

# КОГДА ПОРЯДОК НЕ ИМЕЕТ ЗНАЧЕНИЯ

В «Квантике» № 1 за 2022 год в одной из задач Турнира имени М.В. Ломоносова была дана таблица из 18 словосочетаний с названиями цветов, с указанием, сколько раз они встречаются в корпусе текстов на американском английском языке с 1990 по 2009 год:

|                 |      |                 |     |
|-----------------|------|-----------------|-----|
| black and blue  | 257  | blue and black  | 68  |
| black and red   | 125  | red and black   | 251 |
| black and white | 4541 | white and black | 525 |
| blue and green  | 206  | green and blue  | 194 |
| blue and pink   | 43   | pink and blue   | 95  |
| blue and red    | 174  | red and blue    | 437 |
| green and pink  | 28   | pink and green  | 65  |
| green and white | 184  | white and green | 92  |
| pink and red    | 62   | red and pink    | 40  |

**Словарь:** and – ‘и’; black – ‘чёрный’; blue – ‘синий, голубой’; green – ‘зелёный’; pink – ‘розовый’; red – ‘красный’; white – ‘белый’.

Требовалось узнать, какое словосочетание встречается в тех же текстах чаще другого: *black and pink* или *pink and black*; *blue and white* или *white and blue*; *green and red* или *red and green*; *pink and white* или *white and pink*; *red and white* или *white and red*.

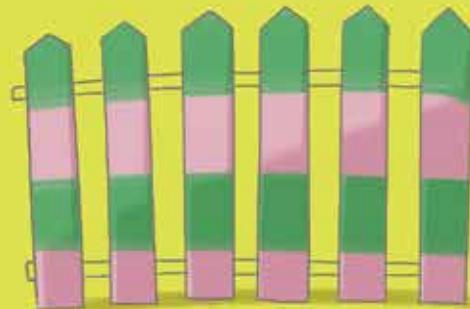
Давайте разбираться. В каком порядке перечислять элементы, порядок которых не важен? Можно подумать, что в случайном, но в реальности это не так. Например, на сайте [problems.ru](http://problems.ru) в условиях нескольких сотен геометрических задач встречается выражение «в треугольнике *ABC*», но нет ни одной задачи, где сказано «в треугольнике *ACB*», «в треугольнике *BAC*» и т. д. Здесь действует алфавитный принцип, а что ещё бывает в таких ситуациях в языке?

Пары цветов – отличный пример того же рода: какая разница, называть ли футболистов «Спартак» красно-белыми или бело-красными, футболистов ЦСКА – красно-синими или сине-красными, а футболистов «Динамо» – бело-голубыми или голубо-белыми? Но мы почему-то говорим *красно-белые* и *красно-синие*, а *голубо-белый* как-то даже и не скажешь.

Такие предпочтения удобно изучать с помощью лингвистических корпусов – больших собраний текстов, по которым можно посчитать, что встречается

## Чудеса ЛИНГВИСТИКИ

Александр Пиперски



# Чудеса ЛИНГВИСТИКИ



ся чаще, а что реже. Строгих правил мы не выведем (и так, и так можно!), но найдём какие-то закономерности. Одна из них и представлена в задаче: не то чтобы по-английски нельзя было сказать *green and pink* ‘зелёно-розовый’, но *pink and green* употребительнее. В итоге прилагательные выстраиваются в шкалу: *pink > red > black > blue > green > white*. Если цвет *X* на этой шкале левее, чем цвет *Y*, то сочетание *X and Y* частотнее, чем *Y and X*. **Ответы:** *pink and black, blue and white, red and green, pink and white, red and white*.

Задача решена, но остаётся вопрос: откуда берётся эта закономерность? Это следствие того, что лингвисты называют *четвёртым законом Бехагеля* в честь немецкого лингвиста Отто Бехагеля (1854–1936) или *законом Панини* в честь древнеиндийского лингвиста Панини (около VI–IV вв. до н. э.): при прочих равных более короткие члены предложения ставятся раньше более длинных. К примеру, по-русски мы скажем *Я подарил ей бусы*, но *Я подарил их Марине*: в первом случае сперва «кому?», а затем «что?», а во втором сперва «что?», а затем «кому?». Дополнения в дательном и винительном падеже можно расположить в любом порядке, но естественнее звучит, когда сперва стоит короткое местоимение, а за ним уже длинное существительное. А теперь попробуйте произнести английские слова из условия задачи и послушайте, какие в них звучат гласные. В *pink, red* и *black* они краткие, в *blue* и *green* – долгие, а в *white* ещё длиннее, там и вовсе дифтонг (сочетание двух гласных звуков – [a] и кратко [и]): это как раз и соответствует нашей шкале.

А в итальянском языке в слове *bianco* ‘белый’ в корне произносится краткий гласный [a], а в слове *nero* ‘чёрный’ – долгий гласный [э], и поэтому ‘чёрно-белый’ по-итальянски будет *bianconero*, а не *nerobianco*. В русском языке ударные гласные не различаются по долготе, но мы смотрим на количество слогов: односложные корни идут раньше двусложных. Мы скорее скажем *красно-зелёный*, а не *зелёно-красный*; *бело-голубой*, а не *голубо-белый* (слово *голубой* не ставят в начале ещё и потому, что у него ударение на окончании, и в *голубо-белый* пришлось бы сдвигать его на *у*, чтобы оно не попало на соединительный гласный).

Если вы внимательно решали задачу, то наверняка заметили, что в условии дано 9 комбинаций цветов, а в задании – 5, то есть пропущена одна комбинация из 15 возможных. Дело в том, что для слов *black* ‘чёрный’ и *green* ‘зелёный’ закономерность нарушается: *black and green* – 65 раз, *green and black* – 74 раза. Можно, конечно, сказать, что этот контрпример разрушает всё, и не пытаться предсказывать более распространённый порядок цветов с помощью шкалы, но всё же простая модель, описывающая абсолютное большинство случаев, – это хорошо, а с небольшим шумом можно и смириться: примерно как когда мы говорим, что Земля имеет форму шара, мы понимаем, что это не совсем правда, но предпочитаем не переусложнять.

В любом случае, статистика по текстам не может быть идеальной по разным причинам:

- Никакой корпус текстов не представляет язык полностью. Подсчёты в задаче выполнены на материале 400 млн слов из текстов 1990–2009 годов, но вполне возможно, что если расширить массив данных, добавив другие годы или даже другие тексты за те же годы, результаты изменятся. Это особенно вероятно в тех парах, где различия невелики: например, *blue and green* ‘синий и зелёный’ 206 ~ *green and blue* 194.
- Отдельные закрепившиеся сочетания могут сильно влиять на результаты. Так, *bianconeri* по-итальянски – это почти всегда футболисты команды «Ювентус», четверть вхождений *green and black* по-английски – это *green and black teas* ‘зелёные и чёрные сорта чая’ и т. д.
- Некоторые примеры на самом деле являются частью более длинного списка: к примеру, *white and blue* обычно встречается в ряду *red, white and blue*.
- Английские сочетания типа *red and white buses* могут обозначать либо двуцветные предметы («красно-белые автобусы»), либо одноцветные предметы двух видов («красные и белые автобусы») – не исключено, что порядок слов в этих значениях устроен по-разному. В русском языке *розово-красный* – это либо двуцветный, либо красный с розовым оттенком, и это тоже может портить простой подсчёт.

Вот сколько разных сложностей порой скрывается за небольшой лингвистической задачей!

Художник Мария Усеинова

