**Бонусное задание (диминутивы и цветовые прилагательные)**

Для того, чтобы объяснить распределение вариантов с диминутивами, следует выдвинуть некоторое количество гипотез, зависящих от разных параметров, чтобы далее проверить их.

*Гипотеза 1.* Распределение суффиксов диминутивов у существительных при цветовых прилагательных будет отличаться от распределения суффиксов диминутивов в корпусе, что покажет влияние конкретных суффиксов на сочетаемость с цветовыми прилагательными.

Для проверки данной гипотезы нами был задан запрос по всем диминутивам в основном подкорпусе НКРЯ, далее была составлена рандомная выборка из 200 вхождений, откуда были вручную удалены некорректно размеченные примеры (например, «бабушка», т. к. в СРЛЯ оно семантически не является диминутивом от «бабка»). Далее из оставшихся 175 вхождений была составлена сводная таблица по суффиксам (см. лист «б\_статистика по суфф. дим. (корп)»). Сравнение топ-суффиксов по таблицам показывает, что существенным различием в них является довольно высокое положение суффикса -ек в выборке с цветовыми прилагательными (отношение топ-1 к -ек в общей выборке = 30,3; в нашей выборке = 1,4). Можно было бы предположить, что суффикс -ек является частотным для сочетания с цветовыми прилагательными, но если обратить внимание на биграммы, в составе которых он встречается, то можно увидеть, что половина употреблений суффикса -ек в нашей выборке приходится на биграмму «аленький цветочек». Такое распределение должно иметь скорее экстралингвистическое объяснение, т. к. «аленький цветочек» является названием известной сказки С. Т. Аксакова, после которой и в контексте которой, видимо, и распространилось данное сочетание. Такие результаты позволяют нам сделать вывод, что суффикс диминутива у существительных, по-видимому, не влияет на сочетаемость с цветовыми прилагательными.

*Гипотеза 2.* Семантика существительных с показателями диминутива будет влиять на их сочетаемость с цветовыми прилагательными.

Для проверки данной гипотезы была составлена сводная таблица, ранжированная по убыванию числа вхождений (occurrence) частотности диминутивных существительных с цветовыми прилагательными в корпусе (см. лист «б\_частотность дим сущ»). Далее для всех существительных в таблице был определён их семантический класс (по самостоятельно введённой классификации). В результате семантической разметки оказалось, что в выборку входит несколько неплохо репрезентированных семантических групп: body part (9 лексем, сумм. 182 вхождения), person (3 лексемы, сумм. 135 вхождений), clothes (4 лексемы, сумм. 80 вхождений), animal (5 лексем, сумм. 66 вхождений) (всего в выборке 36 различных лексем, 827 вхождений). Таким образом, на данной выборке можно выделить несколько семантических групп, которые сочетаются с диминутивными прилагательными чаще представителей других семантических групп.

Также интересное направление исследования полученных данных – сопоставление пар н-грамм, которые различаются только диминутивными показателями. В представленной выборке было найдено 25 таких пар. Интересно, что если посчитать сумму количества документов для таких пар (в данном случае лучше считать сумму по документам, чтобы сочетания, которые часто встречаются в малом количестве документов, не сбивали статистику), то получится, что в топ-10 попадает 5 сочетаний вида «цветовое прилагательное + глаз». Если провести семантическую разметку для таких сочетаний, мы также обнаружим, что в топ попадает много существительных с пометой body part, однако из этих данных сложно делать какие-либо выводы, так как мы не знаем методику сбора данных для сочетаний типа «серый козёл». Однако если это рандомный sampling, то из этого можно сделать вывод, что, возможно, описанное мной в разделе *гипотеза 2* является не аргументом в пользу правильности гипотезы, а закономерностью, соблюдающейся для сочетаемости цветовых прилагательных в целом, а не их диминутивных вариантов с существительными.

Однако сопоставительная таблица н-грамм с диминутивными показателями и без них даёт нам возможность сделать несколько важных наблюдений: среди н-грамм без диминутивных показателей можно найти такие, которые имеют некомпозициональное значение, например, «белый дом» (здание правительства некоторых государств), «красная книга» (книга, в которую занесены вымирающие представители флоры и фауны), т. к. они являются наименованием конкретных объектов реального мира, а не разновидностей того, что названо существительным. В свою очередь среди сочетаний двух диминутивов таких н-грамм не наблюдается, т. е. диминутивы, видимо, менее подвержены фразеологизации в сочетании с цветовыми прилагательными.