



Частотный список токенов: типа

Частотный список биграмм: как бы, абсолютная частота: 60, Range: 9

Частотные списки лемм: абсолютная частота: 55, частотный ранг: 81

Ключевые слова: ранг глагола "понимать": 62

Распределение по корпусу: Слово "горшок" (абсолютная частота: 15), встречается только в одном месте (все упоминания слова "кучкуются" в одном месте). Могу предположить, что это слово встречалось лишь в нескольких видео, довольно редко для "нашего" корпуса.

Частица "дак" (абсолютная частота: 13), встречается одиночно ближе к началу два раза (причём расстояние межу этими вхождениями довольно большое, скорее всего проскальзывало случайно и разово у людей в этих текстах), встречается группой чуть дальше середины (группа => человек довольно часто использовал эту частицу ), встречается ещё большей группой ближе к концу (в этом случае можно сказать, что это слово, которое человек очень часто использует). К последним двум предполагаю, что человек так как расположены уж слишком близко, вряд ли это относится сразу к двум текстам. Плюс, во вкладке Concordance мне показалось, что можно их по смыслу поделить именно на границах групп из Concordance Plot.

Глагол "понимать" (абсолютная частота: 55), распределён почти равномерно по всему корпусу. Только в середине чуть больше получается вхождений.

Коллокаты: По частоте: так - 3, правильно - 5, очень - 6. По статистике силы связи с глаголом: правильно - 1, так - 2, очень -3.

По статистике силы связи с глаголом «правильно» стоит на первом месте, скорее всего, это из-за того, что мы часто употребляем: «Я вас правильно понял?», «Я правильно понимаю, что…». В таком случае, среди этих трёх слов «правильно» явно лидирует, ведь «очень» мы употребляем в основном в контексте: «Я Вас очень понимаю» или «Я не очень понимаю» (в ином контексте я не слышала, чтобы эти слова стояли вместе). Что касается слова «так», то, на мой взгляд, оно попало на второе место из-за того, что порой оно не несёт смысловой нагрузки, в отличии от слова «правильно». И в целом, мы его часто употребляем с любыми глаголами (из всех наречий его можно скомбинировать с любым глаголом, и это будет совершенно нормально), поэтому оно и стоит у нас на первом месте по встречаемости среди, выбранных слов, но уходит на второе по силе связи. Я полагаю, что «правильно» стоит выше «очень» даже в списке, отсортированном по частоте, так как у них одинаковая встречаемость, а далее программа смотрит на их силу связи с глаголом, где, как мы знаем, «понимать» будет выше, чем «очень».

Плюс, сделаем поправку на то, что мы размечали интервью. Становится понятно, что оборот: «Я вас правильно понял?» будет в больших ситуациях употреблён, ведь журналисты часто спрашивают такое. В свою очередь оборот: «Я не очень понимаю», будет встречаться гораздо реже, так как журналист, чтобы переспросить, уточнить что-то будет пользоваться больше первым оборотом. А интервьюируемый скорее всего не будет пользоваться подобными словами, так как он в большинстве случаев просто рассказывает.

\*: Слово, которое чаще всего встречается – «и», всего 68 строчек из них в 39 «и» предшествует слову «вот», а перед «и вот» стоит пауза (в основном большая - //). В большинстве случаев мы так начинаем новую мысль, причём как в новом предложении, так и в части сложного, похоже на значение какого-то итога (пример: и вот Константин Анатольевич) Когда же «и» стоит после «вот», то пауза будет предшествовать «вот», кажется, что это больше о начале предложения, так как в некоторых случаях пауза стоит ещё и после «вот» (между «вот» и «и»), словно говорящий выделяет то, что будет следовать после слова, привлекает внимание, похоже на значение следствия (пример: вот … и … поэтому).