsanwalt im Bett, mit dem Star aller juristischen Stare, mit dem großen Allerweltsadvokaten.»

5) Rasch öffnete die Alte den Hahn, und glühendes Metall strömte zischend und prasselnd in eine kleine Form.

6) Er sah an den Hafenkais ungeheure Dampfdruckkrane die Ruhe, Intelligenz und Riesenkraft dienender Elefanten nachahmen.

Вставьте правильное слово в контекст. Придумайте загадки с использованием многозначных слов в немецком языке. Можно воспользоваться данными НКРЯ.

1) Das Studium der russischen (Sprache\Zunge) führte den jungen Studenten in die russische Emigrantengemeinde in Frankfurt.

2) Die Bauern hörten mit einer unverschämten Offenheit zu, verdrehten die Augen, streckten ihre (Sprachen\Zungen) heraus.

3) Dicht über dem Wasser funkelte etwas in der Lichtsäule wie ein heller (Stern\ Star).

4) Sie liegt mit einem Super-Rechtsanwalt im Bett, mit dem Star aller juristischen (Sterne\ Stare), mit dem großen Allerweltsadvokaten.»

5) Rasch öffnete die Alte den (Hahn\ Kran), und glühendes Metall strömte zischend und prasselnd in eine kleine Form.

6) Er sah an den Hafenkais ungeheure (Hähne\ Krane) die Ruhe, Intelligenz und Riesenkraft dienender Elefanten nachahmen.

Пример загадки на многозначность слов:

Welcher Strauß kann nicht laufen? (ein Blumenstrauß) – Strauß – это «страус» и «букет».

Упражнение 4

Важная роль в изучении иностранных языков отводится овладению фразеологическими единицами и поговорками, которые могут вызывать трудности при переводе. Устойчивые сочетания и пословицы должны переводиться эквивалентом, а не буквально. В противном случае их смысл может быть полностью искажен. Для работы с пословицами можно предложить следующее упражнение, разрабатываемое на основе корпусного материала.

Остановимся, к примеру, на следующих пословицах:

- Klein aber fein.
- Heute rot morgen tot.
- Alte Liebe rostet nicht.
- Jedem Tierchen sein Pläsierchen.

- Wer rastet, der rostet.
- Man soll den Tag nicht vor dem Abend loben.

Выбираем в корпусе DWDS контексты с данными пословицами:

– Überhaupt funktioniert das Kummerkastenfernsehen nach der Devise "Jedem Tierchen sein Pläsierchen" und ist insofern eine Schule der Toleranz.

– Der Tagesspiegel vom 31. März 2000 allgemein/ Betrachtungen zum Aktienboom Alles nur Theater - oder wie funktioniert die Börse ? Wer rastet, der rostet – und kriegt vom Aktienkuchen kein Stück mehr ab.

– »Alte Liebe rostet nicht«, sagt sie, »aber schimmelig kann sie werden.«

– So ist das ja nun einmal in der Anarchie: daß sie nicht nur von den Zwängen verkrusteter Gesellschaftsordnungen befreit (dann wäre es für einen liberalen Intellektuellen leicht, sie zu lieben), sondern daß sie auch dem gegenseitigen Abschießen, einst Freund nun Feind, heute rot morgen tot, nichts entgegenzusetzen hat.

– „Klein aber fein“ könnte die Devise jenes knappen halben Dutzend britischer Automobilproduzenten lauten, deren Produktionszahlen sich zwar neben denjenigen des großen Brudes British Leyland (BL) mehr als bescheiden ausnehmen, deren Modelle jedoch ebensoviel zum guten Ruf der britischen Automobilindustrie beitragen wie etwa der Dauerrenner „Mini“ aus dem Hause BL.

– Unverhofft kommt oft, man soll den Tag nicht vor dem Abend loben, denn meistens kommt es anders, als man denkt, und: Lieber den Spatz in der Hand als die Taube auf dem Dach.

Убираем из выбранных контекстов данные пословицы и формулируем задание: вставить подходящие по смыслу пословицы в предложенные контексты. Подобрать эквиваленты на русском языке.

1) Überhaupt funktioniert das Kummerkastenfernsehen nach der Devise " _____ " und ist insofern eine Schule der Toleranz.

2) Der Tagesspiegel vom 31. März 2000 allgemein/ Betrachtungen zum Aktienboom Alles nur Theater – oder wie funktioniert die Börse ? _____ – und kriegt vom Aktienkuchen kein Stück mehr ab.

3) » _____ «, sagt sie, »aber schimmelig kann sie werden.«

4) So ist das ja nun einmal in der Anarchie: daß sie nicht nur von den Zwängen verkrusteter Gesellschaftsordnungen befreit (dann wäre es für einen liberalen Intellektuellen leicht, sie zu lieben), sondern daß sie auch dem gegenseitigen Abschießen, einst Freund nun Feind, _____, nichts entgegenzusetzen hat.

5) „ _____ “ könnte die Devise jenes knappen halben Dutzend britischer Automobilproduzenten lauten, deren Produktionszahlen sich zwar neben denjenigen des großen Brudes British Leyland (BL) mehr als bescheiden ausnehmen, deren Modelle jedoch ebensoviel zum guten Ruf der britischen Automobilindustrie beitragen wie etwa der Dauerrenner „Mini“ aus dem Hause BL.

6) Unverhofft kommt oft, _____, denn meistens kommt es anders, als man denkt, und: Lieber den Spatz in der Hand als die Taube auf dem Dach.

Упражнение 5

Fragestellung: *Wie erstellt man eine Übung zu Richtungsangaben?*

Gesucht wird nach dem Prädikatsverb *stellen* mit der Richtungsangabe *auf* (weiter auch *in, unter, über, hinter*): Satzmuster: *stellen etw. auf (in/unter) etw.*

Recherchiert wird im **InterCorp**, da dieses auch den Vergleich mit anderen Sprachen ermöglicht. Darüber hinaus ist die Abfrage im InterCorp einfacher als im DeReKo, obwohl es auf den ersten Blick vielleicht nicht so aussieht. Für diejenigen, die sich mit der Syntax der Abfrage vertraut gemacht haben, ist das Generieren der Abfrage simpel. Ihre Bedeutung wird anschließend erklärt.

- **korpus.cz** → Login → KonText → Parallel corpus InterCorp → intercorp_de (ggfs. noch (eine) andere Parallelsprache(n) dazu)

- Query Type: CQL auswählen

- CQL (Suchfeldeingabe): **[lemma="stellen"] []{0,2} [word="auf"] []{0,2} [tag="NN"]**
→ Search etc.

Zu jeder Abfrage bekommt man mehrere tausend Konkordanzzeilen. Es ist empfehlenswert, diese zu mischen (auf dem waagrechten Menübalken: Concordance → Shuffle), dadurch bekommt man Belege aus unterschiedlichen Quellen (siehe in der Abb. 90 die Spalte Quelle). Dann müssen nur diejenigen ausgewählt werden, die wirklich eine Lokalpräposition darstellen. Die meisten Belege repräsentieren nämlich Phraseme: *unter Verdacht stellen, unter (...) Willen stellen, in Frage stellen, in Rechnung stellen* und viele andere mehr.

Gut gewählte Konkordanzen können als Übungen zum Trainieren von gleich mehreren grammatikalischen Aspekten verwendet werden:

Ergänzungen von 1) Artikel; 2) (Personal-)Pronomen; 3) Präpositionen, wie die folgenden Beispiele zeigen:

1) Artikelergänzung:

Ergänzen Sie den passenden Artikel:

Ich packte die Wiege und **stellte** sie (...) **neben** ___ Kopfende meines Bettes. (das)

Die Herren **stellen** sich **in** ___ Ecke und politisieren. (eine)

Kalimat **stellte** sich ihnen **in** ___ Weg. (den)

Spann deinen Schirm auf und **stell** ihn **in** ___ Ausguß, da kann er trocknen. (den)

2) Ergänzung von Pronomen

Ergänzen Sie das passende Personal- oder Possessivpronomen:

Eines Abends **stellte** er sich **unter** ___ Fenster, als sie gerade mit ein paar Freundinnen Tee trank. (ihr)

Stellen Sie sich **hinter** ____, Rücken an Rücken! (mich, ihn, sie)

3) Ergänzung von Präpositionen:

Ergänzen Sie die passende Präposition:

Ich packte die Wiege und **stellte** sie (...) _____ **das Kopfende** meines Bettes. (neben, an, (auf))

Die Herren **stellen** sich _____ **eine Ecke** und politisieren. (in)

Kalimat **stellte** sich ihnen _____ **den Weg**. (in)

Spann deinen Schirm auf und **stell** ihn _____ **den Ausguß**, da kann er trocknen. (in)

Eines Abends **stellte** er sich _____ **ihr Fenster**, als sie gerade mit ein paar Freundinnen Tee trank. (unter)

Stellen Sie sich _____ **mich**, Rücken an Rücken! (hinter)

Die Leerstellen können im elektronischen Übungsportal durch ein Eingabefeld oder Auswahlfensterchen ersetzt werden.

Stellen Sie sich _____ mich, Rücken an Rücken!

Stellen Sie sich **hinter** mich, Rücken an Rücken!

zu
bei

Покажите, каким образом на базе НКРЯ можно составить 3 упражнения для освоения лексики определенного типа. Ваш ответ должен соответствовать следующей схеме:

1. Обозначение и обоснование выбора типа лексики.
2. Описание подкорпуса, в котором будет вестись поиск контекстов.
3. Описание типа упражнения (3 упражнения должны относиться к разным типам).
4. Демонстрация контекстов (не менее 5 для каждого упражнения) до и после редактуры.
5. Итоговый результат: текст задания с контекстами.

Ссылки на интернет-ресурсы, упомянутые в лекции:

Online concordancers: lextutor.ca/conc

Русский учебный корпус: web-corpora.net/RLC

Корпус русских учебных текстов: web-corpora.net/learner_corpus

Russian Error-Annotated Learner English Corpus: realec.org

Словарь глагольной сочетаемости непредметных имен русского языка: dict.ruslang.ru/abstr_noun.php

Литература:

Olga Dontcheva-Navratilova. Lexical Bundles in Academic Texts by Non-native Speakers.

Тема 10. Корпус и междисциплинарные исследования

План

- 1) Корпусные методы в образовании
- 2) Корпусные методы в изучении литературы
- 3) Корпусные методы и юриспруденция
- 4) Корпусные методы и межкультурная коммуникация
- 5) Заключение: перспективы корпусных исследований

Просмотрите видеолекции.

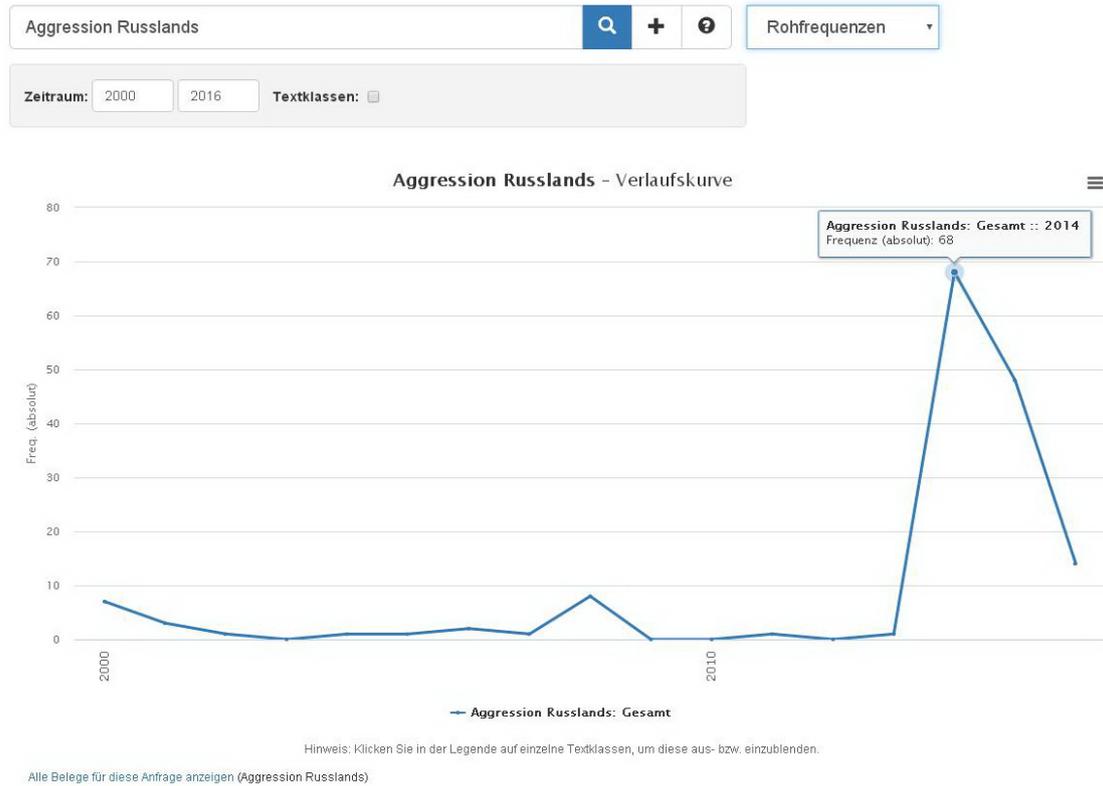
Задания для самостоятельной работы

1. Каким образом в играх, представленных на обучающем сайте «Верные слова», реализуется идеология корпусного преподавания?
2. Прочитайте статью [Ф. Моретти «Операционализация», или функция измерений в современной теории литературы»](#) и определите, какие параметры художественного произведения, с точки зрения автора, поддаются измерению.
3. Какие направления развития корпусной лингвистики в будущем кажутся вам наиболее перспективными?

Практические задания

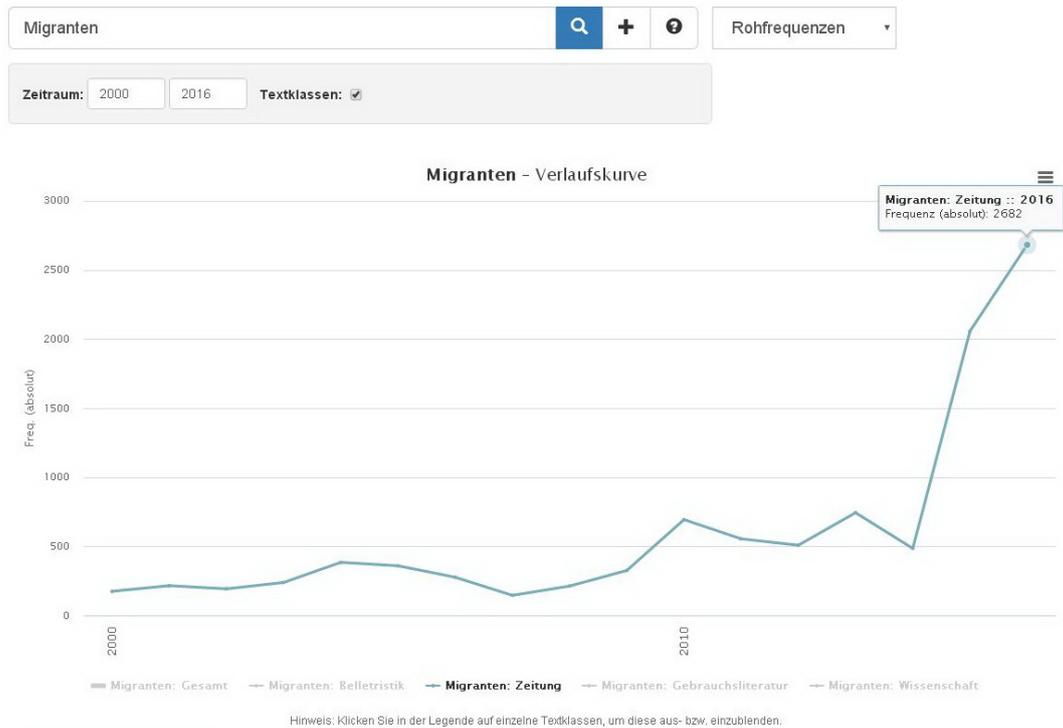
1. Рассмотрите кривую вхождений сочетаний “Aggression Russlands” за период с 2000 по 2016 год из цифровой системы DWDS. Подумайте, какие выводы можно сделать по этим данным. Как и почему изменяется частотность употребления данного словосочетания? Вспомните историю отношений между двумя государствами. С какими событиями могут быть связаны эти изменения? Где могут быть использованы данные выводы?

DWDS – Verlaufskurve



2. Проанализируйте схему. Как и почему изменяется частотность использования слова Migration в немецкой прессе? Как можно использовать данные выводы?

DWDS – Verlaufskurve



3. Подумайте, какие исследования можно было провести на основе кривых частотности употребления слов. Приведите примеры.

4. Проанализируйте наиболее типичные сочетания со словом *Russland*, сгенерированные компьютером. Какие можно сделать выводы на основе этих сочетаний? Где можно применить эти знания?

Typische Verbindungen

computergeneriert



Detailliertere Informationen bietet das [DWDS-Wortprofil zu »Russland«](#).

5. Как можно использовать данные, представленные в упражнении 4, в когнитивных исследованиях? Приведите свои примеры.

Ссылки на интернет-ресурсы, упомянутые в лекции:

Верные слова: vsrussian.com

Cambridge English Corpus: www.cambridge.org/gb/cambridgeenglish/better-learning/deeper-insights/linguistics-pedagogy/cambridge-english-corpus

Проект RusVectōrēs: rusvectores.org/ru

Литература:

Christiane Meierkord, Bebwa Isingoma, Saudah Namyalo. Ugandan English: Its Sociolinguistics, Structure and Uses in a Globalising Post-Protectorate. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2016.

Франко Моретти. Дальнее чтение / пер. с англ. А. Вдовина, О. Собчука, А. Шели. Науч. ред. перевода И. Кушнарера. – М.: Изд-во Института Гайдара, 2016.

Франко Моретти. «Операционализация», или Функция измерений в современной теории литературы.

О. Собчук, А. Шеля. Что такое «дальнее чтение» Франко Моретти? Из интервью Rara Avis.

Patrick Juola. How a Computer Program Helped Show J.K. Rowling write A Cuckoo's Calling.

Проанализируйте представленный в открытом доступе контент образовательного сайта («Верные слова», «Живые страницы», сайт Cambridge English Corpus или любого другого на ваш выбор), использующего корпусные технологии. Покажите, какие возможности предоставляет корпусное обучение преподавателям гуманитарных дисциплин (опишите не менее трёх возможностей).

ОЦЕНОЧНЫЕ СРЕДСТВА

Итоговый тест к видеолекциям

Выберите один или несколько верных вариантов ответа.

Будьте внимательны! В некоторых случаях правильными могут оказаться все предложенные варианты.

1. Цифровая гуманитаристика ставит перед собой задачи

- a) перевести в электронный формат значимые для культуры тексты;
- b) продолжать и совершенствовать филологическую практику медленного чтения;
- c) провести количественный анализ художественных произведений;
- d) популяризовать культурное наследие через визуализацию и создание электронных продуктов.

2. Проект Tolstoy Digital относится к области цифровой гуманитаристики, поскольку

- a) предполагает издание полного 90-томного собрания сочинений писателя;
- b) дает читателю возможность познакомиться не только с самим текстом, но и с комментариями к нему;
- c) включает семантическую разметку, позволяющую связать друг с другом разные элементы текста и метатекста;
- d) является основой для создания инфографических и мультимедийных продуктов.

3. Референтные корпуса используют для того, чтобы с помощью статистических методов в художественном произведении проанализировать

- a) объем отданного персонажу пространства;
- b) количество и силу связей между персонажами;
- c) ключевые слова произведения;
- d) семантическую просодию.

4. Google Translator Toolkit, как и другие CAT, позволяет переводчику

- a) получать мгновенный автоматический перевод всего текста, не нуждающийся в редакторской правке;
- b) пользуясь базой выполненных ранее переводов, выделять в тексте термины и специальную лексику;
- c) находить в тексте оригинала обороты, являющиеся средствами художественной выразительности;
- d) эффективно работать над переводом в сотрудничестве с коллегами.

5. Национальный корпус русского языка может помочь переводчикам

- a) найти возможные переводы лексемы на иностранный язык в параллельных корпусах;
- b) определив время появления лексемы в русском языке, избежать анахронизмов в тексте русского перевода;
- c) сделать выводы о верном произношении иностранного слова;
- d) создать краудсорсинговый переводческий проект.

6. Лингвистический корпус можно определить как

- a) электронную библиотеку, созданную в соответствии с запросами читателя;
- b) коллекцию естественных текстов, отражающих состояние языка в данную эпоху;
- c) сбалансированное, репрезентативное и размеченное собрание текстов;
- d) собрание текстов, организованное для удобства лингвистического исследования.

7. К национальным корпусам относятся

- a) НКРЯ;
- b) BNC;
- c) COCA;
- d) TEI.

8. Национальные корпуса

- a) позволяют исследовать язык нации;
- b) включают только культурно значимые тексты;
- c) не включают тексты, чьи авторы отступают от литературной нормы;
- d) могут включать тексты современных жанров электронной коммуникации.

9. Авторы НКРЯ организовали в составе национального корпуса особый обучающий подкорпус, поскольку

- a) школьная морфологическая разметка отличается от принятой в НКРЯ;
- b) в основном корпусе НКРЯ присутствует ненормативная лексика;
- c) поиск в НКРЯ в ряде случаев выдает омонимичные формы;
- d) орфография в НКРЯ не нормализована.

10. Корпусами русского языка являются

- a) Упсальский корпус;
- b) Тюбингенский корпус;
- c) корпус университета Браун;

d) корпус университета Хельсинки.

11. С точки зрения большинства исследователей, для любого лингвистического корпуса важны

- a) присутствие культурно значимых текстов;
- b) достаточный объем;
- c) наличие подкорпуса звучащей речи;
- d) аннотация.

12. Неприемлемыми для большинства создателей корпусов являются

- a) включение в корпус текстов, искусственно сконструированных его создателями;
- b) отсутствие описания текстов;
- c) единообразие разметки;
- d) создание коллекции текстов для решения конкретной лингвистической задачи.

13. Лингвистический корпус не может заменить словари, поскольку

- a) не дает представления о морфологических и семантических признаках лексемы;
- b) показывает узус, а не норму;
- c) фиксирует употребление, но не анализирует материал;
- d) демонстрирует небольшое, по сравнению со словарем, количество контекстов.

14. Конкорданс

- a) в момент своего появления служил одним из методов интерпретации священных текстов;
- b) является предшественником современных корпусных инструментов;
- c) в современных корпусах реализован форматом KWIC;
- d) прекратил существование в середине XX века.

15. Корпусные лингвисты

- a) признают главенство интуиции и интроспекции;
- b) внимательны к вариативности;
- c) толерантны к норме;
- d) не исследуют формы, не описанные в стандартных грамматиках.

16. Корпуса устной речи, в отличие от корпусов письменной речи, могут включать разметку

- a) семантики слова;
- b) семантики жестов;

- c) речевых сбоев;
- d) интонации.

17. Параллельный корпус

- a) включает тексты оригинала и перевода;
- b) состоит из выровненных фрагментов;
- c) является эффективным инструментом исследования лингвоспецифичной лексики;
- d) позволяет эффективно изучать синонимы.

18. Проект ICE

- a) представляет сравнимые корпуса английского языка;
- b) позволяет исследовать особенности английского языка в странах, где английский является вторым официальным языком;
- c) позволяет исследовать английский язык студентов, изучающих его как неродной;
- d) дает представление не только о письменной, но и об устной речи.

19. Лингвистические корпуса могут быть

- a) диахроническими;
- b) историческими;
- c) многоязычными;
- d) одноязычными.

20. Авторы корпусных грамматик

- a) отказываются от искусственных примеров в пользу контекстов, взятых из реального словоупотребления;
- b) как правило, подбирают примеры в текстах художественной литературы;
- c) описывают как качественные, так и количественные параметры явления;
- d) стремятся составить исчерпывающие списки исключений.

21. Принципы аннотирования корпуса

- a) не важны для пользователя;
- b) должны быть максимально теоретически нейтральны;
- c) демонстрируют себя в наборе тегов;
- d) могут варьироваться в рамках одного корпуса.

22. Автоматическая разметка включает

- a) выделение начальной формы;
- b) приписывание словоформе частеречных признаков;
- c) определение семантики лексемы;
- d) полное снятие омонимии.

23. К метатекстовой разметке относят

- a) сведения об авторе текста;
- b) сведения о тематике текста;
- c) данные о периоде создания текста;
- d) данные о синтаксических признаках сочетаний.

24. Синтаксическая разметка в НКРЯ

- a) проводится в полуавтоматическом режиме;
- b) охватывает все, включенные в корпус тексты;
- c) определяет, каким членом предложения является словоформа;
- d) графически представлена в виде дерева зависимостей.

25. Семантическая разметка в НКРЯ

- a) не предполагает отсутствия омонимии в выдаче примеров;
- b) позволяет найти лексемы с оценочным компонентом значения;
- c) позволяет найти предметные и неперекрестные имена;
- d) проведена полностью вручную.

26. Ключевые слова

- a) вычисляются с помощью референтного корпуса;
- b) это лексемы, чья частотность в тексте превышает обычную частотность в языке;
- c) представляют собой устойчивые сочетания;
- d) полезны при анализе художественного текста.

27. Словари коллокаций

- a) не могут быть представлены в бумажном виде;
- b) составляются с помощью статистических методов;
- c) показывают «силу связи» слов в сочетании;
- d) полезны при обучении иностранному языку.

28. Семантическая просодия

- a) возникает у слова в сочетании;
- b) всегда осознается носителем языка;
- c) может быть нейтральной;
- d) чаще бывает положительной, чем отрицательной.

29. Исследования, основанные на изучении корпусов,

- a) показывают независимость языковых моделей от конкретного регистра;
- b) никогда не используют корпус как источник примеров для уже существующих научных гипотез;
- c) всегда анализируют словоформы, а не лексемы;
- d) обычно связаны с количественными методами.

30. Частотный словарь может давать неверную информацию, если
а) размер корпуса, на материале которого составлен словарь, был недостаточно велик;

- б) корпус был недостаточно сбалансирован;
- в) в корпус входили произведения художественной литературы;
- г) в корпус входили тексты политической тематики.

31. Корпусное преподавание

- а) возникло как направление в XXI веке;
- б) развивалось в России, затем распространилось в Европе и в США;
- в) обязано своим появлением потребностям изучающих английский язык как иностранный;
- г) призвано компенсировать недостаток сведений о языке, представленных в прескриптивных грамматиках.

32. В преподавании лингвистический корпус используется как

- а) база примеров для демонстрации ученикам в компьютерном классе;
- б) инструмент для составления учебных пособий;
- в) инструмент для самостоятельной работы ученика;
- г) инструмент для анализа ошибок студентов.

33. При составлении корпусных учебников для изучающих иностранный язык авторы собирают в корпусе информацию

- а) о частотности слов и сочетаний;
- б) о моделях, характерных для конкретных регистров;
- в) о связи слов и грамматических структур;
- г) о языковой норме.

34. В качестве корпусных материалов для составления учебных пособий можно использовать

- а) «Толковый словарь русского языка» под редакцией Д. Ушакова;
- б) «Новый частотный словарь русской лексики» под редакцией О. Ляшевской и С. Шарова;
- в) Русский учебный корпус RLC;
- г) Учебный корпус REALEC.

35. Learner corpus

- а) представляет собой коллекцию текстов, созданных обучающимися;
- б) обладает дополнительной разметкой по сравнению с традиционными корпусами;
- в) обычно дает возможность найти контексты, демонстрирующие конкретный тип ошибки;
- г) должен включать тексты, отредактированные преподавателем.

36. В ходе овладения формульной речью ученики совершают следующие ошибки:

- a) неверно истолковывают значение сочетания;
- b) заменяют одну из входящих в сочетание лексем;
- c) используют формулы в неподходящем регистре;
- d) используют формулы реже, чем этого требует регистр.

37. Data-driven learning

- a) разработан Т. Джонсом;
- b) предлагает ученику самому распознавать языковые тенденции;
- c) представляет собой индуктивный подход к обучению;
- d) критикуется как чрезмерно легкий и не требующий аналитических усилий от ученика.

38. Достоинства data-driven learning включают

- a) повышение мотивации студента, склонного к самостоятельной работе;
- b) трудоемкость и временную затратность анализа большого количества найденных в корпусе контекстов;
- c) отказ от работы с готовыми формулировками в пользу самостоятельного формулирования правил;
- d) возможность развить у ученика навыки научного исследования.

39. Создание корпусного электронного стилистического тренажера с помощью метода «стилистического эксперимента» предполагает

- a) создание классификации типичных стилистических ошибок;
- b) подбор в корпусе контекстов, в которых возможна ошибка, представленная в классификации;
- c) «ухудшение» контекстов в соответствии с моделью ошибки;
- d) отсутствие какой бы то ни было редактуры найденных в корпусе контекстов.

40. К несомненным достоинствам корпусных методов стоит отнести

- a) зависимость качества исследования от качества программного обеспечения;
- b) возможность проверки гипотезы на обширных примерах;
- c) относительную объективность исследования, обусловленную использованием статистических методов;
- d) возможность провести ревизию классических теорий на базе нового материала.

Тест для самоконтроля

1. Какое определение корпуса является наиболее точным?

а) информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме.

б) репрезентативное собрание текстов, обычно в читаемом машиной формате и включающем информацию о ситуации, в которой текст был произведен, такую как информация о говорящем, авторе, адресате или аудитории.

в) своеобразного рода словесное единство, расположенное на машинном носителе, стандартизовано представляющее словесный материал, позволяющее применять стандартные программы его обработки, характеризующееся конечным размером и репрезентативностью как результатом особой процедуры отбора.

2. Что представляет собой корпусная лингвистика?

а) отрасль компьютерной лингвистики, занимающаяся составлением корпусов текстов для лексикографических целей

б) деятельность, требующаяся для составления и использования корпуса, направленная на исследование естественного использования языка

а. вид деятельности, связанной с использованием текстов для того, чтобы ответить на вопросы о использовании языка

3. Какой корпус содержит только устную речь?

а) London-Lund Corpus

б) Longman-Lancaster Corpus

в) British National Corpus

4. Какой минимальный объем считается репрезентативным для национального корпуса языка?

а) 160 млн. словоупотреблений

б) 1 млн. словоупотреблений

в) 100 млн. словоупотреблений

5. Что обозначает термин «коллокация» в корпусной лингвистике?

а) любое сочетание слов

б) устойчивое сочетание слов

в) сочетание слов, расположенных рядом в тексте

6. Области лингвистических исследований, повлиявшие на развитие современной корпусной лингвистики

а) историческая лингвистика, грамматика, социолингвистика

б) когнитивная лингвистика, лексикография, социолингвистика

в) фонология, лексикография, лингводидактика

- 7. Какие данные можно получить из корпусов текстов?**
- а) суждения о грамматике, экстралингвистическая информация
 - б) экспериментальные данные, данные полевых исследований
 - в) эмпирическая поддержка, информация по частотности, метаинформация (экстралингвистическая информация)
- 8. Что такое конкорданс?**
- а) указатель, связывающий каждое употребление с контекстом
 - б) список встречающихся в тексте словоформ, расположенных в алфавитном порядке
 - в) тип словаря
- 9. Что такое token?**
- а) словоформа
 - б) начальная форма слова
 - в) основа слова
- 10. Какой процесс устанавливает лексему, парадигма которой включает данную словоформу?**
- а) стемминг
 - б) токенизация
 - в) лемматизация
- 11. Для чего нужны нормированные подсчеты?**
- а) они указывают частотность встречаемости определенной словоформы
 - б) они преобразуют количество случаев встречаемости слова по стандартной шкале
 - в) они применяются в лексикографии для выявления новых значений слов
- 12. Чтобы слова можно было ассоциировать друг с другом, они должны:**
- а) примыкать друг к другу или располагаться на некотором расстоянии
 - б) располагаться на некотором расстоянии
 - в) примыкать друг к другу
- 13. В каком из регистров чаще встречаются номинализации с суффиксом –ness?**
- а) в академической прозе
 - б) в художественной литературе
 - в) в устной речи

14. Для каких исследований можно использовать неаннотированные корпуса?

- а) в области морфологии
- б) в области грамматических категорий
- в) в области синтаксиса

15. Какой из регистров демонстрирует предпочтение к статической, а не к динамической упаковке информации, отраженной в большей частоте номинализаций и более высоком соотношении существительное-глагол?

- а) устная речь
- б) художественная литература
- в) академическая проза

ГЛОССАРИЙ

-А-

Аннотация – это приписанная всем единицам выбранного уровня (текст, предложение, словоформа и т. д.) соответствующая лингвистическая информация.

-Б-

n-граммы – устойчивые словосочетания двух, трех и т.д. слов, существующие в языке. На основе анализа частоты употребления n-грамм можно рассчитать вероятность появления того или иного предложения, составленного из n-грамм, в реальном человеческом языке.

Биграмма — это два слова, которые в тексте или в корпусе текстов являются соседними.

Британский национальный корпус (BNC от англ. British National Corpus) – это корпус текстов из 100 миллионов слов, содержащий образцы письменного и разговорного британского английского языка из широкого круга источников. Корпус охватывает британский английский конца XX в., представленный широким разнообразием жанров, и задуман как образец типичного разговорного и письменного британского английского языка того времени.

-В-

Выравнивание параллельного текста – это идентификация соответствующих друг другу предложений в обеих половинах параллельного текста.

-Д-

Дезамбигуация – это разрешение лексической многозначности, способ выбора точного значения слова из имеющихся омонимов с учетом окружающего контекста.

-К-

Ключевое слово (англ. key word in context - KWIC) – слово, заложенное в программу для поиска, результат поиска выстраивается в виде конкорданса.

Коллигация – совокупность морфолого-синтаксических условий, обеспечивающих сочетаемость языковых единиц.

Коллокации (collocates) – слова, с которыми анализируемое слово (часто) встречается вместе.

Коллокация – последовательность слов или терминов, частотность совместного появления которых в корпусе выше, чем ожидаемая вероятность их совместного появления.

Конкорданс – алфавитный перечень всех слов какого-либо текста с указанием контекстов их употребления; тип словаря, представленный в таком виде.

Корпус – информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме.

Корпус устной речи – корпус, в котором собраны образцы устной речи, например, в виде расшифровки магнитофонных записей публичной и частной устной речи, а также транскрипты кинофильмов.

Корпус-менеджер – программное обеспечение для корпуса, которое подсчитывает конкретные словоформы, группы словосочетаний и также может выводить результаты статистики, представить в виде диаграммы.

Корпусная лингвистика – раздел языкознания, занимающийся разработкой, созданием и использованием текстовых корпусов. Термин введён в употребление в 1960-е годы в связи с развитием практики создания корпусов, которому начиная с 1980-х способствовало развитие вычислительной техники.

-Л-

Лемма – нормализованная, основная форма слова, вместе с информацией о построении других форм.

Лемматизация – процесс приведения словоформы к лемме – её нормальной (словарной) форме.

Лингвистическая разметка (англ. tagging, annotation) – это процесс или результат приписывания текстам и их компонентам специальных меток. Разметка даёт возможность идентифицировать тексты по различным параметрам, позволяя осуществлять осмысленный поиск по корпусу. Собственно лингвистическая разметка делится на: морфологическую (выделение аффиксов, сложных слов и т.п.), лемматизацию (указание для каждой словоформы из текста ее исходной формы), морфо-синтаксическую, или частеречную (part-of-speech-tagging) (выделение основ, определение части речи и признаков грамматических категорий), синтаксическую (характер синтаксической связи, тип предложения, член предложения и т.п.), семантическую (снятие семантической омонимии, разрешение анафоры и кореферентности, фиксирование информационной структуры и т.п.), дискурсивную (реплики, коммуникативные акты и т.п.).

Лингвистический корпус - совокупность текстов, собранных в соответствии с определёнными принципами, размеченных по определённому стандарту и обеспеченных специализированной поисковой системой [источник не указан 1537 дней]. Иногда корпусом («корпус первого порядка») называют просто любое собрание текстов, объединённых каким-то общим признаком (языком, жанром, автором, периодом создания текстов).

-М-

Метаразметка – см. Экстралингвистическая разметка.

Морфологическая разметка текста – приписывание словоформам информации об исходной форме слова и о совокупности их грамматических признаков.

Мультимедийный русский корпус (МУРКО) – это подкорпус Национального корпуса русского языка, который содержит устные русские тексты, выровненные с соответствующими аудио- и видеофрагментами. МУРКО позволяет изучать устную речь, используя не только ее письменную фиксацию, но и ее реальное звучание и жестовое сопровождение. МУРКО рассчитан, прежде всего, на фонетистов, интонологов, специалистов по жестикуляции, но будет также полезен преподавателям русского языка, прежде всего, как иностранного.

-Н-

Национальный корпус русского языка (НКРЯ) – доступный для поиска электронный онлайн-корпус русских текстов. Открыт 29 апреля 2004 года. Также доступен для поиска исторический корпус церковнославянских, древнерусских (XI—XIV века) и среднерусских (XV — начало XVIII века) текстов.

Неснятая омонимия. Часть текстов Корпуса представляют собой тексты с неснятой омонимией, то есть в них не разрешена неоднозначность форм с одинаковым написанием (например, печь — инфинитив глагола и существительное, печи — форма родительного, дательного и предложного падежа ед. ч., именительного и винительного мн. ч.). Все такие формы получают весь возможный набор разборов, таким образом, поисковая выдача по этим текстам содержит некоторое количество «шума». Возможно задать ограничение на поиск по корпусу только со снятой или только с неснятой омонимией. В поисковой выдаче рядом с названием текста всегда указывается, снята в нем омонимия или нет.

-О-

Обучающий корпус русского языка (открылся в 2007 г.) — небольшой корпус со снятой омонимией, ориентированный на преподавание русского языка в школе (отобраны произведения из школьной программы, изучаемых в школьном курсе функциональных стилей, размечены словоизменительные типы и другие дополнительные морфологические признаки)

-П-

Параллельный корпус – это двуязычный корпус, то есть текст оригинала и его перевод на какой-то другой язык, причем эти два текста не просто лежат рядом друг с другом, а должны быть выровнены: отдельные фрагменты оригинала должны совпадать с соответствующими фрагментами перевода. Именно это позволяет использовать параллельный корпус как инструмент исследования.

Парсинг – это синтаксический анализ, который автоматически производится парсером – специальной программой или скриптом, процесс сопоставления линейной последовательности лексем (слов, токенов) языка с его формальной грамматикой. Результатом обычно является дерево зависимостей (синтаксическое дерево).

Подкорпус – корпус текстов, являющийся частью более крупного корпуса или набора корпусов.

-Р-

Разметка – приписывание текстам и их компонентам специальных тегов: собственно лингвистических, описывающих лексические, грамматические и прочие характеристики элементов текста, и внешних, экстралингвистических (сведения об авторе и сведения о тексте: автор, название, год и место издания, жанр, тематика).

Репрезентативность – свойство корпуса, заключающееся в статистически достоверном представлении языка или его части и достигаемое за счет необходимого объема и жанрового разнообразия текстов (см. сбалансированность).

-С-

Сбалансированность (англ. balance) – параметр, который определяет, насколько равномерно представлены тексты разных типов (см. репрезентативность).

Семантическая разметка: семантические тэги чаще всего обозначают семантические категории, к которым относится данное слово или

словосочетание, и более узкие подкатегории, специфицирующие его значение. Семантическая разметка корпусов предусматривает спецификацию значения слов, разрешение омонимии и синонимии, категоризацию слов (разряды), выделение тематических классов, признаков каузативности, оценочных и деривационных характеристик и т.д.

Симфония или **Конкорданс**, **конкорданция** (др.-греч. συμφωνία — «созвучие»; лат. concordia — «согласие») — книга, в которой собраны из одного или нескольких сочинений места, состоящие из одних и тех же слов (так называемая конкорданция слов) или содержащие один и тот же смысл (так называемая реальная конкорданция).

Синтаксическая разметка является результатом парсинга, выполняемого на основе данных морфологического анализа. Этот вид разметки описывает синтаксические связи между лексическими единицами и различные синтаксические конструкции (например, придаточное предложение, глагольное словосочетание и т.д.).

Стемминг — это процесс нахождения основы слова для заданного исходного слова.

-Т-

Токен (англ. token) — объект, создающийся из лексемы в процессе лексического анализа («токенизации», от англ. tokenizing). В прикладном программировании понятие токена и его лексема могут не различаться.

Токенизация (англ. tokenization) – выделение в текстовом потоке минимальных фрагментов для последующего анализа (в корпусной лингвистике их принято называть токены (англ. token)).

-Ц-

Цифровая гуманитаристика (англ. Digital Humanities) – это область исследований, обучения и созидания, созданная на стыке компьютерных и гуманитарных наук. Цифровые гуманитарные науки предполагают использование оцифрованных материалов и материалов цифрового происхождения и объединяют методологии из традиционных гуманитарных наук (история, философия, лингвистика, литература, искусство, археология, музыка и т.д.) с компьютерными науками, предоставляя компьютерные инструменты и открывая новые возможности для сбора и визуализации данных, информационного поиска, интеллектуального анализа данных, а также применения матстатистики.

-Ч-

Частиечная разметка (англ. POS tagging) – определение части речи каждого слова.

-Э-

Экстралингвистическая разметка (метаразметка) – сведения об авторе и сведения о тексте: автор, название, год и место издания, жанр, тематика.

CAT (Computer Aided Translation) – [программа](#), которая помогает переводить тексты на компьютере.

Sketch Engine – веб-система, которая позволяет лингвистам исследовать большие корпуса текстов и создавать сложные запросы, чтобы извлекать нетривиальную информацию из этих корпусов.

-С-

Corpus-Based research – метод лингвистического исследования, основанный на корпусах текстов, ориентированный на прикладное изучение языка, его функционирование в реальных средах и текстах.

Corpus-driven research – направление исследования, при котором исследователи не имеют заранее определенной установки, а наоборот, полностью полагаются на данные корпуса, из которого затем уже выводятся теории, нормы и правила.

-К-

KWIC (Key Word in Context) – ключевое слово в контексте

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Богданова С.Ю. Рабочая программа дисциплины «Корпусная лингвистика». Иркутск: Иркутский государственный университет, 2016.
2. Википедия. URL: https://ru.wikipedia.org/wiki/Заглавная_страница
3. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.
4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.
5. Иллюстрация. URL: [рембрандт ван рейн урок анатомии доктора гульпа 1632 г http://www.intelros.ru/readroom/nlo/128-2014/24750-operacionalizaciya-ili-funkciya-izmereniy-v-sovremennoy-teorii-literatury.html](http://www.intelros.ru/readroom/nlo/128-2014/24750-operacionalizaciya-ili-funkciya-izmereniy-v-sovremennoy-teorii-literatury.html)
6. Информационный ресурс: Национальный корпус русского языка www.ruscorgo.ru
7. Копотев М. Введение в корпусную лингвистику. Прага, 2014. Электронный учебник.
8. Левинзон А. И. Введение в корпусную лингвистику. URL: <https://openedu.ru/course/hse/CORPUS/>
9. Морозов Е. А., Уразаева Н. Р. Мигранты и немецкое общество на современном этапе: вектор перемен // Социодинамика. – 2017. – № 11. – С.45–55. DOI: 10.25136/2409-7144.2017.11.23282. URL: http://e-notabene.ru/pr/article_23282.html.
10. Плунгян В. А. Зачем нужен Национальный корпус русского языка. Неформальное введение // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005. С. 6 – 20.
11. Уразаева Н. Р., Морозов Е. А. Концептуализация образа России в немецких СМИ: перемены к лучшему? // Мировая политика. – 2017. – № 3. – С.141–154. DOI: 10.25136/2409-8671.2017.3.24004. URL: http://e-notabene.ru/wi/article_24004.html.

ПРИЛОЖЕНИЯ

Приложение 1. Список корпусов в интернете

1. Национальный корпус русского языка <http://ruscorpora.ru>
2. Компьютерный корпус текстов русских газет конца XX-го века <http://www.philol.msu.ru/~lex/corpus>
3. Корпус русского языка ХАНКО (Хельсинский университет) <http://www.ling.helsinki.fi/projects/hanco/>
4. Корпуса русских текстов на сайте Университета в Лидсе, Великобритания <http://corpus.leeds.ac.uk>
5. Русские корпуса Тюбингенского Университета <http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html>
6. Словарь-корпус языка А.С. Грибоедова <http://www.inforeg.ru/electron/concord/concord.htm>
7. Уппсальский корпус русских текстов <http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html>
8. Банк английского языка (Bank of English) Свободный доступ: <http://www.collins.co.uk/books.aspx?group=153>.
<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
9. Британский национальный корпус <http://www.natcorp.ox.ac.uk/> или <http://sara.natcorp.ox.ac.uk/>
10. Венгерский национальный корпус <http://corpus.nytud.hu/mnsz/>
11. Корпус испанского языка (исторический) <http://www.corpusdelespanol.org/>
12. Корпус латинских текстов «Персей» <http://www.perseus.tufts.edu>
13. Корпус современного датского языка <http://www.korpus2000.dk/>
14. Корпус современного итальянского языка CORIS/CODIS <http://www.cilta.unibo.it/ricerca.htm>
15. Корпус современного китайского языка (LIVAC Synchronous Corpus) <http://www.rcl.cityu.edu.hk/livac/>
16. Мангеймский корпус немецкого языка (Institut für Deutsche Sprache, Mannheim, Germany) <http://corpora.ids-mannheim.de/~cosmas/>
17. Национальный корпус словенского языка <http://www.fida.net/eng/>
18. Польский национальный корпус <http://korpus.ia.uni.lodz.pl/>
19. Словацкий национальный корпус <http://korpus.juls.savba.sk>
20. Хорватский национальный корпус <http://www.hnk.ffzg.hr/>

21. Чешский национальный корпус <http://ucnk.ff.cuni.cz>

22. Эстонский
<http://test.cl.ut.ee/korpused/baaskorpus/1980/index.html.en>

корпус

Приложение 2. Классификация корпусов (по Захарову)

Признак	Типы корпусов
Тип данных	Письменные Речевые Смешанные
Язык текстов	Русский Английский и т.д.
«Параллельность»	Одноязычные Двужычные Многоязычные
«Литературность», специфичность	Литературные Диалектные Разговорные Терминологические Смешанные
Жанр	Литературные Фольклорные Драматургические Публицистические
Доступность	Свободно доступные Коммерческие Закрытые
Назначение	Исследовательские Иллюстративные
Динамичность	Динамические (мониторные) Статические
Разметка	Размеченные Неразмеченные
Характер разметки	Морфологические Синтаксические Семантические Просодические и т.д.
Объем текстов	Полнотекстовые «Фрагментнотекстовые»
Хронологический аспект	Синхронические Диахронические
«Общность»	Общие Одного писателя
Структура	Центральные и архивные Ядерные и периферийные

Приложение 3. Правила пользования немецкими корпусами

Das Digitale Wörterbuch der deutschen Sprache (DWDS)

Abfragemöglichkeiten im DWDS

Der Abstand (ein Leerzeichen) ist mit einem Mittelpunkt (·) markiert.

Eingabe	Ergebnis (Beschreibung)
Arzt	<i>Arzt, Arztes, Ärzte ... (alle flektierten Formen von Arzt)</i>
@Arzt	<i>Arzt (nur diese Wortform)</i>
Arzt*	<i>Arzt, Arztbesuch, Arztberuf ... (mit Arzt- beginnende Wörter)</i>
*arzt	<i>Sportarzt, Hausarzt ... (auf -arzt endende Wörter)</i>
"gut·Arzt"	<i>guter Arzt, bessere Arzt, bester Arzt, besten Arztes, gute Ärzte ... (beide Wörter lemmatisiert)</i>
/arzt/	<i>Zahnarzt, Oberarzt, knarzt, verarzt, verharzt ... (alle Formen mit der Buchstabenkombination arzt)</i>
Sp=V*·with·/arzt/	<i>geharzt, knarzt, knarzten, verharzt, vewartzt (alle Verben mit der Buchstabenkombination arzt)</i>
"das·gute·Beispiel"	<i>das gute Beispiel, das beste Beispiel, die besseren Beispiele ... (alle Formen aller Wörter)</i>
"Kanzler·#1·Schröder"	<i>Kanzler Schröder, Kanzler Gerhard Schröder ... (Kanzler und Schröder im Abstand von höchstens einem Wort)</i>
"Kanzler·#3·Schröder"	<i>Kanzler aller Moleküle Gerhard Schröder will Innovationen fördern; Kanzler und begnadeten Wahlmatador Schröder ... (Kanzler und Schröder im Abstand von höchstens drei Wörtern)</i>
Kanzler· ·Schröder	<i>Alle Sätze, in denen entweder das Wort Kanzler oder Schröder oder beide Wörter vorkommen.</i>
Kanzler·&&!Schröder	<i>Alle Sätze, in denen Kanzler,</i>

\$p=NE·with·Herzog	<i>aber nicht Schröder vorkommen. Roman Herzog, Peter Herzog ... (Eigennamen15 Herzog)</i>
"Einfluss·#2·\$p=NE"	<i>Einfluß Smetanas, Einfluss in der EU ... (Einfluss(β) im Abstand von max. 2 Positionen von einem Eigennamen15)</i>
\$p=NN·with·*lein	<i>Häuflein, Kräutlein, Büchlein ... (Substantive15, die auf -lein enden. Ausgeschlossen.)</i>
\$p=ADJA·with·*sam	<i>schweigsam, ehrsam, wundersam, achtsam ... (Adjektive15 auf -sam)</i>
\$p=ADJA·with·mini*	<i>minimal, minimalistisch, miniaturhaft ... (Adjektive15 auf mini-)</i>
"gehen·#5·aus·with·\$p=PTKV Z"	<i>Das Licht ging aus (...), Mir ging die Luft aus. Wir gehen also nicht von der Sprechhandlung aus (...) ... (Verb15 ausgehen mit abgetrenntem aus, maximaler Abstand 5 Positionen)</i>
"sein·with·\$p=VVFİN·#20·\$p=VVPP·#0·@worden"	
<i>... im September 1980 waren Teile des Streckennetzes stillgelegt worden; das anzuerkennen sei auf Grund der längst überholten These von der Einheit eines Volkes unmöglich gemacht worden; jedoch waren sie gleichzeitig darauf aufmerksam gemacht worden</i>	
<i>(Verb sein als finites Verb15, max. 20 Positionen davon Partizipium eines Vollverbs gefolgt von worden)</i>	
<i>\,</i>	<i>, (Suche nach Komma. Satzzeichen wie ",..?!; werden mit Backslash maskiert!)</i>

Weitere Abfragemöglichkeiten sind unter DWDS → [Hilfe zur Suche](#) zu finden.
Das Korpus ist mit dem Stuttgarter Tagset (STTS) morphologisch getagget und lemmatisiert. Dasselbe Tagset findet man in den meisten Korpora der deutschen Sprache (DeReKo, deutsche Parallele des InterCorp, deWaC).

Wortschatz Leipzig

Abfragemöglichkeiten im Wortschatz – Leipzig

Der Abstand (ein Leerzeichen) ist mit einem Mittelpunkt (·) markiert.

Eingabe

Arzt

Arzt*

***arzt**

böhmisch*·Dorf*

Ergebnis (Beschreibung)

Arzt (nur diese Wortform)

Arzt, Arztabrechnung, Arzttakten, Arztangebot ... (automatische Ordnung nach Formen, alphabetisch absteigend)

Notarzt, Chefarzt, Hausarzt, Zahnarzt ... (automatische Ordnung nach Frequenz, absteigend)

böhmisches Dorf, böhmische Dörfer

Das Deutsche Referenzkorpus (DeReKo)

Das Deutsche Referenzkorpus ist auch unter dem Namen **IDS-Korpus** oder **Cosmas II** bekannt. Die Bezeichnungen sind jedoch nicht ganz richtig: das Institut für Deutsche Sprache in Mannheim entwickelt und verwaltet mehrere Korpora (z.B. auch die DGD), Cosmas II hingegen ist ein vom IDS entwickelter Korpusmanager (Corpus Search, Management and Analysis System, Version II). Unter diesem Manager laufen alle IDS-Korpora der geschriebenen Sprache.

Das DeReKo ist momentan das größte Korpus für Deutsch weltweit, es enthält über 6 Milliarden Textwörter (Stand 2013) und wird sukzessive und regelmäßig erweitert. Die Bedienung aller IDS-Korpora ist ziemlich komplex. Für jede/-n, der/die sich mit der deutschen Sprache befasst, ist es dennoch, ratsam zumindest die Basisfunktionen von Cosmas II (Korpusmanager) sicher zu beherrschen. Es gibt leicht zugängliche Quellen, die die Bedienung sehr gut beschreiben: als Einstieg empfiehlt sich die Anleitung von Stephen Berman (2013). Weiterhin verfügt Cosmas II über ein sehr gutes Hilfe-Portal (Cosmas II 2012). Das Wesentliche (kombiniert aus beiden Quellen und ergänzt) wird hier beschrieben.

Der Zugang zum Korpus der geschriebenen Sprache erfolgt entweder direkt über die Internet-Schnittstelle (<http://www.ids-mannheim.de/cosmas2/web-app/>) oder über eine Windows-Applikation, die auf die Festplatte heruntergeladen werden muss.

Installieren der Cosmas IIwin Suchmaschine

Ins Google- oder Adressen-Suchfeld http://www1.ids-mannheim.de/start/	eingeben
Forschung → laufende Projekte → Cosmas II → Cosmas IIwin	anklicken.
Detaillierte Informationen Applikation herunterladen und installieren (...).	anklicken.

Beim Herunterladen und Installieren der Anleitung folgen. Es ist sehr einfach!

Zugang zur Cosmas IIweb-Schnittstelle

Ins Google- oder Adressen-Suchfeld http://www.ids-mannheim.de/cosmas2/	eingeben
Cosmas IIweb	anklicken.
Aktuelle Version: Applikation starten.	anklicken

Für die Recherche sowohl über die Suchmaschine als auch über das Web benötigt man eine Registrierung. 57

Die Funktionen von CosmasIIwin und CosmasIIweb sind fast ident. CosmasIIwin ermöglicht die sog. „graphische Suche“: das gewünschte Kästchen (links) auf die graue Fläche ziehen, durch einen Doppelklick öffnen und ins Suchfeld den Suchbegriff eingeben. Des Weiteren hat es einige zusätzliche Funktionen, die CosmasIIweb nicht hat. Für den ersten Einstieg wird CosmasIIweb empfohlen, die folgende Beschreibung bezieht sich terminologisch auch auf die Web-Applikation.

Abfragen

Die sprachlichen Elemente, die sich mit Cosmas II abfragen lassen, sind übersichtlich unter: Online Hilfe → Suchanfrage → Zeilenangabe → Beispiele aufgeführt. Die Syntax der häufigsten Abfragen ist hier mit Erklärungen ab Seite 60 zu finden. Im Folgenden werden einige Recherchen Schritt für Schritt beschrieben, somit wird die (oft mühsame) Suche im umfangreichen Hilfe-Portal hinfällig.

ЗАКЛЮЧЕНИЕ

Корпусная лингвистика представляет собой относительно новое направление в лингвистической науке, которое открывает огромные возможности для исследования единиц любого языкового уровня в реальном употреблении, т. е. с учетом того, в какой ситуации то или иное высказывание было произведено. На сегодняшний день созданы большие национальные корпуса многих языков, а также корпуса, предназначенные для узкоспециализированных исследований. Компьютерные технологии позволяют быстро находить и обрабатывать информацию, проводить научные исследования в различных областях наук, а также междисциплинарные исследования.

Развитие информационных технологий в значительной степени облегчило доступ к корпусам с аутентичными текстами и позволило расширить использование лингвистических корпусов в обучении иностранному языку. На базе корпусов преподаватель может создавать практические и контрольно-измерительные материалы. Большую ценность в изучении как родного, так и иностранного языка, имеют самостоятельные эмпирические исследования, нацеленные на формирование универсальных учебных действий у обучающихся. Современная система образования требует не вкладывать в учеников готовые правила, а учить их учиться. Электронные корпуса текстов решают множество практических и исследовательских задач, а также формируют у обучающихся компетентность в коммуникативной и информационной областях и их способность к самообразованию и саморазвитию в соответствии с новым федеральным образовательным государственным стандартом.

Настоящее учебно-методическое пособие составлено к курсу «Введение в корпусную лингвистику» и может быть рассчитано на широкий круг изучающих основы корпусной лингвистики: от студентов до преподавателей и просто заинтересованных лиц. Оно может служить подспорьем в обучении и стимулом для дальнейших самостоятельных исследований и поиска возможностей применения корпусных технологий.

Учебное текстовое электронное издание

Уразаева Наиля Радифовна

КОРПУСНЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ

Учебно-методическое пособие

1,09 Мб

1 электрон. опт. диск

г. Магнитогорск, 2018 год
ФГБОУ ВО «МГТУ им. Г.И. Носова»
Адрес: 455000, Россия, Челябинская область, г. Магнитогорск,
пр. Ленина 38

ФГБОУ ВО «Магнитогорский государственный
технический университет им. Г.И. Носова»
Кафедра романо-германской филологии и перевода
Центр электронных образовательных ресурсов и
дистанционных образовательных технологий
e-mail: ceor_dot@mail.ru