

Время выполнения заданий: 150 минут

**Пишите разборчиво. Кроме решений задач в работе не должно быть никаких пометок. При отсутствии решения ставьте прочерк.
(Максимальное количество баллов — 100)**

Задача 1 (30 баллов)

Для кодирования грамматической информации в компьютерной лингвистике применяются специальные системы помет, которые представляют лингвистические сведения в стандартизированной форме. Одна из таких систем, предназначенных, в частности, для русского языка, носит название MULTEXT-East. Даны некоторые пометы, используемые в этой системе для разметки существительных; к каждой помете приводится по примеру:

Ncmprau — *участников*

Ncfsnn — *поддержка*

Npmsgy — *Владимира*

Ncnsin — *удовольствием*

Ncmpln — *пределах*

Ncfply — *уборщицах*

Ncmsan — *режим*

Npfsay — *Юлию*

Ncfpdn — *группам*

1.1. Схематично опишите, как устроена система помет существительных в системе MULTEXT-East.

1.2. Некоторые из девяти слов, данных в условии, могли бы иметь в системе системы MULTEXT-East и другие пометы. Укажите эти слова и приведите эти пометы.

1.3. Как будут выглядеть в системе MULTEXT-East пометы существительных *Иванову, животные, гордячкой*? Если в каком-то случае возможны несколько вариантов, укажите их все.

Задача 2 (35 баллов)

Некоторые глаголы русского языка (например: *осознавать, мёрзнуть, чихать*) обладают неким свойством (элементом смысла), которое делает невозможным или очень необычным их употребление в ряде конструкций (такую невозможность принято отмечать звёздочкой). В частности:

1) В инфинитивных предложениях: **А Петя ну осознавать свои ошибки! *А Маша давай мёрзнуть!* Хотя возможно: *А Петя ну бежать! А Маша давай петь!*

2) В безличных предложениях с субъектом в дательном падеже и сказуемым в возвратной форме: **Мне с трудом осознавалось. *Пете хорошо чихается на улице. Хотя возможно: Сегодня мне бежалось гораздо легче. Пете хорошо работается на новом месте.*

3) В настоящем времени в значении запланированного будущего: **Завтра Петя осознаёт всю тяжесть своего положения. *На следующей неделе мёрзну по 2 часа в день. Хотя возможно: Завтра Маша бежит 3 км. На следующей неделе работаю по 8 часов в день.*

4) В комбинации с обстоятельствами цели (с предложениями *ради, для*): **Петя осознаёт это для неё. *Она чихает ради справки. Хотя возможно: Он работает ради зарплаты. Маша поёт для него.*

2.1. Объясните, что это за свойство.

2.2. В каких случаях глаголы, обладающие этим свойством, всё же могут употребляться в указанных конструкциях?

2.2. Приведите ещё 2 глагола, обладающих рассматриваемым свойством.

2.3. В каких ещё конструкциях глаголы *осознавать, мёрзнуть, чихать* и им подобные ведут себя отлично от остальных глаголов (не обладающих рассматриваемым свойством)? Для каждой найденной вами конструкции приведите по 2 примера.

Задача 3 (35 баллов)

В Национальном корпусе русского языка содержатся тексты общим объёмом более 500 млн слов. В нём можно осуществлять разные поисковые запросы. Ниже дана верхняя часть списка, полученная по некоторому запросу. Одно слово при этом заменено на X.

- 1 9333 *известно* X
- 2 9038 *ясно* X
- 3 6198 *видно* X
- 4 5663 *хорошо* X
- 5 4875 *понятно* X
- 6 4382 *очевидно* X
- 7 3063 *возможно* X
- 8 2565 *равно* X
- 9 2179 *естественно* X
- 10 2175 *интересно* X
- 11 1696 *мало* X

- 12 1475 *странно* X
13 1284 *вероятно* X
14 1264 *похоже* X
15 1236 *несомненно* X

3.1. Как был сформулирован запрос?

3.2. Что за цифры приведены во втором столбце списка?

3.3. Какое слово заменено на X?

3.4. Приведите ещё 3 примера слов, которые, скорее всего, тоже оказались высоко в этом списке, и 3 примера слов, которые оказались в его нижней части или вовсе отсутствуют в нём.

3.5. Некоторые из вышеуказанных 15 слов оказались в списке по иной причине, чем остальные. Назовите эти слова и объясните причины, по которым они оказались в верхней части списка.