Все таблицы находятся на листах «Дополнительные козлики» и «Дополнительные козлы»

Для многих метрик важна не только абсолютная частота биграммы, но и отношение количества вхождений биграммы к абсолютной частотности каждого из её компонентов, поэтому я проанализировала это отношение для сочетаний с диминутивными суффиксами и без. В таблице ниже приведены все слова, значение которых по данной формуле >0.01. Жирным выделены устойчивые сочетания.

|  |  |  |  |
| --- | --- | --- | --- |
| **(серенький) козлик** | **0.0795** | **(красный) крест** | **0.0665** |
| **(аленький) цветочек** | **0.0343** | (седой) волос | 0.0291 |
| (серенький) пиджачок | 0.0197 | (седой) ус | 0.0249 |
| (седенький) волосик | 0.0137 | (белое) платье | 0.0248 |
| (беленький) цветочек | 0.0101 | (чёрный) волос | 0.0238 |
|  |  | (седая) борода | 0.0237 |
|  |  | **(чёрный) хлеб** | **0.0234** |
|  |  | (чёрное) платье | 0.0195 |
|  |  | (чёрный) дым | 0.0181 |
|  |  | (белый) платок | 0.0179 |
|  |  | (чёрный) ус | 0.0156 |
|  |  | **(белый) хлеб** | **0.0149** |
|  |  | (чёрная) борода | 0.0146 |

Вывод: с прилагательными цвета относительно часто встречаются понятия, признак цвета для которых является одним из основных (волосы, одежда) и устойчивые сочетания.

Интересно, что среди относительно наиболее частотных сочетаний встречается оттенок, которым можно описать только одну из представленных семантических групп – седой для волос (и метонимический перенос на людей, но в частотный список они не вошли). При этом чаще всего встречаются не редкие оттенки, которые наиболее часто употребляются только с определённой семантической группой (например, рыжий для волос или карий для глаз), а более базовые цвета – чёрный и белый.

Ту же процедуру я проделала для прилагательных. Здесь получилось гораздо больше сочетаний, удовлетворяющих условию >0.01, поэтому для этого анализа возьмём условие >0.03

|  |  |  |  |
| --- | --- | --- | --- |
| **аленький (цветочек)** | **0.1637** | карий (глаз) | 0.4884 |
| седенький (старичок) | 0.1324 | седой (волос) | 0.0985 |
| седенькая (бородка) | 0.0494 | русый (волос) | 0.0946 |
| голубенький (глазок) | 0.0460 | голубой (глаз) | 0.0884 |
| тёмненькая (фигурка) | 0.0366 | мутный (глаз) | 0.0544 |
|  |  | серый (глаз) | 0.0455 |
|  |  | седая (голова) | 0.0439 |
|  |  | румяное (лицо) | 0.0318 |
|  |  | седая (борода) | 0.0315 |

Интересно, что в данном случае напротив проявляются прилагательные, которые сочетаются только с одной семантической группой (что довольно ожидаемо), например, карий для глаз или русый для волос. Базовые цвета напротив имеют очень разнообразную сочетаемость, поэтому в данный список не входят.

Также интересно, что в этом списке почти нет устойчивых сочетаний – прилагательные в них как правило являются базовыми цветами с широкой сочетаемостью.

всё, больше ничего осмысленного до дедлайна я всё равно не успею. как говорится, это немного, но это честная работа. хорошего дня и спасибо за проверку!

p.s. kibuzi – это козочка на суахили. это слово очень широко известно в узких кругах...