

## [Ссылка на таблицу](#)

### Бонус

На этом листе в сводной таблице отражена частотность каждого из прилагательных (отдельно для прилагательных с диминутивным суффиксом и прилагательных без), в том числе частотность коллокаций с существительными.

Помимо абсолютной частоты OCCURANCE и количества документов, в которых была найдена коллокация, я рассчитал относительную частотность REL.FREQUENCY (равно отношению абсолютной частотности к общему количеству вхождений всех коллокаций) и метрику, показывающую частотность коллокации на документ (рассчитанную как отношение OCCURANCE к DOCUMENTS). Чем выше значение OCCUR. PER DOCUMENT, тем более часто встречается слово. Если значение равно единице или меньше -- слово редкое и встречается эксклюзивно и не повторяется в одном тексте. Для REL.FREQUENCY и OCCUR. PER DOCUMENT данные заполняются средним значением, а не их суммой. То есть для каждого прилагательного отображается его средняя относительная частотность и среднее вхождение в документ в составе всех имеющихся коллокаций.

Среди прилагательных с диминутивными суффиксами самым частотным (REL.FR) является 'аленький' (при этом связанный с 'цветочек') -- это легко объяснить наличием сказки с таким названием. На втором месте -- 'седенький', что можно объяснить попыткой смягчить потенциально обидное высказывание. Среди прилагательных без этого суффикса чаще всего встречаются достаточно устойчивые сочетания вроде 'голубые глаза', 'серые глаза', 'карие глаза', а также 'черные глаза/хлеб/волосы/человек (например, стихотворение Есенина)' и 'белый дом'. Заметим, что многие из этих коллокаций существуют зачастую во множественном числе (карие глаза, но не карий глаз; черные волосы, но не черный волос).

### Бонус 2(')

Бонус 2 -- исходная таблица для 2', в которой обе таблицы из ДЗ сведены в одну, при этом каждой присвоен тег диминутивности или базовости, в отдельный столбец вынесены базовые формы (с помощью регулярок). На листе Бонус 2' отображена частотность сочетания прилагательных с диминутивными суффиксами и без, а также существительных с каждым из диминутивных суффиксов по отдельности.

Рассмотрим те прилагательные, которые в корупсе встречались и с суффиксами, и без. Выделим бирюзовым те, которые чаще употреблялись с суффиксом, и те, что без -- фиолетовым. Интересно, что прилагательные, обозначающие темные цвета -- зеленый, красный, черный -- чаще употребляются без диминутивного суффикса, а светлые -- белый, голубой, желтый -- с ним.

Во второй таблице сочетаемости сущ. с суффиксами выделим красным те, у которых процент сочетаемости с суффиксами меньше 50. Можно предположить, что не выделенные красным слова чаще могут иметь уменьшительные формы благодаря их семантике. Например, многие из них обозначают живых существ ('заяц', 'собака'), предметы, которые относятся к быту и окружению, имеющие свойства, позволяющие создавать уменьшительные версии (например, 'книга' -> 'книжка', 'платье' -> 'платьице').