

## Тест

### Выберите один или несколько верных вариантов ответа.

1. Тексты, составляющие лингвистический корпус,
  - ☐ организованы для удобства чтения.
  - ☐ сконструированы лингвистами.
  - ☐ не включают речь необразованных носителей языка.
  - ☒ размечены лингвистами и программистами.
2. В лингвистический корпус могут входить
  - ☒ тексты электронной коммуникации.
  - ☒ записи устной речи детей.
  - ☒ транскрипции теле- и радиопередач.
  - ☒ произведения художественной литературы.
3. Для того, чтобы электронная коллекция текстов могла считаться корпусом,
  - ☐ в ней не должно быть текстов, посвященных неактуальным для общества проблемам.
  - ☐ она должна отражать состояние языка в определенный период времени.
  - ☒ она должны быть размечена.
  - ☒ в ней должны быть исправлены ошибки носителей языка.
4. Обязательными свойствами хорошего лингвистического корпуса являются
  - ☒ представительность.
  - ☒ присутствие разметки.
  - ☐ снятая грамматическая омонимия.
  - ☒ баланс между разными типами текстов.
5. Создавая репрезентативный корпус, лингвисты должны ответить на вопросы:
  - ☒ Какие тексты считать значимыми для данной культуры?
  - ☒ Отражают ли оригинальные тексты художественной литературы состояние языка эпохи?
  - ☒ Можно ли, изучая язык периода или жанра, опираться на переводные тексты?
  - ☒ 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 «\_\_\_\_\_».
9. Нормализация значений обычно предполагает вычисление отношения количества вхождений к объему корпуса, умноженного на \_\_\_\_\_.
10. Пропорциональное представление в корпусе текстов разных типов называется «\_\_\_\_\_» корпуса.

## Выберите единственный правильный вариант ответа.

11. Что означает понятие «накопительная сбалансированность»?
- ☒ В большом корпусе в некоторых случаях можно пренебречь представлением текстов разных типов в соотношениях, сопоставимых с соотношениями в естественном языке.
  - ☐ С увеличением числа текстов, входящих в корпус, увеличивается необходимость разнообразить состав включаемых в него документов.
  - ☐ Лингвистический корпус может быть относительно небольшим, важно, чтобы в нем были представлены разные подязыки в пропорциях, соответствующих аналогичным пропорциям в естественном языке.
  - ☐ Понятия «сбалансированность» и «репрезентативность» никак не связаны; хороший корпус должен обладать обеими характеристиками.
12. Что отличает разметку Обучающего корпуса от аннотации других частей НКРЯ?
- ☐ В ней присутствует синтаксическая разметка.
  - ☒ Она проведена в соответствии со школьной грамматикой.
  - ☐ Она позволяет вести поиск по речевым действиям.
  - ☐ Она целиком произведена вручную.
13. Чем отличается лингвистический корпус от электронного словаря?
- ☐ Электронный словарь не предлагает контекстов употребления слова.
  - ☒ В лингвистическом корпусе языковая норма отражена только опосредованно.
  - ☐ Материалы лингвистического корпуса не дают представления о том, насколько слово распространено в языке.
  - ☐ В электронном словаре могут содержаться сведения о семантике слова.
14. Какой вид поиска невозможен для пользователя в поисковых системах Яндекс и Google?

- ☐ Поиск в пределах одного сайта.
- ☐ Поиск точных форм.
- ☐ Поиск словоформ одной лексемы.
- ☒ Поиск смысловых категорий.

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

- ☐ Яндекс Переводчик
- ☒ Google Books Ngram Viewer
- ☐ ABBYY LINGVO
- ☐ Google Translator Toolkit