

### Задание 1 (4 балла)

В этом задании вам предлагается провести корпусное исследование с опорой на Национальный корпус русского языка. Каждый раз, когда в задании о запросе к корпусу, приводите запрос настолько точно, чтобы можно было воспроизвести.

В русском языке можно обнаружить две конструкции с (практически) одинаковым значением:

- «слово в сравнительной степени + *обычного*» (пример (1)),
- «слово в сравнительной степени + *обыкновенного*» (пример (2)).

(1) *Она выглядит лучше обычного, как-то вся светится радостью.* [Н. П. Каманин. Дневник (1963)]

(2) *Старик изредка взглядывал на него, чаще обыкновенного взмахивал головой, но ничего не говорил.* [П. П. Бажов. Через межу (1934)]

Описанные конструкции примечательны тем, что *обычный* и *обыкновенный*, хотя и являются прилагательными, с точки зрения значения скорее сопоставляются с обстоятельствами. Предложение (1) можно понять как (1)', а не как (1)".

(1)' *Она выглядит лучше, чем обычно, как-то вся светится радостью.*

(1)" *Она выглядит лучше, чем обычный человек, как-то вся светится радостью.*

Прилагательное в такой конструкции всегда стоит в родительном падеже и среднем роде (совпадает здесь с мужским) и может не соотноситься по смыслу ни с каким существительным среднего или мужского рода, как в примерах (1) и (2).

1.1. (2 балла) Составьте для каждой конструкции запрос, который позволяет найти все её употребления в корпусе.

- а. Приведите эти запросы. Они могут быть сформулированы разными способами — достаточно привести один любой для каждой конструкции.
- б. Вероятно, в результате запросов будут найдены не только подходящие конструкции, но и другие случаи. Коротко опишите их и приведите примеры.

В дальнейших пунктах не уточняйте запрос, чтобы исключить такие лишние случаи, а считайте, что их случаев слишком мало, чтобы они повлияли на результат.

1.2. (1 балл) Когда в языке есть две единицы, которые употребляются в одном и том же контексте, всегда хочется найти между ними различия. В случае рассматриваемых конструкций, корпус, по-видимому, не позволяет найти разницу в их значении. Он однако даёт возможность сформулировать другую закономерность, связанную с их историческим развитием и объясняющую существование двух конструкций с одинаковым значением.

- а. В каком году у каждой из конструкций наибольшая частотность (без сглаживания).
- б. Сформулируйте историческую закономерность, о которой идёт речь в условии. Сделайте предположение об их дальнейшем развитии с точки зрения допустимости в языке.

1.3.\* (1 балл) Описанная вами историческая закономерность в развитии двух конструкций, по всей видимости, не является результатом исторической случайности, а может получить объяснение, связанное с историческим развитием самих слов *обычный* и *обыкновенный*. Предложите такое объяснение.

## Задание 2 (2 балла)

Перед вами [таблица](#), содержащая метаинформацию о русских стихотворениях.

- а. (1 балл) Укажите, какой автор в этой таблице написал стихотворение с самым длинным заголовком и длину этого заголовка.
- б. (1 балл) Какая средняя длина у заголовков 1800–1849 и 1900–1949 годах? В какой период времени заголовки в среднем длиннее?

Произведите предварительную очистку данных, если это необходимо. Считайте длину в символах. Приложите к ответу файл или ссылку на файл, в котором видно, как получен результат, в том числе промежуточные шаги. Можете использовать любой удобный вам инструмент для выполнения задания.

## Задание 3 (2 балла)

3.1. (0.5 балла) Выберите произведение крупной формы на русском языке, которое представляет для вас интерес. Можете взять сборник произведений короткой формы. Найдите выбранное произведение в интернете и скачайте в формате .txt или переведите в соответствующий формат. Приведите ссылку на этот файл (если вы его изменяли, загрузите соответствующую версию на диск / в файлообменник и приложите на него ссылку).

Не забудьте вписать название произведения в [таблицу](#), чтобы не повторяться.

3.2. (0.5 балла) Составьте частотный список слов этого произведения, приведите 10 наиболее часто использующихся. В нескольких словах скажите, насколько репрезентативны эти частотные слова по отношению к этому произведению.

3.3. (1 балл) Составьте список ключевых выбранного произведения. Используйте [SinTagRus](#) в качестве Reference corpus и параметры Log-Likelyhood (4-term),  $p < 0.05$  (+ Bonferroni), Dice (параметры по умолчанию, если вы пользуетесь AntConc, но стоит проверить их в настройках на всякий случай).

- а. Приведите 10 наиболее важных ключевых слов.
- б. Кратко проанализируйте, что найденные ключевые слова говорят о произведении.

## Задание 4\* (1 балл)

Ознакомьтесь с [материалами](#), посвящёнными регулярным выражениям. Составьте регулярное выражение, которое ищет заданные формы прилагательного, указанного для вас на [странице](#). Тестировать регулярное выражение можно в гугл-документах или на [сайте](#).

Может ли это регулярное выражение найти ещё какие-то последовательности символов? Если да, приведите любой пример, если нет, просто укажите это.