Составление частотных словарей, конкордансов и двуязычных словарей языка писателя как элемент подготовки филологов

Н. В. Сивенкова
Санкт-Петербургский государственный университет

Аннотация
Современные компьютерные технологии существенно облегчают кропотливую работу по составлению словарей различного типа в части подготовки материала для дальнейшего анализа и интерпретации. Широкое распространение текстовых редакторов и электронных таблиц позволило включить в цикл лексикологических и лексикографических дисциплин, преподаваемых студентам, такой модуль, как "Составление словаря (словника / частотного словаря / конкорданса / двуязычного писательского словаря)". Для выполнения студентами самостоятельной работы нет необходимости в специальном программном обеспечении. Самостоятельная лексикографическая деятельность мотивирует будущих филологов более уважительно относиться к слову как единице языка, выявлять стилистические нюансы, размышлять над причинами различий в частоте употреблений, обращать внимание на контекстные связи. Получаемые в итоге корпусы учебных словников или ряды словарных статей могут быть использованы в научно-исследовательской деятельности.

Ключевые слова: частотный словарь, конкорданс, авторская лексикография, подготовка филологов.

Современные компьютерные технологии существенно облегчают кропотливую работу по составлению словарей различного типа в части подготовки материала для дальнейшего анализа и интерпретации. Широкое распространение текстовых редакторов и электронных таблиц позволило включить в цикл лексикологических и лексикографических дисциплин, преподаваемых студентам, "Составление словаря (словника / частотного словаря / конкорданса / двуязычного писательского словаря)".

Изучение иностранного языка уже много десятилетий для многих, если не всех учащихся начинается с выписывания иностранного слова и его перевода в особую тетрадку, которая носит громкое имя "Тетрадь для записи иностранных слов". Процесс этот не очень почитаем студентами, увы. Что еще обиднее, такие словники, глоссарии воспринимаются как рабочий, черновой материал и выкидываются за ненадобностью по окончании обучения. В то же время использование электронных

средств могло бы изменить эту ситуацию. Например, даже в рамках курса "домашнее чтение" важно не просто выписывать варианты перевода слова, но изучать лексемы как составную часть языка и стиля писателя. Возможно, для будущих переводчиков технических текстов это и является излишним, но при подготовке филологов обращение к авторской лексикографии необходимо. Комплексное, многогранное изучение *языка писателя* является сейчас актуальным. Одним из замечательных инструментов для такого рода исследований являются частотные словари и словари-конкордансы, число которых все возрастает.

М. Л. Гаспаров, расшифровывая конкорданс как расширенный словарь языка писателя, обращает внимание на то, что "все мельчайшие оттенки значения каждого слова выступают при этом так исчерпывающе ясно, как ни в каком другом словаре. Изучать по-настоящему язык и литературу, не имея таких конкордансов, невозможно" [1, с. III]. Обратим внимание на справедливое определение конкорданса именно как *словаря языка* писателя, а не просто индекса (словника, списка слов), выводимого в результате того или иного запроса при автоматической обработке текста, хотя и такие определения имеют место, и под конкордансом иногда понимается просто список примеров со ссылками на источник, сформированный в результате поиска языкового выражения по корпусу текстов.

История конкорданса насчитывает уже не одно десятилетие, но в последние годы интерес к такого рода словарям существенно возрос. Появление большего числа конкордансов повышает и требования к ним. Так, по мнению некоторых исследователей, сейчас уже нет необходимости в составлении списков лексем без контекста, толкования или грамматической информации: "Подобные конкордансы в эпоху компьютерной лексикографии могут быть получены почти механически и не представляют особого интереса. Ценность им придает только дополнительная лингвистическая информация, которая не выводится автоматически, а является результатом труда лингвиста" [4, с. III]. Именно при такой постановке вопроса конкорданс из механически "препарированного" корпуса текстов превращается в инструмент исследования языка писателя.

Одним из ранних конкордансов к текстам русской классической литературы, полученных при помощи компьютерных программ, является "Конкорданс (указатель слов в контексте) к роману Ф. М. Достоевского 'Преступление и наказание'" [2]. При этом под *контекстом* понимается отрезок в 49 символов (22 знака до начала искомого слова плюс 27 знаков справа от первой буквы в искомом слове). Безусловно, для поиска искомого слова в корпусе текстов подобный "контекст" можно считать достаточным, но, к сожалению, он мало пригоден для толкования или прояснения значения слова, равно как и для уточнения особенностей словоупотребления.

Другой важной вехой среди русскоязычных конкордансов, полученных благодаря компьютерным технологиям, можно считать "Конкорданс к стихам А. С. Пушкина" Дж. Т. Шоу [5], который включает все словоформы кириллицей, за исключением тех, что были сочтены непечатными. По каждой из них приводятся сведения о том, сколько раз она употреблена в тексте, строки, в которых она употреблена. Индекс содержит сокращенное название произведения и номер строки в нем. Использование строки (более или менее определенного смыслового целого), а не механическое "отсечение" знаков до и после искомого слова, дает более удовлетворительный с точки зрения репрезентативности контекст, однако и такой подход не всегда себя оправдывает, особенно если учесть, что в словаре индексируется, но не воспроизводится целиком так называемая "непечатная" лексика. Несмотря на мелкие "огрехи" такого рода, нельзя не оценить по достоинству упомянутый "Конкорданс к стихам А. С. Пушкина", ибо именно этот словарь, судя по всему, является одной из основных моделей для построения новых конкордансов к русской поэзии.

Сейчас существует уже немало хороших и интересных конкордансов на материале русской литературы. В болгаристике в этом отношении есть некие лакуны. В связи с этим возникла необходимость подготовки электронного конкорданса к поэзии П. К. Яворова [3], причем было принято решение отойти от традиции минимизации иллюстрации к словоупотреблению. Недостаточность "контекста" объемом в 22 знака справа / слева, в строку и т. п. (при том, что подобные "контексты" легко выделяются компьютерной программой) очевидна: без обращения по адресу, указанному в словарной статье, значение слова может быть невыводимо. Однако и полный, всеохватывающий контекст, максимально иллюстрирующий оттенки значения для конкорданса-словника, является неоправданной "роскошью". В качестве компромиссного варианта было принято решение приводить контекст, соответствующий предложению. Это, с одной стороны, осложнило работу (автоматический режим превращается едва ли не в ручной), а с другой — открыло новые перспективы.

Использование электронной формы для конкорданса имеет ряд достоинств. Во-первых, только электронные словари позволяют быструю сортировку, перегруппировку, выборочную композицию анализируемого материала. Во-вторых, именно в виде таблиц (баз данных) наиболее удобно проводить сопоставительный анализ, сравнение отдельных фрагментов текстов, составление сводных таблиц. Конкорданс не является самоцелью и не предполагает получения "конечного результата" (а именно это обычно ассоциируется с печатным изданием). Это прежде всего мощный лексикографический инструмент исследования языка писателя. И в электронном виде этот инструмент намного более удобен, нежели в виде книги. Отметим, что

рынок быстро реагирует на запросы, и сейчас в интернете уже можно найти платные программы для составления конкордансов.

Создание студентами глоссариев, частотных словарей, конкордансов к отдельным произведениям (для этого необязательно использовать специальные программы, достаточно Word и Excel) является сложным, но весьма эффективным учебным заданием для будущих филологов. Хочется надеяться, что компьютерная лексикография находится еще в самом начале своего пути и ее бурное развитие приведет к дальнейшему развитию языкознания в целом.

Список литературы:

1. Гаспаров М. Л. Предисловие к Конкордансу к стихотворениям М. Кузмина // Конкорданс к стихотворениям М. Кузмина / сост. А. В. Гик. М. : Языки русской культуры, 2005. С. III.
2. Конкорданс (указатель слов в контексте) к роману Ф. М. Достоевского "Преступление и наказание" / сост. А. Андо, Я. Ураи, Т. Мотидзуки. Хоккайдо : Центр славяноведения, 1994.
3. Сивенкова Н. В. Электронный словарь-конкорданс к поэзии П. К. Яворова // XVI Державинские чтения: Современные и исторические проблемы болгаристики и славистики : материалы XL Международной филологической конференции, 14–18 марта 2011 г. СПб., 2011. С. 87–95.
4. Словарь языка А. С. Грибоедова / сост. А. Е. Поляков. М., 2008. С. III.
5. Шоу Дж. Т. Конкорданс к стихам А. С. Пушкина. М. : Языки русской культуры, 2000.

Сведения об авторе: Наталья Владимировна Сивенкова, старший преподаватель, СПбГУ; email: n.sivenkova@gmail.com

WORKING OUT FREQUENCY DICTIONARIES, CONCORDANCES AND BILINGUAL DICTIONARY OF AN AUTHOR AS AN ELEMENT OF TEACHING STUDENT-PHILOLOGISTS

Natalya V. Sivenkova

St. Petersburg State University teacher; email: n.sivenkova@gmail.com

Abstracts

The article presents some materials allowing to consider the work on compiling dictionaries and vocabularies of various types as a part of preparing the material for further analysis and interpretation. Students should be involved in creating such lexicographical issues as concordances and author's dictionaries, especially if they aim to be philologists. The modle is called "Making a dictionary/vocabulary (glossary / frequency dictionary / concordance / bilingual dictionary of an author)". There is no need

for the students to obtain special software in order to accomplish the task themselves. The educational glossaries, dictionaries or groups of lexical entries, which appear after such lexicographical exercises, can be used in further research activities.

Key words: glossary, frequency dictionary, concordance, bilingual dictionary of an author.

References:

1. Gasparov, M. L. (2005). Predislovie k Konkordansu k stikhotvoreniiam M. Kuzmina [Introduction to the Concordance of M.Kuzmin's poetry]. *Konkordans k stikhotvoreniiam M.Kuzmina* (compiled by A. V. Gik). M.: Iazyki russkoi kul'tury. P. III.
2. *Konkordans (ukazatel' slov v kontekste) k romanu F. M. Dostoevskogo "Prestuplenie i nakazanie"* [Concordance (index of the words in the context) to the F.Dostoevskiy's novel "Crime and Punishment"]. (1994). (compiled be A. Ando, Ia. Urai, T. Motidzuki). Khokkaido: Tsentr slavianovedeniia.
3. Sivenkova, N. V. (2011). Elektronnyi slovar'-konkordans k poezii P. K. Iavorova [Machine-readable concordance to the P. K. Yavorov's poetry]. *XVI Derzhavinskie chteniia: Sovremennye i istoricheskie problemy bolgaristiki i slavistiki: Materialy XL Mezhdunarodnoi filologicheskoi konferentsii, 14–18 marta 2011 g.* SPb. P. 87–95.
4. *Slovar' iazyka A. S. Griboedova* [Dictionary of the A. S. Griboedov's language]. (2008). (compiled be A. E. Poliakov). M. P. III.
5. Shou, Dzh. T. (2000). *Konkordans k stikham A. S. Pushkina.* M.: Iazyki russkoi kul'tury.

Возможности параллельного корпуса НКРЯ при обучении специальной лексике

И. А. Терентьева
Воронежский институт МВД России

Аннотация

Корпусная лингвистика является одним из активно развивающихся направлений. За последнее время были созданы национальные корпусы практически всех языков. Русский язык не стал исключением, сейчас доступен для работы Национальный корпус русского языка. Помимо основного корпуса, создан и корпус параллельных текстов, где для каждого предложения на иностранном языке дан перевод этого предложения на русский язык и наоборот. Параллельный корпус можно использовать не только для изучения базовой лексики, но и для обучения специальной и академической лексике. Корпус