Корпус языка – это большое собрание особым образом размеченных текстов на каком-то языке. Если в названии корпуса есть слово «национальный», это обычно значит, что в такой коллекции есть тексты разных жанров и годов.

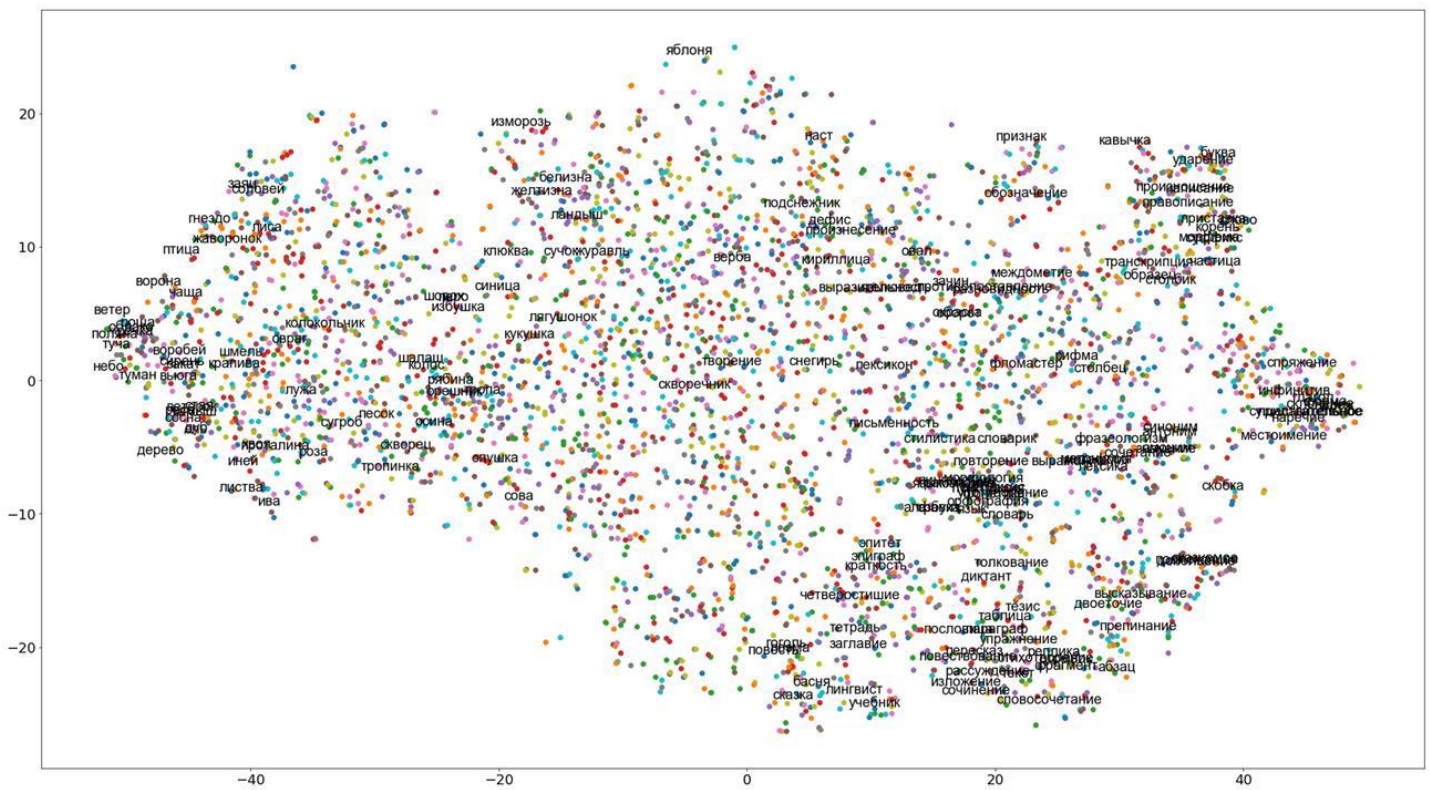
Национальный корпус русского языка (далее – **НКРЯ**) можно найти на сайте https://ruscorpora.ru/new/index.html.

Корпусные методы могут использоваться:

- при анализе учебных материалов;
- при составлении наиболее отвечающих целям и задачам обучающихся и учителей упражнений для уроков русского языка (а также на уроках иностранного языка);
- при решении (см. олимпиаду «Что может корпус?») и составлении олимпиадных заданий по языкам;
- для разработки и проведения уроков-лабораторных, на которых обучающиеся самостоятельно проводят языковое исследование.

Использование корпусных методов для анализа учебных материалов

Вот яркий пример анализа учебных материалов при помощи корпусных методов:



Эту картинку дала на выходе программа, которая ищет в текстах термины. Такая программа принимает на вход целевой и референционный корпус. Целевым корпусом в данном случае были все тексты из шести учебников для 5 класса (ФГОС), а референционным корпусом служил НКРЯ.

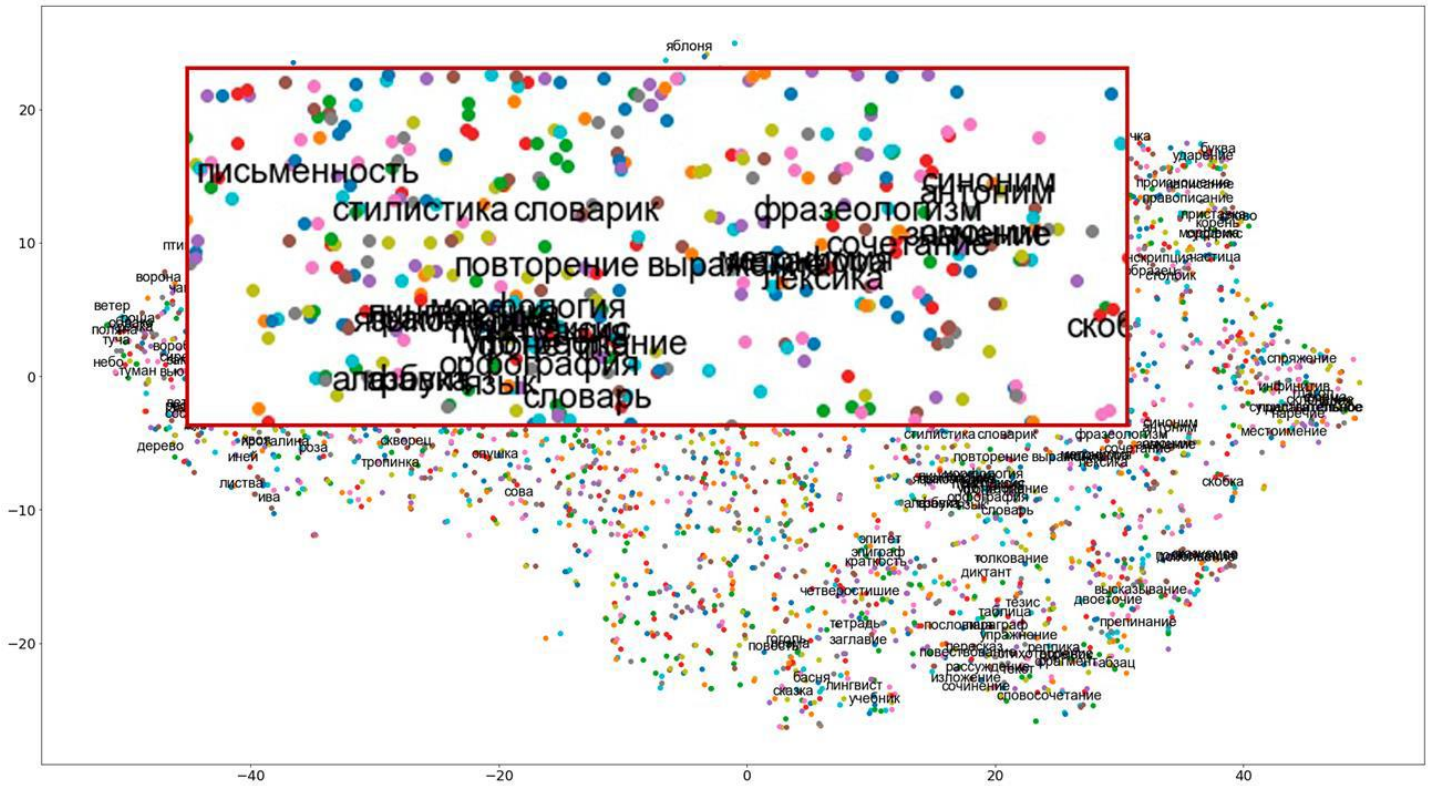
Далее для каждого слова, встретившегося в целевом и референционном корпусе, высчитывается частотность. Допустим, что референционный корпус состоит из 1 000 000 слов, а целевой – из 100 000. В целевом корпусе слово *человек* (в разных формах) встретилось 257 раз, а в референционном – 2 723 раза. В таком случае частотность данного слова в целевом корпусе рассчитывается так:

$$257 \div 100\,000 = 0.00257 \approx 0.26\%$$

Расчёт для целевого корпуса:

$$2723 \div 1\,000\,000 = 0.002723 \approx 0.27\%$$

Частотность слова *человек* примерно одинакова в обоих корпусах. Значит, это не термин, иначе бы его частотность в целевом корпусе была много больше.

Из всех найденных таким образом терминов программа создает граф. Чем ближе точки, обозначающие разные слова, тем чаще эти слова встречаются друг с другом.

Как мы видим, в правой части графика действительно представлены термины, изучаемые на уроках русского языка. Их мы и ожидали увидеть: