Выберите один или несколько верных вариантов ответа.

1. Тексты, составляющие лингвистический корпус, организованы для удобства чтения, сконструированы лингвистами, размечены лингвистами и программистами.

2. В лингвистический корпус могут входить тексты электронной коммуникации, записи устной речи детей, произведения художественной литературы.

3. Для того, чтобы электронная коллекция текстов могла считаться корпусом, она должна отражать состояние языка в определенный период времени, она должны быть размечена.

4. Обязательными свойствами хорошего лингвистического корпуса являются представительность, присутствие разметки, 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 раздел корпуса, по которому велся поиск.

7. Согласно материалам лекции, основная часть Национального корпуса русского языка включает более 600 млн. слов.

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

9. Нормализация значений обычно предполагает вычисление отношения количества вхождений к объему корпуса, умноженного на 1000000.

10. Пропорциональное представление в корпусе текстов разных типов называется «*репрезентативностью*» корпуса.

Выберите единственный правильный вариант ответа.

11. Что означает понятие «накопительная сбалансированность»? Понятия «сбалансированность» и «репрезентативность» никак не связаны; хороший корпус должен обладать обеими характеристиками.

12. Что отличает разметку Обучающего корпуса от аннотации других частей НКРЯ? Она проведена в соответствии со школьной грамматикой.

13. Чем отличается лингвистический корпус от электронного словаря? Электронный словарь не предлагает контекстов употребления слова.

14. Какой вид поиска невозможен для пользователя в поисковых системах Яндекс и Google?Поиск словоформ одной лексемы.

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