

**федеральное государственное автономное образовательное учреждение высшего образования
Первый Московский государственный медицинский университет имени И.М. Сеченова
Министерства здравоохранения Российской Федерации
(Сеченовский Университет)**

Институт иностранных языков для профессиональных целей

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Корпусная лингвистика

основная профессиональная Высшее образование - бакалавриат - программа бакалавриата

45.00.00 Языкознание и литературоведение

45.03.04 Интеллектуальные системы в гуманитарной сфере

Профиль: Интеллектуальные системы в социологии и медицине

Интеллектуальные системы в лингвистике и медицине

Раздел дисциплины (тема): **Основы корпусной лингвистики: Применение корпусных технологий в лингвистике**

Вид оценочного средства: *Тестовое задание*

№	Вопрос и варианты ответа	эталон ответа	уровень применения
1.	Что такое лингвистический корпус? А. Большой, структурированный и размеченный массив текстов в электронном виде. В. Любой набор текстов на жестком диске компьютера. С. Электронная библиотека художественной литературы.	А	ПА
2.	Какое свойство корпуса означает, что он содержит тексты разных жанров в пропорциях, соответствующих их доле в языке? А. Сбалансированность. В. Полнота. С. Хронологичность.	А	ПА
3.	Как называется процесс добавления в текст метainформации (например, об авторе, годе создания, жанре)? А. Метаразметка. В. Токенизация. С. Лемматизация.	А	ПА
4.	Что такое конкорданс? А. Список всех употреблений слова в контексте, обычно выровненный по центру. В. Частотный список всех слов корпуса. С. Грамматический словарь, прилагаемый к корпусу.	А	ПА

5.	Какой корпус считается первым в мире компьютерным корпусом? А. Брауновский корпус (Brown Corpus). В. Британский национальный корпус (BNC). С. Национальный корпус русского языка (НКРЯ).	А	ПА
6.	Чем отличается корпус от простого архива текстов? А. Наличием лингвистической разметки и метаданных. В. Большим объемом. С. Возможностью скачать тексты.	А	ПА
7.	Что такое параллельный корпус? А. Корпус, содержащий тексты на одном языке и их переводы на другой язык. В. Корпус, содержащий тексты только на двух языках. С. Корпус, где тексты расположены в две колонки.	А	ПА
8.	Как в корпусной лингвистике называется элементарная единица текста (слово, знак препинания)? А. Токен. В. Тип (type). С. Лемма.	А	ПА
9.	Что изучается с помощью коллокаций? А. Устойчивые сочетания слов (частотная сочетаемость). В. Грамматическую структуру предложений. С. Историю происхождения слов.	А	ПА
10.	Что означает аббревиатура PoS-tagging? А. Разметка частей речи (Part of Speech tagging).	А	ПА

	В. Программное обеспечение для корпусов. С. Стандарт обмена данными.		
11.	Для чего используется мера ассоциации T-score или MI (взаимная информация)? А. Для определения силы связи между словами в коллокации. В. Для подсчета общего количества слов в тексте. С. Для автоматического перевода.	А	ПА
12.	Какой тип корпуса предназначен для изучения изменений в языке с течением времени? А. Диахронический корпус. В. Синхронический корпус. С. Мониторный корпус.	А	ПА
13.	Что такое лемма? А. Начальная (словарная) форма слова. В. Конкретное употребление слова в тексте. С. Грамматический признак слова.	А	ПА
14.	Что представляет собой разметка по ключевым словам (KWIC)? А. Вывод искомого слова в окружении контекста слева и справа. В. Список самых редких слов в корпусе. С. Морфологический разбор слова.	А	ПА
15.	Какой из перечисленных корпусов является национальным корпусом русского языка? А. НКРЯ. В. BNC. С. COCA.	А	ПА

16.	Что такое нормализация частоты (ipm)? A. Приведение частоты к общему знаменателю (например, на миллион слов) для сравнения текстов разного объема. B. Подсчет только уникальных слов. C. Удаление шума из текста.	A	ПА
17.	Какая программа чаще всего используется как бесплатный конкордансер для локальных текстов? A. AntConc. B. Microsoft Word. C. Adobe Reader.	A	ПА
18.	Что понимают под снятием омонимии в корпусе? A. Автоматическое или ручное приписывание слову правильного грамматического значения в контексте. B. Удаление всех повторяющихся слов. C. Поиск слов, одинаково звучащих.	A	ПА
19.	Какой корпус пополняется постоянно новыми текстами и отражает текущее состояние языка? A. Мониторный корпус. B. Статический корпус. C. Учебный корпус.	A	ПА
20.	Что такое N-грамма? A. Непрерывная последовательность из N элементов (слов или символов). B. Программа для построения графиков. C. Специальный вид запроса.	A	ПА

21.	Чем характеризуется репрезентативность корпуса? А. Способностью адекватно представлять язык или его вариант. В. Наличием графических иллюстраций. С. Большим объемом в гигабайтах.	А	ПА
22.	Какие данные содержит метаразметка текста? А. Информацию об авторе, времени создания, жанре. В. Частеречную принадлежность каждого слова. С. Синтаксические связи между словами.	А	ПА
23.	Что такое тип (type) в корпусном анализе? А. Уникальная словоформа или лексема, в отличие от каждого конкретного употребления (токена). В. Тип предложения (вопросительное, повествовательное). С. Размер шрифта в исходном документе.	А	ПА
24.	Для чего используется сопоставимый корпус? А. Для сравнения языков через тексты, подобранные по сходным критериям, но не являющиеся переводами. В. Для сравнения оригинала с его переводом. С. Для сопоставления устной и письменной речи.	А	ПА
25.	Что обозначает термин "ключевые слова" (keywords) в корпусной лингвистике? А. Слова, частота которых статистически значимо выше в данном тексте по сравнению с основным корпусом. В. Самые редкие слова в тексте. С. Заголовки разделов и глав.	А	ПА
26.	Какой уровень разметки отвечает за связи между	А	ПА

	словами в предложении? А. Синтаксический (парсинг). В. Морфологический. С. Фонетический.		
27.	Что такое тэгсет (tagset)? А. Полный набор меток, используемых для разметки корпуса. В. Компьютерная сеть для лингвистов. С. Набор инструментов для анализа.	А	ПА
28.	Какой корпус содержит транскрипции звучащей речи? А. Устный корпус. В. Параллельный корпус. С. Диахронический корпус.	А	ПА
29.	Что такое коллигация? А. Связь грамматических категорий (сочетаемость на грамматическом уровне). В. Лексическая сочетаемость слов. С. Сочетаемость звуков в речи.	А	ПА
30.	Что изучает корпусная лексикография? А. Составление словарей с использованием корпусных данных. В. Правила написания слов. С. Происхождение слов в корпусе.	А	ПА

Раздел дисциплины (тема): **Использование лингвистических корпусов: Язык поисковых запросов в корпусной лингвистике**

Вид оценочного средства: *Тестовое задание*

№	Вопрос и варианты ответа	эталон ответа	уровень применения
1.	Что такое конкорданс в корпусной лингвистике? А. Список всех употреблений слова в контексте, обычно выровненный по центру В. Алфавитный список всех слов корпуса С. Программа для морфологической разметки текста	А	ПА
2.	Какой формат представления конкорданса называется KWIC (Key Word In Context)? А. Искомое слово выводится в центре с контекстом слева и справа В. Слова выводятся только в алфавитном порядке С. Результаты группируются по частям речи	А	ПА
3.	Что такое частотный список (frequencylist)? Перечень языковых единиц, упорядоченный по частоте их встречаемости в корпусе. Список самых редких слов языка. Алфавитный указатель слов в корпусе.	А	ПА
4.	Для чего используется логарифмическое правдоподобие (log-likelihood) в корпусном анализе? Для сравнения частот слов в двух разных корпусах и выявления статистически значимых различий. Для построения синтаксических деревьев. Для перевода текстов с одного языка на другой.	А	ПА

5.	Что такое семантическая разметка корпуса? Разметка, приписывающая словам метки смысловых категорий (например, 'человек', 'животное', 'артефакт'). Разметка, указывающая на часть речи. Разметка ударений в словах.	А	ПА
6.	Какой из этих корпусов является корпусом современного американского английского? COCA (Corpus of Contemporary American English). BNC (British National Corpus). НКРЯ	А	ПА
7.	Что изучается с помощью KWIC-конкорданса? Окружение слова, его типичные контексты употребления. Только левое окружение слова. Этимологию слова.	А	ПА
8.	Какая проблема возникает при автоматической морфологической разметке? Грамматическая омонимия (неоднозначность разбора). Отсутствие знаков препинания. Слишком мелкий шрифт в исходных текстах.	А	ПА
9.	Что такое лемматизация? Процесс приведения словоформы к её начальной (словарной) форме — лемме.	А	ПА

	Процесс разбиения текста на предложения. Процесс удаления стоп-слов.		
10.	Что такое подкорпус? Часть корпуса, отобранная по определенным критериям (жанру, времени, автору) для решения конкретной задачи. Уменьшенная копия корпуса. Корпус, содержащий только приставки и суффиксы.	А	ПА
11.	Что показывает мера взаимной информации (MI) при анализе коллокаций? Насколько сильна связь между словами (насколько часто они встречаются вместе по сравнению со случайностью), часто выделяя терминологические сочетания. Только частоту совместной встречаемости. Расстояние между словами в тексте.	А	ПА
12.	Каков примерный объем Корпуса современного американского английского (COCA)? А. Более 560 миллионов слов В. 100 миллионов слов С. 1 миллиард слов	А	ПА
13.	Каков примерный объем Британского национального корпуса (BNC)? А. 100 миллионов слов В. 560 миллионов слов С. 50 миллионов слов	А	ПА
14.	Какое жанровое распределение в корпусе COCA? А. Примерно поровну (20%) между устной речью,	А	ПА

	художественной литературой, журналами, газетами и научными текстами В. 90% письменные тексты, 10% устная речь С. Только газетные тексты		
15.	Каково соотношение устной и письменной речи в BNC? А. 10% устная речь, 90% письменная В. 50% устная речь, 50% письменная С. Только письменная речь	А	ПА
16.	Какое преимущество дает больший объем СОСА по сравнению с BNC? А. Возможность изучать низкочастотные конструкции и редкие коллокации В. Более быстрый поиск С. Удобный интерфейс	А	ПА
17.	Какие слова можно изучать с помощью СОСА, но сложнее изучать с помощью BNC? А. Неологизмы последних десятилетий (website, blog, globalization) В. Древнеанглийскую лексику С. Только термины	А	ПА
18.	Как часто пополняется корпус СОСА? А. Ежегодно (по 20 млн слов в год) В. Раз в десятилетие С. Никогда не пополняется	А	ПА
19.	Какие подкорпусы входят в состав Национального корпуса русского языка (НКРЯ)? А. Основной, синтаксический, поэтический, устный, обучающий, диалектный, параллельный В. Только основной и газетный	А	ПА

	С. Только письменные тексты		
20.	Что представляет собой синтаксический корпус НКРЯ (Treebank)? А. Корпус с размеченными деревьями зависимостей для каждого предложения В. Корпус только простых предложений С. Корпус текстов по синтаксису	А	ПА
21.	Чем Обучающий корпус НКРЯ отличается от Основного? А. В нем используется адаптированная грамматическая разметка (1,2,3 склонение; 1,2 спряжение) и включены тексты школьной программы В. Он содержит только стихи С. Он не имеет разметки	А	ПА
22.	Какая дополнительная ступень разметки введена в Устном корпусе НКРЯ, отсутствующая в Основном? А. Разметка реплик с социологическими параметрами (пол, возраст, год рождения говорящего) В. Разметка по жанрам С. Автоматическая разметка	А	ПА
23.	Какой тип поиска стал возможен благодаря социологической разметке в Устном корпусе НКРЯ? А. Поиск по полу, возрасту или году рождения говорящего В. Поиск по географическим названиям С. Поиск только по авторам текстов	А	ПА
24.	Какие исторические корпуса входят в состав НКРЯ? А. Древнерусский (XI-XIV вв.), старорусский	А	ПА

	(XV-XVII вв.), корпус берестяных грамот В. Только корпус XVIII века С. Только советский корпус		
25.	Что представляет собой мультимедийный корпус НКРЯ? А. Корпус, где текст синхронизирован с видео- или аудиозаписью, а в некоторых фильмах размечены жесты В. Корпус только аудиозаписей С. Корпус с картинками	А	ПА
26.	Какие основные функции (инструменты) присутствуют в программе AntConc и Sketch Engine? А. Конкорданс, список слов, коллокации, N-граммы, ключевые слова В. Только переводчик С. Только проверка орфографии	А	ПА
27.	Что позволяет определить функция "коллокации" (Collocates) в корпусных программах? А. Устойчивые сочетания слов, статистически значимые связи между словами В. Частоту отдельных слов С. Грамматическую структуру предложений	А	ПА
28.	Для чего используется функция "N-граммы" (N-grams)? А. Для поиска повторяющихся последовательностей из N слов (например, биграммы, триграммы) В. Для построения графиков частотности С. Для перевода текста	А	ПА
29.	Какая платформа предоставляет больше возможностей для выбора персональных настроек среди корпусных программ? А. Sketch Engine	А	ПА

	B. AntConc C. WordSmith Tools		
30.	Что такое коллокация? A. Статистически устойчивое сочетание двух или более слов. B. Любое сочетание слов в предложении. C. Грамматическая конструкция.	A	ПА
31.	Какой корпус содержит тексты, созданные в разные исторические эпохи? A. Диахронический корпус (например, Исторический корпус русского языка). B. Мониторный корпус. C. Параллельный корпус	A	ПА
32.	Что такое транскрипция в контексте устного корпуса? A. Письменная запись звучащей речи по определенным правилам. B. Переписывание текста от руки. C. Перевод текста на другой язык.	A	ПА
33.	Какая программа является примером корпусного менеджера? A. Sketch Engine. B. Microsoft Excel. C. Adobe Photoshop.	A	ПА
34.	Что означает помета "снятая омонимия" в НКРЯ? A. Для каждого слова в контексте вручную или автоматически выбран правильный грамматический вариант. B. Все слова в корпусе имеют только одно значение. C. Из корпуса удалены все многозначные слова.	A	ПА
35.	Какое требование предъявляется к текстам для	A	ПА

	включения их в корпус? А. Машиночитаемый формат. В. Наличие иллюстраций. С. Наличие твердой обложки в оригинале.		
36.	Что такое синтаксический парсинг? А. Автоматический синтаксический анализ, приписывание предложению структуры зависимостей или составляющих. В. Разбиение текста на абзацы. С. Проверка орфографии	А	ПА
37.	Что подразумевается под "корпусной компетенцией" исследователя? А. Умение правильно ставить запросы, интерпретировать данные и владение методами корпусного анализа. В. Умение коллекционировать тексты. С. Знание всех существующих корпусов мира	А	ПА
38.	Что такое мультимодальный корпус? А. Корпус, включающий не только текст, но и аудио, видео, а иногда и жестикуляцию. В. Корпус на нескольких языках. С. Корпус, размеченный несколькими способами	А	ПА
39.	Какой из этих параметров относится к метаразмечке текста? А. Пол автора. В. Окончание существительного. С. Вид глагола.	А	ПА
40.	В чем отличие T-score от MI (взаимной информации)? А. T-score подчеркивает частотность сочетания (доверительный интервал), а MI — силу связи (не случайность встречи). В. T-score используется только для глаголов, а MI — для существительных.	А	ПА

	С. Т-score — это устаревший показатель, а MI — современный		
--	--	--	--

Вид оценочного средства: *Собеседование по темам (вопросам) для самостоятельной подготовки*

критерий	параметры оценивания
полнота и логичность изложения	соответствие содержания ответа поставленному вопросу, структурированность и последовательность изложения
степень освоения теоретического материала	корректность использования терминологии, опора на изученные источники (учебные пособия, научные публикации)
самостоятельность мышления	способность интерпретировать изученный материал, избегая дословного воспроизведения текста
аргументация личной позиции	умение формулировать и обосновывать собственное мнение по дискуссионным аспектам темы

Оценка «удовлетворительно»/ зачтено Ответ полностью соответствует установленным критериям: демонстрирует глубокое понимание материала, логичность изложения, корректное использование терминологии, опору на изученные источники и способность к аргументированной самостоятельной интерпретации.

Ответ в целом удовлетворяет предъявляемым требованиям, однако содержит отдельные недостатки: незначительные погрешности в отборе или анализе материала, отдельные нарушения логики изложения. Вместе с тем, общая логика и аргументация сохраняются.

Оценка «неудовлетворительно»/ не зачтено

Содержание источника понято в основных чертах, однако ответ строится преимущественно на репродуктивном воспроизведении материала с опорой на один источник. Самостоятельный анализ и обобщение отсутствуют.

Ответ свидетельствует о неспособности студента к самостоятельному анализу материала: терминологический аппарат не используется или применяется некорректно, изложение отличается нелогичностью и неполнотой. Оценка выставляется также в случаях, когда ответ не соответствует поставленному вопросу либо обнаруживает полное непонимание темы.

Вопрос	эталон ответа	уровень применения
1. Чем принципиально отличается поиск информации в корпусном менеджере от поиска в обычной поисковой системе (например, Google или Яндекс)? Какие задачи можно решить с помощью корпусного менеджера, но нельзя (или сложно) решить с помощью поисковика?	1. Цель поиска: Поисковик ищет документы по теме, корпусный менеджер ищет лингвистические единицы (слова, конструкции) в документах. 2. Язык запросов: Корпусный менеджер позволяет искать по леммам, грамматическим признакам, с учётом расстояния; поисковик — только по словоформам. 3. Разметка: Корпус содержит лингвистическую разметку (части речи, синтаксис), поисковик — нет. 4. Результат: Корпус выдает конкорданс (контексты) и статистику, поисковик — список ссылок на документы. 5. Задачи корпусного менеджера: изучение грамматической сочетаемости, подсчет частотности разных форм, анализ коллокаций, диахронические исследования.	ПА
2. Представьте, что вам нужно изучить все формы родительного падежа множественного числа существительных в русском языке. Какой инструмент (конкордансер или корпусный менеджер) вы выберете и почему? Какие ограничения могут возникнуть при ручном сборе материала без использования корпуса?	1. Инструмент: Корпусный менеджер с грамматической разметкой (например, НКРЯ). 2. Почему: можно составить запрос: существительное + падеж=родительный + число=множественное. Программа выдаст все примеры автоматически. 3. Ограничения ручного сбора: трудоемкость и огромные временные затраты; субъективность и неполнота выборки (исследователь может пропустить редкие формы); невозможность получить точные статистические данные; отсутствие гарантии репрезентативности	ПА
3. Почему формат KWIC (Key	1. Информативность KWIC: позволяет увидеть слово в его	ПА

Word In Context) считается наиболее информативным для представления результатов поиска? Какие закономерности можно увидеть, сортируя конкорданс по левому или правому контексту?	типичном окружении, быстро сравнивая левый и правый контекст. Экономит время, так как не нужно открывать каждый документ целиком. 2. Сортировка по левому контексту помогает выявить типичные подлежащие или определения, стоящие перед словом. 3. Сортировка по правому контексту помогает выявить типичные дополнения или управление (для глаголов). Результат: позволяет увидеть синтагматические паттерны (коллигации и коллокации) на одном экране	
4. В чем разница между поиском по точной словоформе и поиском по лемме? Приведите примеры исследовательских задач, где обязателен поиск по лемме, и задач, где необходим поиск по конкретной словоформе.	1. Поиск по словоформе: ищет только одну конкретную форму слова (например, "столом"). Нужен, когда исследуется конкретная грамматическая форма. 2. Поиск по лемме: ищет все грамматические формы слова (стол, стола, столу...). Нужен, когда исследуется лексема в целом. 3. Задачи для леммы: изучение семантики слова, его сочетаемости, частотности концепта. 4. Задачи для словоформы: изучение грамматических особенностей (например, как часто существительные употребляются в творительном падеже).	ПА
5. Какие возможности открывает перед исследователем использование грамматических тегов (поиск по частям речи и грамматическим признакам)? Приведите пример запроса, который невозможно составить без использования	1. Возможности: позволяет искать не конкретные слова, а целые классы слов или грамматические конструкции. 2. Пример запроса: "Найти все прилагательные в превосходной степени, стоящие перед существительным в родительном падеже". Без тегов это невозможно, так как слова могут быть разными, но грамматическая конструкция — общей.	ПА

грамматической разметки.		
6. Сравните логику работы операторов "И" и "ИЛИ" в языке запросов. В каких случаях для поиска коллокаций удобнее использовать запрос с оператором расстояния (например, +1, -2), а не простой перебор слов?	1. "И" (AND): Оба условия должны выполняться одновременно. Сужает поиск. 2. "ИЛИ" (OR): должно выполняться хотя бы одно условие. Расширяет поиск. 3. Оператор расстояния: Нужен, когда важен порядок слов и дистанция между ними (например, прилагательное + существительное). Позволяет найти "интересная книга", но не "книга была интересная", если задано расстояние 1. 4. Простой перебор не учитывает порядок и расстояние, выдавая много шума.	ПА
7. Сравните интерфейсы и функционал НКРЯ и СОСА (или BNC). Какие функции, доступные в зарубежных корпусах, вы хотели бы видеть в русских корпусах, и наоборот?	1. Сильные стороны НКРЯ: глубокая грамматическая разметка, множество подкорпусов (поэтический, диалектный, обучающий), удобный поиск по снятой омонимии. 2. Сильные стороны СОСА/BNC: мощные инструменты для анализа коллокаций (сортировка по MI, T-score), удобное сравнение частот по жанрам и годам, интеграция с N-граммами. 3. Что взять из СОСА в НКРЯ: более наглядную визуализацию коллокаций и графиков изменения частотности по годам. 4. Что взять из НКРЯ в СОСА: более детальную грамматическую разметку (особенно для славянских языков).	ПА
8. Почему при работе с зарубежными корпусами	1. Жанровый состав влияет на лексику (в газетах одни слова, в художественной литературе —	ПА

(COCA, BNC) важно учитывать их жанровый состав и хронологические рамки? Приведите пример исследовательской задачи, для которой BNC подойдет лучше, чем COCA, и наоборот.	другие). 2. Хронологические рамки важны для изучения неологизмов или устаревшей лексики. 3. BNC лучше для изучения классического британского варианта конца XX века (1980-1993). 4. COCA лучше для изучения современного американского варианта, неологизмов последних лет (мониторный, ежегодно пополняется).	
9. Каковы преимущества и недостатки использования веб-интерфейсов корпусов по сравнению с локальными корпусными менеджерами (например, AntConc)? В каких случаях исследователю может потребоваться скачать данные и работать с ними локально?	1. Преимущества веб-интерфейсов: не требуют установки, всегда обновляются, содержат готовую разметку, большой объем данных. 2. Недостатки веб-интерфейсов: ограниченный функционал (нельзя применить свой тип анализа), зависимость от интернета, ограничения на объем выгрузки. 3. Когда нужен локальный анализ: работа с собственными текстами, применение специфических статистических методов, отсутствие интернета, создание учебных материалов.	ПА
10. Чем устный корпус принципиально отличается от письменного с точки зрения лингвистической разметки? Какие исследовательские задачи можно решить только с помощью устного корпуса (например, подкорпуса «Один речевой день» в НКРЯ)?	1. Отличия разметки: транскрипция с учетом особенностей устной речи (хезетации: "э-э", "м-м"), обрывы слов, фонетические особенности, просодическая разметка (иногда), социолингвистическая разметка говорящих. 2. Задачи: изучение хезитационных явлений (слов-паразитов), анализ разговорного синтаксиса (неполные предложения), исследование социальной вариативности речи (как говорят мужчины/женщины, разных возрастов).	ПА

11. Как социолингвистическая разметка (пол, возраст, год рождения говорящего) в устном корпусе расширяет возможности исследователя? Какие гипотезы о речевом поведении разных групп можно проверить с ее помощью?	Расширение возможностей: позволяет связать языковые явления с социальными характеристиками говорящих. Гипотезы для проверки: употребляют ли женщины определенные слова чаще мужчин? Меняется ли частота использования заимствований с возрастом? Используют ли молодые люди больше сленга и слов-паразитов ("типа", "короче")? Как меняется речевое поведение с течением времени (сравнение записей одного человека в разные годы)?	ПА
12. Сравните функциональные возможности AntConc и Sketch Engine. Для каких задач (исследовательских, учебных, переводческих) каждый из этих инструментов подходит лучше	AntConc: Бесплатный, легкий, работает локально. Лучше для: работы с собственными небольшими корпусами, учебных целей, базового анализа (частотные списки, конкордансы, простые коллокации). Sketch Engine: Платный, мощный, облачный, огромные готовые корпуса. Лучше для: профессиональных исследований, лексикографии (составление словарей), переводческой работы (поиск переводных эквивалентов в параллельных корпусах), сложного статистического анализа.	ПА
13. Какие этапы подготовки текстов необходимы перед загрузкой их в локальный конкордансер (например, AntConc)? Почему простые	Этапы подготовки: сбор текстов, очистка от "мусора" (тителы, колонтителы, лишние пробелы, html-теги), сохранение в формате .txt в кодировке UTF-8.	ПА

текстовые файлы (.txt) предпочтительнее для анализа, чем файлы .docx или .pdf?	Почему .txt предпочтительнее: не содержит служебной информации о форматировании (шрифты, отступы), которая мешает анализу. Универсальный формат, читается любой программой. Файлы .docx/.pdf могут содержать скрытые символы и требуют дополнительного парсинга для извлечения чистого текста.	
14. Если бы перед вами стояла задача создать собственный мини-корпус текстов определенного автора или жанра, какие инструменты вы бы использовали для сбора, очистки, разметки и анализа данных? Опишите полный цикл работы	1. Сбор: скачивание текстов из открытых источников (библиотеки, сайты), сканирование и OCR (распознавание) для бумажных книг. 2. Очистка: текстовый редактор (блокнот, Notepad++) для удаления лишнего, регулярные выражения для автоматической чистки. 3. Разметка (опционально): если нужна автоматическая морфологическая разметка — использование MyStem (для русского) или TreeTagger. 4. Анализ: загрузка .txt файлов в AntConc (или аналоги) для построения конкордансов, частотных списков, выявления коллокаций и ключевых слов. 5. Интерпретация: формулирование выводов на основе полученных данных.	ПА
15. В чем ценность мультимедийного корпуса (например, МУРКО) для лингвистических исследований? Какие аспекты коммуникации становятся доступны для анализа благодаря синхронизации текста, аудио и видео?	1. Ценность: позволяет изучать невербальные компоненты коммуникации в синхронизации с речью. 2. Аспекты анализа: жесты, мимика, интонация, паузы, визуальный контекст ситуации. 3. Пример: можно изучать, какие жесты сопровождают дейктические слова ("здесь", "там", "этот"), или как интонация меняет смысл	ПА

	высказывания.	
--	---------------	--

Раздел дисциплины (тема): **Использование лингвистических корпусов: Статистическая обработка корпусных данных**

Вид оценочного средства: *Собеседование по темам (вопросам) для самостоятельной подготовки*

критерий	параметры оценивания
полнота и логичность изложения	соответствие содержания ответа поставленному вопросу, структурированность и последовательность изложения
степень освоения теоретического материала	корректность использования терминологии, опора на изученные источники (учебные пособия, научные публикации)
самостоятельность мышления	способность интерпретировать изученный материал, избегая дословного воспроизведения текста
аргументация личной позиции	умение формулировать и обосновывать собственное мнение по дискуссионным аспектам темы

Оценка «удовлетворительно»/ зачтено

Ответ полностью соответствует установленным критериям: демонстрирует глубокое понимание материала, логичность изложения, корректное использование терминологии, опору на изученные источники и способность к аргументированной самостоятельной интерпретации.

Ответ в целом удовлетворяет предъявляемым требованиям, однако содержит отдельные недостатки: незначительные погрешности в отборе или анализе материала, отдельные нарушения логики изложения. Вместе с тем, общая логика и аргументация сохраняются.

Оценка «неудовлетворительно»/ не зачтено

Содержание источника понято в основных чертах, однако ответ строится преимущественно на репродуктивном воспроизведении материала с опорой на один источник. Самостоятельный анализ и обобщение отсутствуют.

Ответ свидетельствует о неспособности студента к самостоятельному анализу материала: терминологический аппарат не используется или применяется некорректно, изложение отличается

нелогичностью и неполнотой. Оценка выставляется также в случаях, когда ответ не соответствует поставленному вопросу либо обнаруживает полное непонимание темы.

№	Вопрос	эталон ответа	уровень применения
1.	Почему простое перечисление найденных примеров (цитат) недостаточно для серьезного лингвистического исследования? Что добавляет статистическая обработка данных?	1. Статистика позволяет перейти от отдельных наблюдений к объективным закономерностям. 2. Она дает возможность измерить, насколько то или иное явление типично или редко для языка в целом. 3. Статистика позволяет сравнивать разные тексты, жанры или периоды не на глаз, а на основе точных числовых данных (нормализованных частот, критериев значимости). 4. Без статистики высок риск субъективных выводов, основанных на случайно попавшихся примерах.	ПА
2.	Объясните разницу между понятиями "токен" и "тайп" (тип). Приведите пример для небольшого текста (например, "дом, который построил Джек").	1. Токен — это каждое конкретное употребление слова в тексте (общее количество слов). 2. Тайп (тип) — это уникальная словоформа или лексема (количество разных слов). 3. Пример: В предложении "Дом стоит у дома" — 4 токена (дом, стоит, у, дома), но 3 тайпа (дом, стоит, у), так как слово "дом" представлено двумя разными токенами (им.п. и род.п.).	ПА
3.	Что такое TTR (Type-TokenRatio) и о чем говорит его значение? В	1. TTR — это отношение количества уникальных слов (типов) к общему	

	каких случаях сравнение TTR может быть некорректным?	количеству слов (токенов). Показывает лексическое разнообразие текста. 2. Высокий TTR — богатый словарь, разнообразная лексика (характерно для поэзии, художественных текстов). 3. Низкий TTR — лексическая бедность, много повторов (характерно для разговорной речи, некоторых жанров официальных документов). 4. Некорректность сравнения: TTR сильно зависит от объема текста (чем текст длиннее, тем TTR естественно падает). Сравнивать можно только тексты примерно равной длины или использовать стандартизированный TTR (STTR).	
4.	Чем частотный словарь отличается от алфавитно-частотного? Для каких практических задач удобнее каждый из них?	1. Частотный словарь: слова упорядочены по убыванию частоты (от самых частотных к редким). Удобен для выделения ядра лексики, определения слов, покрывающих большую часть текста. 2. Алфавитно-частотный словарь: слова расположены по алфавиту, но рядом указана частота. Удобен для быстрого поиска конкретного слова и проверки его частоты.	ПА
5.	Какие слова (части речи) обычно занимают первые строчки в частотном словаре любого языка и почему? Что такое "стоп-слова" и зачем их иногда исключают из	1. Первые строчки: служебные слова (предлоги, союзы, частицы) и местоимения. 2. Причина: они выполняют грамматические функции, встречаются в любом тексте	ПА

	анализа?	независимо от темы, их количество в языке ограничено. 3. Стоп-слова — это список самых частотных служебных слов, которые исключают из анализа, когда исследователя интересует содержательная (знаменательная) лексика — существительные, прилагательные, глаголы.	
6.	Что такое гапакс (hapaxlegomenon)? Почему изучение гапаксов важно для лингвистов и лексикографов?	1. Гапакс — слово, встретившееся в корпусе только один раз. 2. Важность для лингвистов: гапаксы часто являются ключевыми для понимания идиостиля автора, могут быть авторскими неологизмами (окказионализмами). 3. Важность для лексикографов: большое количество гапаксов в корпусе — показатель того, что корпус хорошо отражает редкую лексику языка. Гапаксы помогают выявлять редкие и устаревшие слова.	
7.	Сформулируйте суть закона Ципфа (Zipf's law). Как этот закон помогает понять устройство языка?	1. Суть закона: Произведение частоты слова на его ранг (место в частотном списке) есть величина приблизительно постоянная ($F \times i \approx \text{const}$). 2. Следствие: Очень частотных слов мало, они покрывают большую часть текста. Большинство слов языка — редкие. 3. Понимание языка: Закон показывает, что язык — это экономичная система, где небольшой набор единиц	ПА

		используется постоянно, а огромный запас лексики — лишь изредка.	
8.	Как можно использовать закон Ципфа для оценки того, насколько корпус "хорош" (репрезентативен)?	1. В "хорошем", большом и сбалансированном корпусе распределение частот слов должно приближаться к идеальному распределению Ципфа. 2. Сильные отклонения от закона могут указывать на то, что корпус перекошен (например, состоит из текстов одной тематики, где определенные слова встречаются неестественно часто).	ПА
9.	Что описывает нормальное распределение (гауссово) и применимо ли оно к распределению частот слов? Почему?	1. Нормальное распределение описывает явления, где большинство значений группируется вокруг среднего (рост людей, баллы за тест). 2. К распределению частот слов оно неприменимо. 3. Причина: распределение слов подчиняется закону Ципфа — это "степенное" распределение, где очень много редких явлений и очень мало частотных (так называемый "длинный хвост").	
10.	Почему для разных исследовательских задач могут требоваться данные о частоте разных единиц (словоформ, лексем, грамматических категорий)? Приведите примеры	1. Частота словоформ нужна для изучения грамматики (например, как часто используется форма прошедшего времени vs настоящего). 2. Частота лексем нужна для изучения лексики и семантики (например, как часто в языке встречается	ПА

		концепт "счастье" во всех его формах). 3. Частота грамматических категорий (частей речи, падежей) нужна для типологических исследований, определения жанровой специфики (в научных текстах больше существительных).	
11.	Что можно узнать о тексте, подсчитав частоту разных частей речи в нем? Приведите пример сравнения научного и художественного текста.	1. Научный текст: высокая частотность существительных (терминология), отглагольных существительных, низкая частотность глаголов и междометий. 2. Художественный текст: более высокий процент глаголов (динамика), прилагательных (образность), местоимений, наличие междометий. 3. Эти данные позволяют объективно (числами) подтвердить интуитивные представления о стилях и даже автоматически определять жанр текста.	ПА
12.	Как используя данные о частоте грамматических категорий, можно проследить эволюцию языка? Приведите пример с возвратными глаголами на "-ся" в русском языке.	1. Нужно сравнить частоту интересующей категории в подкорпусах разных исторических периодов. 2. Пример с "-ся": можно проследить, увеличивалась или уменьшалась частотность возвратных глаголов с XVIII по XXI век. Если частотность растет, значит, категория становится более продуктивной, если падает — возможно, она вытесняется другими конструкциями.	

		3. Аналогично можно изучать изменение частоты кратких прилагательных, деепричастий и т.д.	
13.	Какой тип корпуса необходим для изучения изменения частот слов во времени? Почему обычный (синхронический) корпус для этого не подходит?	1. Необходим диахронический корпус (например, исторический подкорпус НКРЯ, СОНА для английского), который содержит тексты, равномерно распределенные по разным временным срезам (десятилетиям, векам). 2. Обычный корпус (синхронический) представляет язык в один временной срез (обычно современный) и не позволяет увидеть динамику изменений.	ПА
14.	Что можно узнать, исследуя не только изменение частоты слова, но и изменение его контекстов (коллокаций) в диахронии? Приведите пример со словом "спутник".	1. Изменение контекстов показывает семантический сдвиг — как менялось значение слова. 2. Пример "спутник": до 1957 года: контексты — "спутник в пути", "спутник жизни" (попутчик, товарищ). После 1957 года: появляются и становятся доминирующими контексты — "искусственный спутник", "спутник Земли", "запуск спутника" (космический аппарат). Анализ коллокаций по годам позволяет точно определить момент и причины семантического сдвига.	ПА

15.	Какие экстралингвистические (внеязыковые) факторы могут влиять на изменение частоты слов в диахроническом корпусе? Приведите 2-3 примера.	Социально-политические изменения: резкий рост частоты слов "перестройка", "гласность" в конце 1980-х; падение частоты слов "товарищ", "партия" после 1991 года. Технологический прогресс: появление и резкий рост частоты слов "интернет", "смартфон", "блог" в конце XX — начале XXI века. Исторические события: всплеск частоты слов "война", "фронт", "победа" в корпусах, отражающих период 1941-1945 годов. Культурные тенденции: рост частоты заимствований (англицизмов) в периоды активных международных контактов.	

Раздел дисциплины (тема): **Использование лингвистических корпусов: Лингвистические корпусные исследования**

Вид оценочного средства: *Проект*

Конечный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий, которые позволяют оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитических, исследовательских навыков, умений практического и творческого мышления. Может выполняться в индивидуальном порядке или группой обучающихся.

Критерии оценивания:

критерий	параметры оценивания
-----------------	-----------------------------

цель и задачи	Сформулированы четко, достижимы, соответствуют теме.
выбор инструмента	Корпус и/или программа (AntConc, НКРЯ и др.) выбраны адекватно задаче.
методика и запросы	Поисковые запросы описаны, корректны, достаточны для сбора материала
наглядность данных	Есть скриншоты конкордансов, примеры контекстов, таблицы или графики.
анализ и выводы	Выводы основаны на данных, а не на общих рассуждениях, есть попытка интерпретации.
оформление отчета	Структура соблюдена (введение, основная часть, выводы), текст грамотный.

Оценка «удовлетворительно»/ зачтено

Ответ полностью соответствует установленным критериям: демонстрирует глубокое понимание материала, логичность изложения, корректное использование терминологии, опору на изученные источники и способность к аргументированной самостоятельной интерпретации.

Ответ в целом удовлетворяет предъявляемым требованиям, однако содержит отдельные недостатки: незначительные погрешности в отборе или анализе материала, отдельные нарушения логики изложения. Вместе с тем, общая логика и аргументация сохраняются.

Оценка «неудовлетворительно»/ не зачтено

Содержание источника понято в основных чертах, однако ответ строится преимущественно на репродуктивном воспроизведении материала с опорой на один источник. Самостоятельный анализ и обобщение отсутствуют.

Ответ свидетельствует о неспособности студента к самостоятельному анализу материала: терминологический аппарат не используется или применяется некорректно, изложение отличается нелогичностью и неполнотой. Оценка выставляется также в случаях, когда ответ не соответствует поставленному вопросу либо обнаруживает полное непонимание темы.

№	Тема проекта	уровень применения
	БЛОК 1: ЛЕКСИЧЕСКАЯ СЕМАНТИКА И СЛОВОУПОТРЕБЛЕНИЕ	
1.	Анализ синонимического ряда методом корпусных коллокаций <i>Задача:</i> выбрать группу синонимов (например, «быстрый — стремительный — скоростной» или «important — essential — crucial»). С помощью корпусного менеджера (НКРЯ, СОСА, BNC) извлечь типичные коллокации для каждого члена ряда. Выявить	ПА

	дифференциальные признаки — с какими существительными сочетается каждый синоним, какие семантические классы слов образуют устойчивые связи. Результат представить в виде таблицы сочетаемости и краткого анализа различий.	
2.	Исследование семантического сдвига слова в диахронии <i>Задача:</i> выбрать слово, которое изменило значение за последние 50-100 лет (например, «товарищ», «спутник», «виртуальный», «gay», «awful»). На материале диахронического корпуса (НКРЯ, СОНА — Corpus of Historical American English) проследить динамику частотности и изменение контекстов употребления по десятилетиям. Выявить момент и возможные причины семантического сдвига.	ПА
3.	Контрастивный анализ употребления лексемы в разных жанрах <i>Задача:</i> выбрать слово (например, «свобода», «демократия», «кризис»). Сравнить его частотность, коллокации и контексты употребления в двух разных подкорпусах — например, в газетном и научном (или в художественном и устном). Объяснить выявленные различия с точки зрения жанровой специфики	ПА
4.	Корпусное исследование полисемии: семантическая структура многозначного слова <i>Задача:</i> выбрать многозначное слово (например, «ключ», «язык», «голова»). Извлечь из корпуса 100-150 контекстов употребления. Провести семантическую классификацию всех употреблений, выделив основные лексико-семантические варианты. Определить частотность каждого значения и типичные контексты реализации. Построить семантическую сеть слова.	ПА
5.	Анализ ключевых слов художественного текста <i>Задача:</i> Выбрать небольшое художественное произведение (рассказ, повесть). Создать его электронную версию и загрузить в программу AntConc. С помощью инструмента Keywords сравнить частотность слов с корпусом-референсом (например, с основным корпусом НКРЯ или BNC). Выявить ключевые слова, определяющие тематику, проблематику и стилистику текста. Интерпретировать их роль в структуре произведения	ПА
	БЛОК 2: ГРАММАТИЧЕСКИЕ КОНСТРУКЦИИ И МОРФОЛОГИЯ	
6.	Вариативность грамматических форм: корпусное исследование <i>Задача:</i> выбрать грамматическую вариативную пару (например, «жду поезд» vs «жду поезда»; «чашка чая» vs «чашка чаю»; «на Украину» vs	ПА

	«в Украину»). На материале НКРЯ проследить динамику соотношения вариантов по десятилетиям. Определить, какой вариант является доминирующим в современном употреблении, и выявить факторы (лексические, стилистические), влияющие на выбор варианта.	
7.	Словообразовательные модели: продуктивность аффиксов <i>Задача:</i> выбрать словообразовательный аффикс (например, суффикс -ость, -щик, префикс анти-, супер-). На материале корпуса выявить все лексемы, образованные с помощью этого аффикса. Составить частотный список, определить наиболее продуктивные модели. Выявить, с какими основами (исконными/заимствованными, конкретными/абстрактными) сочетается аффикс	ПА
8.	Анализ грамматической омонимии и способы ее разрешения <i>Задача:</i> выбрать словоформы, регулярно проявляющие грамматическую омонимию (например, «стекло» (сущ. и гл.), «печь», «знакомый»). Извлечь из НКРЯ контексты со снятой и неснятой омонимией. Проанализировать, какие контекстуальные факторы (лексическое окружение, синтаксическая позиция) помогают снять омонимию. Оценить точность автоматического снятия омонимии в корпусе.	ПА
9.	Конструкции с отрицанием: корпусное исследование <i>Задача:</i> исследовать особенности употребления конструкций с двойным отрицанием («не могу не сказать», «нельзя не заметить») или с усилительным отрицанием («ничуть не», «отнюдь не»). Выявить частотность, жанровое распределение, типичные коллокации и семантические классы глаголов/прилагательных, входящих в эти конструкции.	ПА
	БЛОК 3: ДИСКУРСИВНЫЕ И СОЦИОЛИНГВИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ	ПА
10.	Лингвокультурный анализ концепта <i>Задача:</i> выбрать культурно значимый концепт (например, «душа», «судьба», «тоска», «privacy», «fairness»). На материале корпуса (НКРЯ	ПА

	для русского, COCA/BNC для английского) собрать языковые репрезентации концепта (ключевая лексема и ее синонимы, дериваты, устойчивые сочетания). Проанализировать метафорическую сочетаемость и выявить когнитивные признаки, формирующие структуру концепта в национальном сознании .	
11.	Особенности устной речи: корпусное исследование дискурсивных маркеров <i>Задача:</i> на материале устного подкорпуса НКРЯ исследовать употребление дискурсивных маркеров («ну», «вот», «значит», «типа», «как бы»). Выявить их функции (хезетация, структурирование дискурса, смягчение категоричности и т.д.), частотность в зависимости от социальных параметров говорящих (пол, возраст).	ПА
12.	Гендерная специфика речевого поведения <i>Задача:</i> сравнить речевое поведение мужчин и женщин на материале устного корпуса. Выявить статистически значимые различия в употреблении частей речи, модальных слов, дискурсивных маркеров, лексики определенных семантических полей. Сделать выводы о соответствии выявленных различий стереотипным представлениям о мужской и женской речи.	ПА
13.	Языковая репрезентация социально значимых феноменов в медиадискурсе <i>Задача:</i> выбрать социально значимое явление (миграция, пандемия, экология, цифровизация). На материале газетного подкорпуса или специализированного медиакорпуса проанализировать, какие языковые средства (метафоры, оценочная лексика, ключевые коллокации) используются для его репрезентации. Выявить доминирующие дискурсивные стратегии и их эволюцию во времени.	ПА
	БЛОК 4: ПЕРЕВОДОВЕДЕНИЕ И КОНТРАСТИВНАЯ ЛИНГВИСТИКА	
14.	Сопоставительный анализ переводных соответствий на материале параллельного корпуса <i>Задача:</i> выбрать лексему или грамматическую конструкцию в русском языке и проанализировать способы ее передачи на английский язык (или наоборот) с использованием параллельного подкорпуса НКРЯ.	ПА

	Выявить основные переводческие стратегии, наиболее частотные соответствия, случаи безэквивалентности. Построить типологию переводческих решений.	
15.	Корпусное исследование языка начинающих переводчиков (учебный корпус переводов) <i>Задача:</i> создать небольшой учебный корпус студенческих переводов (или использовать имеющийся ресурс). Сравнить его с корпусом профессиональных переводов (например, из параллельного корпуса) для выявления типичных ошибок, калек, неоправданных лексических и грамлексических повторов. Определить наиболее проблемные зоны при переводе с иностранного языка на родной.	ПА

ТЕСТОВЫЕ ВОПРОСЫ

№	Вопрос и варианты ответа	эталон ответа	уровень применения
1.	Основной функцией науки является: А. объяснительная функция В. коммуникативная функция С. социальная функция D. когнитивная функция	А	ПА
2.	Основные функции языка: А. когнитивная, коммуникативная, социальная В. фатическая, социальная С. экспрессивная, эстетическая D. коммуникативная, метаязыковая	А	ПА
3.	Основными принципами современного научного исследования являются: А. объективность, достоверность и воспроизводимость полученных данных	А	ПА

	В. субъективность и отсутствие воспроизводимости полученных данных С. достоверность и воспроизводимость полученных данных Д. объективность и воспроизводимость полученных данных		
4.	Объектом изучения лингвистики является: А. язык В. человек С. лексикографический справочник Д. социум	А	ПА
5.	Что лежит в основе когнитивного подхода к языку? А. понимание и изучение языка как средства познания мира, формирования и выражения мысли, хранения и организации знания в человеческом сознании, обмена знаниями В. понимание и изучение языка как средства познания мира С. понимание и изучение языка как средства формирования и выражения мысли Д. понимание и изучение языка как средства хранения и организации знания в человеческом сознании	А	ПА
6.	В центре антропоцентрической парадигмы знаний находится: А. человек В. язык С. лингвистический корпус Д. описание научных проблем	А	ПА
7.	Особое место в обработке поступающей к человеку информации занимает: А. язык В. способ передачи информации С. когниция Д. сознание человека	А	ПА

8.	Одним из основных инструментов познания является: А. язык В. социум С. когниция D. человек	А	ПА
9.	Корпусная лингвистика - комплексная научная дисциплина, которая использует компьютерные технологии для: А. выяснения природы языка, оптимизации коммуникации и решения иных лингвистических задач, создает корпуса текстов В. оптимизации коммуникации и решения иных лингвистических задач С. создает корпуса текстов для выяснения природы языка D. решения математических задач	А	ПА
10.	Корпусная лингвистика – А. раздел компьютерной лингвистики В. раздел лексикологии С. раздел лексикографии D. раздел семантики	А	ПА
11.	Корпусный менеджер – А. поисковая система, которая включает в себя набор средств для поиска данных, получения статистической информации и предоставления результатов пользователю в доступной форме В. поисковая система, которая включает в себя набор средств для поиска данных С. поисковая система для получения статистической информации D. человек, продающий лингвистические корпуса	А	ПА
12.	Корпусный менеджер должен: А. выполнять простые и сложные поисковые	А	ПА

	запросы, сортировать списки результатов, выдавать статистическую информацию В. сортировать списки результатов С. выполнять простые и сложные поисковые запросы Д. выдавать статистическую информацию		
13.	Конкорданс – А. список всех употреблений слова в контексте со ссылками на источник В. корпусный менеджер С. сформулированный запрос к корпусным данным Д. пример контекстуального употребления слова	А	ПА
14.	Корпуса могут использоваться для: А. получения справок о характеристиках лексических единиц, статистических данных, о лингвистических категориях и метаданных, жанрово-стилистических характеристиках В. получения справок о характеристиках лексических единиц С. получения статистических данных Д. получения данных о жанрово-стилистических характеристиках лексических единиц	А	ПА
15.	Представительный массив языковых данных за определенный период позволяет: А. изучать динамику процессов изменения лексического состава слова, проводить анализ лексико-грамматических характеристик в разных жанрах и у разных авторов В. проводить анализ лексико-грамматических характеристик С. получить интерпретацию полученных результатов Д. заменить интеллектуальную обработку корпусных данных	А	ПА
16.	Корпусные методы лежат в основе:	А	ПА

	А. культурометрии В. социологии С. социолингвистики Д. психолингвистики		
17.	Корпусно-ориентированный (corpus-based) подход предполагает: А. использование корпусов для проверки лингвистических теорий или гипотез В. получение статистической информации С. проведение дефиниционного анализа Д. проведение лексикографического анализа	А	ПА
18.	Корпусно-управляемый (corpus-driven) подход предполагает: А. использование корпуса как главного и единственного источника теорий о языке В. получение статистической информации С. проведение семантического анализа Д. проведение лексикографического анализа	А	ПА
19.	Объектом корпусной лингвистики является: А. корпус текстов В. лексикографический источник С. лексема Д. лексическая единица	А	ПА
20.	Предмет корпусной лингвистики: А. теоретические основы и практические механизмы создания и использования представительных массивов языковых данных, предназначенных для лингвистических исследований в интересах широкого круга пользователей В. лексический состав языка С. морфемы Д. фонемы	А	ПА
21.	Корпусная лингвистика может быть представлена в виде совокупности методов, процедур и ресурсов, имеющих дело с: А. эмпирическими данными в лингвистике В. статистическими данными в лингвистике С. математическими данными в лингвистике	А	ПА

	D. стилистическими данными в лингвистике		
22.	В настоящее время электронные корпуса часто используются в: A. исследованиях языкового многообразия (диалектов, социодиалектов) B. статистических исследованиях C. в стилистических исследованиях D. в морфологических исследованиях	A	ПА
23.	Термин «корпус» обычно обозначает: A. собрание текстов конечного фиксированного размера B. нелимитированное собрание текстов C. собрание лексикографических данных D. собрание статических данных	A	ПА
24.	Представительность корпуса получила название: A. репрезентативность B. аннотированность C. разметка D. аннотация	A	ПА
25.	Основными проблемами при создании корпуса являются: A. проблема объема и отбора материала B. проблема выбора статического критерия C. проблема создания корпусного менеджера D. проблема аннотирования	A	ПА
26.	Достаточное и пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов представляют: A. репрезентативность и сбалансированность B. аннотирование и стилистический анализ C. лексикографический и стилистический анализ D. использование корпусного менеджера и сбалансированность	A	ПА
27.	Достоверность полученных материалов определяют: A. репрезентативность и сбалансированность B. аннотирование и стилистический анализ	A	ПА

	С. лексикографический и стилистический анализ Д. использование корпусного менеджера и сбалансированность		
28.	Способность отражать все свойства языка или подязыка определяют: А. репрезентативность и сбалансированность В. аннотирование и стилистический анализ С. лексикографический и стилистический анализ Д. использование корпусного менеджера и сбалансированность	А	ПА
29.	Универсальные корпуса отражают: А. всё многообразие языка статистически корректно В. картину бытования языка фрагментарно С. всю картину бытования языка статистически некорректно Д. речевую деятельность языка	А	ПА
30.	Специфичные корпуса отражают: А. бытование некоторого языкового или культурного явления в общественной речевой практике В. существование пословиц в языке С. существование политических метафор в газетной речи Д. существование диалектов	А	ПА
31.	По способу фиксации речевого материала существуют корпуса: А. письменные, устные, смешанные В. письменные С. устные Д. смешанные	А	ПА
32.	По цели создания существуют корпуса: А. многоцелевые, специализированные, исследовательские, иллюстративные В. исследовательские и иллюстративные	А	ПА

	С. специализированные и исследовательские Д. многоцелевые		
33.	Специализированные корпуса служат для: А. решения специальной задачи В. для описания лингвокогнитивных особенностей слов С. для статистического подсчета Д. для создания корпусного менеджера	А	ПА
34.	Методология конструирования корпуса зависит от: А. от его типа В. от выбранного статистического критерия С. от корпусного менеджера Д. от языка	А	ПА
35.	Корпусы предназначены для: А. неоднократного применения многими пользователями В. однократного применения С. применения ограниченным числом пользователей Д. фиксации лингвистического материала	А	ПА
36.	Примерами метаинформации (контекстуальной информации) могут служить: А. возраст, пол, географическое происхождение участников речевого акта и т.п. В. дефиниции слова С. примеры контекстуального использования слова Д. статистические данные	А	ПА
37.	Разметка корпуса –	А	ПА

	А. приписывание текстам и их компонентам дополнительной информации – метаданных самого разного типа В. лексикографический анализ С. контекстуальный анализ Д. дискурсивный анализ		
38.	Лингвистические теги в разметке описывают: А. лексические, грамматические характеристики элементов текста В. статистические данные С. результаты поиска с помощью корпусного менеджера Д. тип корпуса	А	ПА
39.	Экстралингвистические теги в разметке описывают: А. сведения об авторе и о тексте: автор, название, год и место издания В. лексические, грамматические характеристики элементов текста С. статистические данные Д. тип корпуса	А	ПА
40.	В результате работы программ автоматической морфологической разметки каждой лексической единице приписываются: А. грамматические характеристики, набор граммов (род, число, падеж, одушевленность /неодушевленность, переходность/непереходность и т.п.) В. синтаксические характеристики С. семантические характеристики Д. пунктуационные характеристики	А	ПА
41.	В результате работы программ автоматической синтаксической разметки фиксируются: А. синтаксические связи между словами и словосочетаниями, а также	А	ПА

	синтаксические характеристики (тип предложения, синтаксическая функция слова или словосочетания) В. грамматические характеристики С. семантические характеристики Д. набор грамем		
42.	При создании корпусов для снятия неоднозначности (disambiguity) используется: А. комбинация автоматического и ручного способа обработки В. ручной способ обработки С. автоматический способ обработки Д. корпусный менеджер	А	ПА
43.	Относительное количество неправильно приписанных лемм и грамматических признаков выше: А. в веб-корпусах В. в специализированных корпусах С. в универсальных корпусах Д. устных корпусах	А	ПА
44.	Лингвистическая разметка – это: А. аннотация, основанная на лингвистических характеристиках текста В. стилистический анализ С. лексикографический анализ Д. пунктуационный анализ	А	ПА
45.	С точки зрения технологии, разметка может быть: А. автоматической, ручной или автоматической с ручной правкой В. только автоматической С. только ручной Д. только автоматической с ручной правкой	А	ПА
46.	Морфологическая разметка — это: А. частеречная разметка В. семантическая разметка С. стилистическая разметка	А	ПА

	D. грамматическая разметка		
47.	Морфологические метки включают: A. признак части речи и признаки грамматических категорий, свойственных данной части речи B. стилистические признаки C. семантические признаки D. исключительно признаки грамматических категорий	A	ПА
48.	Синтаксическая разметка является результатом: A. парсинга (автоматического синтаксического анализа), выполняемого на основе данных морфологического анализа B. семантического анализа C. лексикографического анализа D. грамматического анализа	A	ПА
49.	Синтаксическая разметка описывает: A. синтаксические связи между лексическими единицами и различные синтаксические конструкции B. грамматические связи C. семантические связи D. стилистические особенности	A	ПА
50.	Синтаксическая разметка зависит от: A. принятой формальной синтаксической модели B. корпусного менеджера C. лексикографического источника D. семантических особенностей	A	ПА
51.	Синтаксическая разметка A. раскрывает характер синтаксических связей между единицами B. приписывает словам семантическую категорию C. указывает ударение и интонацию D. фиксирует библиографические и	A	ПА

	типологические характеристики		
52.	Морфологическая разметка А. частеречная разметка В. приписывает словам семантическую категорию С. раскрывает характер синтаксических связей между единицами D. указывает ударение и интонацию	А	ПА
53.	Семантическая разметка А. приписывает словам семантическую категорию В. указывает ударение и интонацию С. фиксирует библиографические и типологические характеристики D. раскрывает характер синтаксических связей между единицами	А	ПА
54.	Семантические теги обозначают: А. семантические категории В. грамматические признаки С. стилистические особенности D. лингвокогнитивные особенности	А	ПА
55.	Семантическая разметка предусматривает: А. спецификацию значения слов, разрешение омонимии и синонимии, категоризацию слов (разряды) В. описание характера синтаксических связей между единицами С. описание грамматических признаков D. описание стилистических особенностей	А	ПА
56.	Экстралингвистическая разметка: А. фиксирует библиографические и типологические характеристики В. приписывает слова семантическую	А	ПА

	категорию С. раскрывает характер синтаксических связей между единицами Д. указывает ударение и интонацию		
57.	Экстралингвистическая разметка включает: А. внешнюю, «интеллектуальную» разметку (библиографические характеристики, типологические, тематические, социологические характеристики) В. спецификацию значения слов, разрешение омонимии и синонимии, категоризацию слов (разряды) С. описание характера синтаксических связей между единицами Д. описание грамматических признаков	А	ПА
58.	Набор метаданных определяет: А. дополнительные возможности поиска, предоставляемые корпусами исследователям В. спецификацию значения слов, разрешение омонимии и синонимии, категоризацию слов (разряды) С. описание характера синтаксических связей между единицами Д. описание грамматических признаков	А	ПА
59.	Метаразметка необходима для: А. выявления взаимосвязи языка и условий его существования, а также для отбора и изучения отдельных подмножеств языка В. проведения семантического анализа С. проведения лексикографического анализа Д. описания стилистических особенностей лексики	А	ПА
60.	Дискурсивная разметка А. служит для обозначения пауз, повторов, оговорок В. фиксирует библиографические и типологические характеристики С. приписывает словам семантическую	А	ПА

	категорию D. раскрывает характер синтаксических связей между единицами		
61.	Анафорическая разметка A. учитывает референтные связи, устанавливает связи между номинациями и их объектами B. фиксирует библиографические и типологические характеристики C. приписывает словам семантическую категорию D. раскрывает характер синтаксических связей между единицами	A	ПА
62.	Просодическая разметка A. указывает ударение и интонацию B. фиксирует библиографические и типологические характеристики C. приписывает словам семантическую категорию D. раскрывает характер синтаксических связей между единицами	A	ПА
63.	Парсинг – A. выстраивание синтаксических зависимостей B. вычленение в потоке речи значимых единиц C. погрешность в конечных результатах анализа D. обнаружение основы слова, благодаря которому прослеживаются принципы словообразования сходных слов	A	ПА
64.	Токенизация – A. вычленение в потоке речи значимых	A	ПА

	единиц (словоформ) В. погрешность в конечных результатах анализа С. обнаружение основы слова, благодаря которому прослеживаются принципы словообразования сходных слов Д. процесс приписывания любой форме слова его начальной формы		
65.	Стемминг – А. обнаружение основы слова, благодаря которому прослеживаются принципы словообразования сходных слов В. процесс приписывания любой форме слова его начальной формы С. вычленение в потоке речи значимых единиц Д. погрешность в конечных результатах анализа	А	ПА
66.	Лемматизация – А. процесс приписывания любой форме слова его начальной формы В. вычленение в потоке речи значимых единиц С. погрешность в конечных результатах анализа Д. обнаружение основы слова, благодаря которому прослеживаются принципы словообразования сходных слов	А	ПА
67.	Классификация корпусов по цели: А. многоцелевые и специализированные В. динамические и статистические С. свободно доступные и коммерческие Д. полнотекстовые и «фрагментотекстовые»	А	ПА
68.	Классификация корпусов по типу языковых данных: А. письменные, устные (речевые), смешанные В. многоцелевые и специализированные С. динамические и статистические Д. свободно доступные и коммерческие	А	ПА

69.	Классификация корпусов по «литературности»: А. литературные, диалектные, разговорные, терминологические, смешанные В. письменные, устные (речевые), смешанные С. многоцелевые и специализированные Д. динамические и статистические	А	ПА
70.	Классификация корпусов по жанру: А. литературные, фольклорные, драматургические, публицистические В. письменные, устные (речевые), смешанные С. многоцелевые и специализированные Д. динамические (мониторные) и статистические	А	ПА
71.	Классификация корпусов по назначению: А. исследовательские, иллюстративные В. размеченные, неразмеченные С. многоцелевые и специализированные Д. динамические и статистические	А	ПА
72.	Классификация корпусов по динамичности: А. динамические (мониторные) и статистические В. размеченные, неразмеченные С. многоцелевые и специализированные Д. полнотекстовые и «фрагментотекстовые»	А	ПА
73.	Классификация корпусов по разметке: А. размеченные, неразмеченные В. многоцелевые и специализированные С. динамические и статистические Д. свободно доступные и коммерческие	А	ПА
74.	Классификация корпусов по характеру разметки: А. морфологические, синтаксические, семантические, анафорические, просодические и т.д. В. свободно доступные, коммерческие, закрытые С. полнотекстовые и «фрагментотекстовые» Д. многоцелевые и специализированные	А	ПА

75.	Классификация корпусов по доступности: А. свободно доступные, коммерческие, закрытые В. полнотекстовые и «фрагментотекстовые» С. многоцелевые и специализированные D. динамические и статистические	А	ПА
76.	Классификация корпусов по объему текстов: А. полнотекстовые и «фрагментотекстовые» В. многоцелевые и специализированные С. динамические и статистические D. свободно доступные и коммерческие	А	ПА
77.	Исследовательские корпуса создаются с целью: А. изучения различных аспектов функционирования языка В. для описания лингвокогнитивных особенностей лексики С. для описания стилистических особенностей D. для описания морфологических особенностей	А	ПА
78.	Иллюстративные корпуса создаются с целью: А. подтвердить и обосновать уже полученные результаты В. изучить различные аспекты функционирования языка С. описать стилистические особенности D. описать морфологические особенности	А	ПА
79.	Иллюстративные корпуса служат: А. для выделения из них показательных лингвистических примеров, подтверждающих те или иные языковые (речевые, текстовые) факты В. для подтверждения и обоснования уже полученных результатов С. для изучения различных аспектов функционирования языка D. для описания стилистических особенностей	А	ПА
80.	Примером статического корпуса является: А. авторский корпус - коллекция текстов	А	ПА

	писателей В. Национальный корпус русского языка (НКРЯ) С. Британский национальный корпус (BNC) Д. корпус современного американского языка (COCA)		
81.	Примером неограниченного (постоянно развивающегося) мониторингового корпуса является: А. корпус современного американского языка (COCA) В. авторский корпус - коллекция текстов писателей С. корпусный менеджер корпуса НКРЯ Д. Британский национальный корпус (BNC)	А	ПА
82.	Мониторные (постоянно развивающиеся) корпуса: А. играют огромную роль в строении словарей, поскольку позволяют лексикографам отслеживать новые слова, проникающие в язык В. помогают при проведении различных диахронических исследований С. служат для изучения семантических особенностей Д. служат для проведения морфологического анализа	А	ПА
83.	Динамические корпуса предназначены для: А. проведения различных диахронических исследований В. изучения семантических особенностей С. проведения морфологического анализа Д. проведения лексикографического анализа	А	ПА
84.	Параллельные корпуса используются для: А. разработки эффективных методов перевода, в том числе машинного В. изучения семантических особенностей С. проведения морфологического анализа Д. проведения лексикографического анализа	А	ПА
85.	Параллельные корпуса используются для:	А	ПА

	А. составления двуязычных и многоязычных терминологических словарей В. изучения семантических особенностей С. проведения морфологического анализа Д. проведения лексикографического анализа		
86.	Параллельные корпуса используются для: А. сравнительных исследований языков в области лексикологии, грамматики, стилистики, диалектологии, переводоведения В. изучения семантических особенностей С. проведения морфологического анализа Д. проведения лексикографического анализа	А	ПА
87.	С помощью параллельных корпусов можно: А. переводить терминологические и фразеологические единицы текста В. проводить диахронические исследования С. изучать семантические особенности Д. проводить морфологический анализ	А	ПА
88.	С помощью параллельных корпусов можно: А. устранять полисемию лексических единиц путем компьютерного анализа контекста многозначного слова В. проводить диахронические исследования С. изучать семантические особенности Д. проводить морфологический анализ	А	ПА
89.	С помощью параллельных корпусов можно: А. формировать семантические поля и терминологические системы В. проводить диахронические исследования С. изучать семантические особенности Д. проводить морфологический анализ	А	ПА
90.	С помощью параллельных корпусов можно: А. изучать универсалии перевода В. проводить диахронические исследования С. изучать грамматические особенности Д. проводить морфологический анализ	А	ПА

91.	При исследовании параллельных корпусов могут успешно применяться: А. инструменты автоматической классификации лексики В. автоматическая разметка С. ручная разметка Д. использование корпусного менеджера	А	ПА
92.	Лингвистические корпуса незаменимы для: А. выявления разнообразных количественных характеристик и исследования частотности любых языковых явлений В. проведения диахронического исследования С. изучения семантических особенностей Д. проведения лексикографического анализа	А	ПА
93.	Частотность фиксирует: А. количество употреблений определённой единицы на заданный объём других единиц В. грамматические особенности лексической единицы С. стилистические особенности лексической единицы Д. морфологические особенности лексической единицы	А	ПА
94.	Выровненный параллельный корпус – это корпус, где тексты на одном языке и тексты на другом строго соотнесены: А. по фразам и предложениям В. по семантическим особенностям С. по грамматическим особенностям Д. по морфологическим особенностям	А	ПА
95.	Лемма – А. начальная форма слова, приписанная словоформе в результате лемматизации В. начальная форма слова С. графическая единица с учетом лексического значения Д. графическая единица с учетом стилистических особенностей	А	ПА

96.	Парсер – А. компьютерная программа для морфологического и синтаксического анализа слов В. компьютерная программа для морфологического анализа слов С. компьютерная программа для синтаксического анализа слов Д. корпусный менеджер	А	ПА
97.	Аннотированный корпус – А. корпус, в котором произведена разметка данных по одному или нескольким признакам В. корпус, который сопровождается корпусным менеджером С. корпус, в котором произведена разметка данных по одному признаку Д. неразмеченный корпус	А	ПА
98.	Снятие омонимии – А. выбор правильного прочтения слова, верное определение его значения в ситуации, когда потенциально возможно более чем одно прочтение В. выбор грамматических характеристик слова С. определение стилистических характеристик слова Д. процедура разметки лингвистического корпуса	А	ПА
99.	Стандарты кодирования – А. приемы унификации разметки корпусов, используемые для возможности их правомерного сопоставления В. правила чтения слов С. принципы функционирования корпусного менеджера Д. правила использования лингвистического корпуса	А	ПА
100.	Стем – А. основа слова, выявленная в результате стемминга	А	ПА

	В. корень слова С. значимая часть слова Д. грамматическая основа		
103	Токен – А. минимальный фрагмент в тексте, подлежащий анализу, выявленный в результате токенизации В. значимая часть слова С. грамматическая основа Д. основа слова, выявленная в результате стемминга	А	ПА
102	Тег – А. условное обозначение морфологического признака, характера реализации или проявления определенной грамматической категории конкретной словоформы, выявленной в процессе лемматизации В. основа слова, выявленная в результате стемминга С. минимальный фрагмент в тексте, подлежащий анализу, выявленный в результате токенизации Д. грамматическая основа	А	ПА
101	Коллокация – А. устойчивое сочетание слов в корпусе В. идиома С. грамматическая основа Д. основа слова	А	ПА
104	Ранговое распределение – А. расположение в порядке убывания характеристик (значений) некоторой величины В. учет стилистических характеристик С. количество употреблений единицы на миллион слов Д. коллокации	А	ПА
105	Общая частота – А. количество употреблений единицы на	А	ПА

	миллион слов В. расположение в порядке убывания характеристик (значений) некоторой величины С. средняя частота по всему корпусу Д. количество сегментов, на которые разбит корпус		
--	---	--	--

ВОПРОСЫ ОТКРЫТОГО ТИПА

№	вопрос	эталон ответа	уровень применения
1.	лингвистическая дисциплина, занимающаяся разработкой общих принципов построения и использования текстовых корпусов с применением компьютерных технологий	корпусная лингвистика	ПА
2.	большой, структурированный, размеченный массив языковых данных в машиночитаемом формате	корпус	ПА
3.	свойство корпуса адекватно представлять язык или его вариант, для описания которого он создавался	репрезентативность	ПА
4.	свойство корпуса, достигаемое включением текстов разных жанров, типов и тематик в пропорциях, отражающих их реальное соотношение в языке	сбалансированность	ПА
5.	представление текста в электронном виде, пригодном для автоматической обработки компьютером	машиночитаемый формат	ПА
6.	процесс добавления в тексты корпуса металингвистической информации (морфологической, синтаксической и т.д.)	разметка (аннотация)	ПА
7.	добавление информации о тексте в целом (метаданные), например, об авторе, годе создания, жанре	метаразметка	ПА
8.	элементарная единица текста (слово, знак препинания, символ) в компьютерной обработке	токен	ПА
9.	уникальная словоформа или лексема в тексте, в отличие от	тип	ПА

№	вопрос	эталон ответа	уровень применения
	конкретного употребления (токена)		
10.	список всех употреблений искомого слова или конструкции в контексте, обычно выровненный по центру	конкорданс	ПА
11.	слова, частота которых в данном тексте или корпусе статистически значимо выше по сравнению с другим, более общим корпусом	ключевые слова	ПА
12.	корпус, представляющий язык на определенном этапе (или в истории) и стремящийся к максимальной полноте и репрезентативности (например, Национальный корпус русского языка)	национальный корпус	ПА
13.	корпус, состоящий из текстов на одном языке и их переводов на один или несколько других языков	параллельный корпус	ПА
14.	мультязычный корпус, содержащий как оригинальные тексты, так и их переводы для всех представленных языков	двусторонний параллельный корпус	ПА
15.	корпус текстов на двух или более языках, подобранных по сходным критериям (жанр, время, тематика), но не являющихся переводами друг друга	сопоставимый корпус	ПА
16.	корпус, включающий транскрипции устной речи	устный корпус	ПА
17.	корпус, включающий тексты из разных исторических периодов, и позволяющий изучать изменения в языке во времени	диахронический корпус	ПА
18.	динамический корпус, который постоянно пополняется новыми текстами	мониторный корпус	ПА
19.	корпус текстов, созданных изучающими язык	учебный корпус	ПА
20.	корпус текстов определенного жанра, автора или тематики	специализированный корпус	ПА

№	вопрос	эталон ответа	уровень применения
21.	поисковая система, программное обеспечение для управления корпусом, поиска и извлечения информации из него	корпусный менеджер	ПА
22.	программа для создания конкордансов	конкордансер	ПА
23.	программа для установления соответствий (выравнивания) между предложениями или словами в параллельных корпусах	выравниватель	ПА
24.	веб-страница или программа, через которую пользователь взаимодействует с корпусом, задает запросы и получает результаты	интерфейс корпуса	ПА
25.	процесс присвоения каждому слову в тексте метки с указанием его части речи и грамматических характеристик	морфологическая разметка	ПА
26.	добавление информации о синтаксической структуре предложения (например, в виде дерева зависимостей или составляющих)	синтаксическая разметка (парсинг)	ПА
27.	добавление информации о значении слов (например, семантические классы, тематические группы)	семантическая разметка	ПА
28.	разметка, отражающая структуру дискурса (связи между предложениями, тема-рематические отношения)	дискурсивная разметка	ПА
29.	метка, присваиваемая единице текста в процессе разметки	тэг (tag)	ПА
30.	метка, присваиваемая единице текста в процессе разметки	тэгсет	ПА
31.	количественная характеристика, показывающая, сколько раз данная единица встречается в корпусе	частота	ПА

№	вопрос	эталон ответа	уровень применения
32.	статистический показатель, используемый для оценки силы связи между словами (коллокатами)	мера ассоциации	ПА
33.	одна из наиболее распространенных мер ассоциации	T-score	ПА
34.	популярная мера ассоциации, показывающая, насколько часто два слова встречаются вместе по сравнению со случайностью	взаимная информация (MI, Mutual Information)	ПА
35.	устойчивое сочетание слов, статистически значимая связь двух или более слов	коллокация	ПА
36.	непрерывная последовательность из n элементов (слов, символов) в тексте (например, биграмма — последовательность из двух слов)	N-грамма	ПА
37.	частота, приведенная к стандартному объему текста (например, на миллион слов) для корректного сравнения текстов разного размера	нормализованная частота	ПА
38.	характеристика встречаемости языковой единицы по разным частям корпуса (жанрам, периодам, авторам)	распределение	ПА
39.	словарная форма слова (для существительных — именительный падеж, для глаголов — инфинитив)	лемма	ПА
40.	конкретная грамматическая форма слова (например, "столом", "иду")	словоформа	ПА
41.	представление искомого слова в окружении нескольких слов слева и справа, как в конкордансе	контекст (KWIC)	ПА
42.	изменение значения слова, которое можно отследить с помощью диахронического корпуса	семантический дивиг	ПА
43.	первый в мире компьютерный корпус текстов (1960-е гг.), положивший начало современной корпусной лингвистике	Брауновский корпус	ПА

№	вопрос	эталон ответа	уровень применения
		(Brown Corpus)	
44.	британский аналог Брауновского корпуса	корпус LOB (Lancaster-Oslo-Bergen Corpus)	ПА
45.	один из первых крупных национальных корпусов английского языка	Британский национальный корпус (BNC)	ПА
46.	крупнейший и наиболее часто используемый мониторинговый корпус английского языка	Корпус современного американского английского языка (COCA)	ПА
47.	более широкая область, включающая корпусную лингвистику, машинный перевод и другие направления автоматической обработки языка	компьютерная лингвистика	ПА
48.	общее название для компьютерных методов анализа и синтеза текстов	автоматическая обработка текста	ПА
49.	наука о составлении словарей, которая исторически тесно связана с корпусными методами	лексикография	ПА
50.	требование к создателям корпуса о тщательной, квалифицированной подготовке текстов	филологическая компетентность	ПА