

К РАЗРАБОТКЕ ЦИФРОВОГО КЛАССИФИКАТОРА ЯЗЫКОВОГО МАТЕРИАЛА

Кабанова Екатерина Анатольевна

Студент (магистр)

Нижегородский государственный лингвистический университет им. Н.А. Добролюбова,
Нижний Новгород, Россия

E-mail: kabanova365@gmail.com

На данном этапе развития лингвистических учений существует ряд классификаций, охватывающих разные типы словосочетаний и обладающих различными структурами. В основе нашего исследования лежит классификация, которая базируется на таких параметрах как множественность/единичность, свобода/связанность компонентов и переносность значения. Материал исследования отобран путем анализа специальных текстов различной направленности (технический, юридический, коммерческий), неспециальных текстов (художественный), а также занимающих двойственное положение публицистических текстов.

Поиск словосочетаний в специальных текстах различной тематики проводился с помощью переводческой системы Memsource, которая в дальнейшем позволила извлечь словосочетания из исследуемых текстов и создать базу данных в виде электронных таблиц. Извлечение по заданным критериям происходит достаточно быстро, но полученный материал, как и любая терминологическая база, нуждается в непосредственной обработке. Это обусловило потребность в совершенствовании данного инструмента для профессиональной работы вышеуказанной системы, а также привлечении и разработке иных программных средств (например, на языке Python), которые могли бы послужить совершенствованию систем переводческой памяти и полностью позволили бы автоматизировать отбор языкового материала, упорядочить его по заданным критериям. Такая процедура обработки языкового материала может быть подчинена различным целям: для обучения переводу на раннем этапе для вычленения словосочетаний как единиц перевода, для их тщательного изучения и точного понимания их значения в тексте, а также систематизации словосочетаний с дальнейшим выявлением их наиболее актуальных типов для обучения аутентичной речи на иностранном языке; для составления лексикографических источников и создания лингвистической терминологической базы, которая сможет в дальнейшем найти свое применение в переводческой практике.

Вопрос о классификации словосочетаний неоднократно рассматривался в научных работах лингвистов отечественных и зарубежных школ (И.С. Парина [1, с. 51-58], М.Б. Чиков [2, с. 119-143], Ф.-Й. Хаусманн [3] и др.). Попытки систематизировать словосочетания, однако, этим не ограничились, а наоборот, с развитием информационных технологий только набирают обороты. Создание новейшего цифрового классификатора словосочетаний должно быть подчинено цели обучения их употреблению, цели систематизации, поиска и отбора наиболее актуальных для изучения словосочетаний, которые, как известно, и составляют основу аутентичной речи. В настоящее время ведется разработка цифрового инструмента для разметки словосочетаний. Предполагаемый алгоритм авторазметки должен выглядеть следующим образом: словосочетание размечается соответствующими тегами (свободное, серийное, единичное и т.п.). При этом пользователь может свободно добавлять, убирать и изменять критерии. Размеченные словосочетания помещаются в базу и извлекаются из нее по запросу. Конечный продукт разработки - новейшая электронная

база с классификатором, в котором словосочетания разбиты и по их основным признакам, и по словам, и по семантической сочетаемости; некая система, которая постоянно будет пополняться словосочетаниями, с возможностью их дальнейшей систематизации и отслеживания, что, в свою очередь, окажет большую помощь в обработке, хранении, изучении языкового материала.

Наше исследование началось с рассмотрения различных текстов на предмет присутствия в них устойчивых словосочетаний (фразеологизмов). Это было вызвано предположением о наименьшей фразеологизированности специальных текстов и, соответственно, наиболее фразеологизированном характере неспециальных текстов. Общее количество двухсловных сочетаний - 391. Отметим сразу, что большинство этих словосочетаний состояло именно из двух слов (не считая артиклей), т.е. это были сочетания существительных с прилагательными и существительных с переходными глаголами. Наречные словосочетания не рассматривались, выбор был ограничен субстантивными; это объясняется тем, что явление коллокации, которое для обеспечения аутентичности речи на изучаемом иностранном языке представляется центральным, рассматривается как правило именно на основе существительных.

Результаты количественного анализа представлены в таблице 1 (рис. 1).

Как мы видим, доля серийных и единичных сочетаний в специальных текстах значительно ниже, чем в художественном и публицистических текстах. Это объясняется тем, что фразеологизмы несут в себе преимущественно экспрессивно-оценочный компонент, присутствие которого в специализированных текстах является недопустимым.

В таблице 2 (рис. 2) представлено несколько примеров из исследуемых текстов.

С уверенностью можно сказать, что останутся пустыми первые столбцы, в которых принцип свободы словосочетаний соотносится с переносностью: свободные словосочетания переносными не бывают. Что касается единичных словосочетаний с переносным значением и главного, и зависимого слова, всего словосочетания, то это идиомы. Все ячейки в этой строке перед ними также останутся пустыми, так как переносное значение всего словосочетания может быть только у единичных словосочетаний.

Конечно, вопрос о единичности - закрытости - открытости словосочетания или группы требует дальнейшего изучения, выяснения четких критериев открытости - закрытости на основе корпусного анализа с применением новейших цифровых технологий. В качестве перспективы исследования можно отметить более подробный анализ специальных текстов и неспециальных на наличие в них коллокаций и иных фразеологических единиц, а также полную автоматизацию процесса поиска словосочетаний посредством использования иных цифровых инструментов. На примере таблиц выше мы видим, насколько неудобен такой способ представления и поиска словосочетаний; использование программы MS Excel частично оптимизирует эти процессы, однако обработать в ней большое количество материала достаточно проблематично - а ведь речь пока идет лишь о нескольких сотнях, а не о тысячах словосочетаний. По этой причине возникла острая необходимость создания классификатора для удобной разметки и затем - онлайн-репозиторий с онлайн-конструктором для удобного поиска и представления языкового материала.

Источники и литература

- 1) Парина И. С. Описание фразеологизмов в двуязычном словаре как переводческая проблема // Электронный ресурс Интернет: <https://cyberleninka.ru/article/n/opisanie-frazeologizmov-v-dvuyazychnom-slovaire-kak-perevodcheskaya-problema>
- 2) Chikov M. Kollokationen mit Raumverben im Deutschen im Vergleich zum Russischen. Muttersprache. 2017. Т. 127. № 1-2.

- 3) Hausmann, Franz J. Die Kollokationen im Rahmen der Phraseologie – systematische und historische Darstellung // Электронный ресурс Интернет: https://www.zaa.uni-tuebingen.de/wp-content/uploads/05-Hausmann_ffinal.pdf

Иллюстрации

Тип текста	Свободные сочетания	Серийные сочетания	Двойствен. сочетания	Единичные сочетания	Общ. кол-во слов-й
Все тексты	358	16	1	16	391
Худ. текст	58	7	0	7	72
Публ. текст 1	44	2	0	3	49
Публ. текст 2	16	2	1	2	20
Публ. текст 3	33	0	0	2	35
Техн. текст	69	1	0	0	70
Юр. текст	90	2	0	1	93
Комм. текст	48	2	0	1	51

Рис. 1. Результаты количественной выборки

Множеств-ть /един-ть Перен-ть	Свободные сочетания		Серийные сочетания		Двойств. сочетания	Единичные сочетания
	Со слабой связью	С сильной связью	Условно-открытые	Условно-закрытые		
У главного слова						
Метафорическая	hochtrabende Schwarten	Unsicherheit schüren		ein feste Burg		die Glanzzeit erleben
Метонимическая						die Sprache wiederfinden
Нет	eine klare Absage	das taktische Vorgehen	sanfte Stimme			den Schaden nehmen
У зависимого слова						
Метафорическая	nackte Panik		Spannungen abbauen	die bittere Wahrheit		Pläne schmieden
Метонимическая						der globale Süden
Нет	eine kalte Brise		Lippen kräuseln		Keine Ahnung haben	die territoriale Integrität

Рис. 2. Результаты качественной выборки