Тест

**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

при помощи разметки создатели корпуса передают пользователю свои теоретические установки.

прежде, чем включать тексты в корпус, лингвисты тщательно изучают их и делают выводы об их содержании.

начиная работу в корпусе, пользователь по умолчанию соглашается принять на веру принципы, положенные в основу аннотации.

каждое языковое явление должно быть интерпретировано создателями корпуса как правильное или неправильное и соответствующим образом помечено.

2. Аннотирование – это

обработка текста в целях облегчения его изучения.

необходимый этап создания лингвистического корпуса.

присвоение каждому элементу текста особого тэга.

сведения, размещенные на главной странице корпуса.

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

разметчики описывают характеристики, присущие элементам текста.

принципы разметки – это внутренняя информация, важная только для создателей корпуса.

теоретические установки, положенные в основу аннотации, должны быть максимально нейтральными.

разметка не выводится на экраны пользователей корпуса.

4. Автоматическая разметка

начинается с деления текста на минимальные фрагменты.

невозможна на уровне семантики.

предполагает отнесение словоформы к начальной форме.

позволяет быстро провести аннотирование по частям речи.

5. Экстралингвистическая разметка

предполагает демонстрацию сведений об авторе.

не затрагивает содержания текста.

позволяет вести поиск в отобранной по социолингвистическим параметрам группе текстов.

в НКРЯ выражается в виде «паспорта текста».

6. Морфологическая разметка в НКРЯ

произведена с помощью компьютерных программ.

началась с выработки командой лингвистов стандартных решений в спорных ситуациях.

единообразна во всех частях НКРЯ.

предполагает присвоение тэгов каждой единице текста.

7. Синтаксическая разметка в НКРЯ НЕ

проведена в соответствии с традиционным выделением членов предложения.

графически представлена в виде дерева зависимостей.

проведена компьютерной программой, затем доработана вручную.

охватывает все части НКРЯ.

8. Семантическая разметка в НКРЯ

помечает семантические признаки грамматических разрядов.

помечает семантические признаки словообразовательных показателей.

не затрагивает собственно семантические признаки.

предполагает присвоение одной лемме одного семантического тэга.

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

для корпуса МУРКО в составе НКРЯ были разработаны дополнительные принципы описания текстов, например, создана разметка значений жестов.

элементы корпуса, представляющего визуальные образы, должны быть описаны иначе, чем элементы письменного или аудио текста.

мультимодальные корпуса должны быть частично закрытыми из соображений этики.

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

одной лемме.

двум леммам.

трем леммам.

четырем леммам.

**Впишите слово.**

11. В качестве минимального элемента разметки в НКРЯ принята \_\_\_\_\_\_\_\_, то есть графическая единица, не обладающая, в отличии от лексемы, 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 \_\_\_\_\_\_\_\_.