

В. К. Щербин

СРАВНИТЕЛЬНАЯ ЛЕКСИКОГРАФИЯ И ПАРАЛЛЕЛЬНАЯ ТЕКСТОГРАФИЯ: АНАЛИЗ ВЗАИМОСВЯЗЕЙ

Как свидетельствуют историки лексикографии [1, с. 189], первоначальной формой словарной деятельности была текстовая лексикография (или глоссография), которая развивалась в виде разнообразных глосс (толкований непонятных слов) на полях или между строками в текстах древних рукописей XI–XVI столетий. Со временем практика глоссирования непонятных слов в текстах настолько усовершенствовалась и специализировалась, что стали различаться следующие виды глосс: толковые, текстологические, этимологические, ссылочные [2, с. 216]; эквивалентные [3, с. 222], грамматические [4, с. 62] и др. Именно эти отличающиеся виды глосс послужили прообразами тех лексикографических параметров (толкового, иллюстративного, этимологического, отсылочного, переводного, грамматического и др.), доминированием которых в том или ином словаре сегодня определяется тип такого словаря [5, с. 29]. Для данной статьи наибольший интерес представляют иллюстративный и переводной параметры, в основе которых, по мнению Е. М. Верещагина и В. Г. Костомарова, лежат приемы контекстного изъяснения и эквиваленции [3, с. 222].

На наш взгляд, «прием контекстного изъяснения слова (путем включения его в контекст литературной цитаты, пословицы, поговорки, специально созданного лексикографом типового предложения) выступает одним из ключевых лексикографических параметров в справочных изданиях таких типов, как толковые, исторические словари, конкордансы, словари языка писателя и др.» [5, с. 29]. Причем наиболее широкое и разнообразное использование иллюстративный (цитатный, текстологический) параметр или «контекст-глосса» (по выражению Ю. Н. Трубочева) получил не в общеязыковых справочниках, а в двуязычных терминологических словарях, в которых, как правило, ставится задача дать максимально точное и по возможности однозначное толкование понятийного содержания описываемого термина и его иноязычного эквивалента. С этой целью в двуязычных терминологических словарях используются следующие типы научных цитат: 1) фрагменты научных текстов (терминированные словосочетания и сложные номинативные единства); 2) цитаты-дефиниции из других словарей; 3) авторские глоссы из научных текстов; 4) текстовые единицы понимания; 5) контекстные изъяснения [6].

К примеру, научные цитаты первого типа (фрагменты научных текстов) активно использовались в двух выпусках «Тематического русско-белорус-

ского словаря языка научных отчетов» (далее ТРБСЯНО) [7; 8]. Так, в первом выпуске указанного словаря (в первой части его макроструктуры) насчитывается около 4 тысяч словарных статей со стандартной микроструктурой (русская реестровая единица – белорусский эквивалент – порядковый номер текстового фрагмента из второй части словаря, в котором данная реестровая единица употребляется). Во второй части макроструктуры указанного справочника «подаются «оправдательные» цитаты для всех без исключения реестровых русских единиц и их белорусских эквивалентов из первой части словаря. В качестве таких цитат используются полностью эквивалентные по содержанию текстовые фрагменты научных отчетов на русском и белорусском языках, которые приводятся в виде параллельных колонок на одной и той же странице словаря. Каждая пара таких колонок имеет свой порядковый номер (от 1 до 510). При этом, как легко заметить, общее количество «оправдательных» цитат во второй части словаря уступает общему количеству реестровых единиц из первой части словаря, поскольку каждый текстовый фрагмент используется в качестве иллюстрации в среднем для 10 реестровых единиц» [7, с. 11]. Кроме того, объем отдельных текстовых фрагментов в ТРБСЯНО колебался от 2–3 до 60-ти слов и более в зависимости от степени известности реестрового термина. Как правило, менее известные термины иллюстрировались и семантизировались посредством использования более объемных текстовых фрагментов.

Научная общественность Беларуси положительно восприняла новшества, представленные в указанном словаре: 1) иллюстрацию и семантизацию русских реестровых единиц и их белорусских эквивалентов при помощи текстовых фрагментов научных отчетов, совокупность которых представлена в макроструктуре словаря отдельной, второй частью (ее содержание упорядочено при помощи цифрового указателя); 2) возможность использования второй части словаря в качестве русско-белорусской сопоставительной грамматики. Свидетельством позитивной оценки ТРБСЯНО стало значительное количество заявок на его получение, поступивших от академических институтов физико-математического и физико-технического профиля. Более того, один из читателей словаря, доктор технических наук А. И. Добролюбов, используя макроструктурную схему ТРБСЯНО, составил близкий по типу словарь английского языка [9] и даже запатентовал метод его создания в России [10]. И хотя уже в самом названии словаря А. И. Добролюбова декларируется его принципиальная новизна, в структурном отношении данный словарь как две капли воды похож на ТРБСЯНО, изданный на несколько лет раньше.

Определяя перспективы дальнейшего использования иллюстративного параметра в словарях разных типов, можно сделать следующий вывод: в современных условиях, 1) когда наличие Интернета и электронных корпусов национальных языков позволяет практически мгновенно находить необходимый фрагмент текста, в рамках которого используются интересные исследователя слова, термины и другие языковые единицы; 2) когда требования международного авторского права и словарной этики становятся все более жесткими в отношении случаев плагиата и компиляции чужих высказываний

в справочных изданиях; 3) когда юридически корректное использование высказываний известных ученых придает справочному изданию ореол научности; 4) когда источниками цитат все чаще становятся авторские тексты, представленные в электронной форме, что фактически исключает ошибки в цитировании этих текстов и позволяет представить их в оригинале, не вызывая сомнений обоснованность прогноза о том, что частота и объемы использования цитат в словарях разных типов со временем будут только увеличиваться [6, с. 168].

Более того, по мере развития инструментальных возможностей современной корпусной лингвистики и расширения практики использования цитат в словарях разных типов, можно смело прогнозировать развитие ряда новых лексикографических направлений: словарно-ориентированной текстографии, конкордансной и гипертекстовой лексикографии. К примеру, в области авторской лексикографии сегодня уже нередким является включение в состав словаря языка писателя (в виде приложения) полнообъемных художественных текстов самого писателя (см., например, [11]). В свою очередь, в рамках диалектной лексикографии организован выпуск региональных текстографических сборников, неотъемлемыми компонентами которых являются а) подобранные по географическому и тематическому принципам образцы диалектных и этнолингвистических текстов; б) лексические и фразеологические словари говоров (см., например, [12]).

Успешно развивается в Беларуси и конкордансная лексикография, саму основу которой составляет широкое использование иллюстративного параметра, поскольку конкорданс к тому или иному тексту (совокупности текстов) – это не что иное, как полный перечень выявленных в данном тексте словоупотреблений с контекстами их использования и адресными характеристиками. К настоящему времени в виде CD-дисков выпущены: 1) электронный «Канкарданс Кандрата Крапівы» (2011); 2) электронный «Канкарданс беларускай мовы XIX стагоддзя» (2016), являющийся комплексом словарей, которые были составлены на основе корпуса текстов, включающего все произведения белорусской художественной литературы XIX века (всего 497 текстов).

Наконец, импульс развитию гипертекстовой лексикографии дало появление Интернета и лежащих в его основе сетевых технологий. Именно широкое использование современных компьютерных гипертекстовых технологий позволило издательской компании «Encyclopaedia Britannica» создать в электронном и печатном вариантах такой объемный и комплексный (линейно-нелинейный) гипертекст, как «Великие книги западной цивилизации». 60 томов печатного варианта данного гипертекста содержат 517 линейных прецедентных текстов 130 авторов, к которым прилагается уникальный двухтомный индекс «Синтопикон». При помощи индекса содержание указанных линейных текстов группируется по 3020 «революционным идеям», наиболее повлиявшим на развитие западного мира [13, с. 32]. Создание данного гипертекста наглядно показало, что «объем памяти и быстродействие современных ЭВМ уже таковы, что позволяют, в отличие от описанных выше печатных словарей цитат, хранить на машинных носителях огромное

количество полнообъемных линейных текстов, осуществляя при этом мгновенный вход в эти тексты практически через любое слово, встречающееся в них» [Там же, с. 32].

В свою очередь, «прием эквиваленции является, пожалуй, наиболее распространенным средством семантизации реестровых слов в таких типах словарей, как переводные (двуязычные и многоязычные), синонимические и др.» [5, с. 29]. Причем, если раньше в переводных словарях использование приема эквиваленции ограничивалось рамками отдельных слов и словосочетаний, то в микроструктуре словарной статьи сравнительных (двуязычных и многоязычных) словарях концептов нередко представлены полностью эквивалентные (параллельные) тексты на разных языках. Иногда объем этих текстов составляет несколько страниц (см., например, русскоязычные, польскоязычные и англоязычные части биографических словарных статей в польском пятитомном справочнике «*Idee w Rosji. Идеи в России. Ideas in Russia: Leksikon rosyjsko-polsko-angielski*» (1999–2003)).

В Беларуси, где еще в советский период широко практиковалось одновременное использование как белорусского, так и русского языков, имеющих сегодня статус государственных языков, нередким является издание полностью эквивалентных по содержанию белорусскоязычного и русскоязычного справочников. Например: пятитомные «Беларуская ССР. Кароткая энцыклапедыя» (1978–1981) и «Белорусская ССР. Краткая энциклопедия» (1979–1982); «Слоўнік фізічных і астранамічных тэрмінаў» А. И. Болсуна и Е. Н. Рапановича (1979) и их же «Словарь физических и астрономических терминов» (1986) и т.д. Аналогичным образом дело обстоит и с учебной литературой Беларуси: «учебники представляют собой параллельные тексты, т.е. учебник по какому-либо предмету, написанный на белорусском языке, является переводом того же самого учебника, написанного на русском языке, и наоборот» [14, с. 44].

Сравнительная лексикография Беларуси может получить дополнительный импульс в своем развитии, если в каждой области профессиональных знаний будет создана необходимая текстовая база в виде электронных корпусов параллельных текстов на государственных языках Беларуси. Наличие такой текстовой базы содействовало бы также качественному выполнению всех требований Закона «О языках в Республике Беларусь». Кроме того, всегда следует помнить о том, что «полноценность и адекватность спектра лексикографических систем в значительной степени определяют уровень и достоверность извлечения информации и знаний из текстов различного состава, структуры и назначения» [15, с. 38].

Масштабная работа по формированию таких электронных корпусов параллельных текстов в нашей республике ведется на протяжении многих лет. Особенно весомый вклад в теорию и практику создания указанных корпусов параллельных текстов сделан профессором А. В. Зубовым, который, начиная с защиты в 1969 г. кандидатской диссертации на тему «Переработка текста естественного языка в системе «человек–машина», уже много лет занимается компьютерной обработкой текстов. В частности, для рассматриваемой нами темы А. В. Зубов не только дал точные и исчерпывающие

определения многим ключевым понятиям в этой области (см., например, следующее определение: «Корпус параллельных текстов состоит из множества текстов, написанных на каком-либо исходном языке, и текстов-переводов этих исходных текстов на один или несколько других языков» [16, с. 132]), но и раскрыл взаимосвязи, существующие между параллельной текстографией и сравнительной лексикографией: «...корпуса параллельных текстов ... позволяют: 1) автоматически строить двуязычные и многоязычные словари; 2) автоматически создавать и пополнять словари для систем машинного перевода...» [Там же, с. 136].

Таким образом, рассмотрение взаимосвязанного развития двух направлений прикладной лингвистики (сравнительной лексикографии и параллельной текстографии) показало следующее:

- в рамках обоих направлений рассматриваются совокупности текстов: объект сравнительной лексикографии – это тексты дву- и многоязычных словарей разных типов, т.е. нелинейные тексты; объект параллельной текстографии – это параллельные тексты на разных языках, обработанные специальным образом (тэггирование, лемматизированная, морфологическая, синтаксическая и иная разметка [17]), т.е. линейные тексты;

- главной функцией сравнительной лексикографии является обслуживание тех совокупностей линейных разноязычных текстов, которые репрезентируют соответствующие национальные культуры. На наш взгляд, «удельный вес нелинейных текстов (словарей, энциклопедий, справочников и т.п. гипертекстов) в общей совокупности разножанровых текстов, репрезентирующих ту или иную национальную культуру, будет непрерывно увеличиваться, обеспечивая этим самым целостность и информационную доступность данной культуры» [18, с. 19–20];

- перечень основных функций параллельной текстографии намного шире. В их числе: познавательная, трансляционная, учебная, информационная и иные, более специализированные функции. К примеру, представителям сравнительной лексикографии электронные корпуса параллельных текстов помогают: а) находить переводные эквиваленты многоязычных слов [19]; б) вырабатывать алгоритм перевода терминологических словосочетаний [20]; в) автоматически устранять полисемию лексических единиц [16, с. 136]; г) переводить несвободные словосочетания [21], а также решать множество иных словарных задач.

ЛИТЕРАТУРА

1. Ковтун, Л. Языкознание у восточных славян / Л. Ковтун // История лингвистических учений. Позднее Средневековье. – СПб. : Наука, 1991. – С. 182–207.
2. Жураўскі, А. І. Тыпы і асаблівасці глос у “Евангеллі” В. Цяпінскага / А. І. Жураўскі // Пр. Інстытута мовазнаўства АН БССР. – Мінск : Выд-ва АН БССР, 1960. – Вып. VII. – С. 213–226.
3. Верещагин, Е. М. Лингвострановедческая теория слова / Е. М. Верещагин, В. Г. Костомаров. – М. : Русский язык, 1980. – 320 с.
4. Анічэнка, У. В. Лексікаграфічная дзейнасць Скарыны / У. В. Анічэнка // Питання східнослов'янської лексикографії XI–XVII ст. : матер. симп. – Київ : 1979. – С. 58–64.

5. *Щербин, В. К.* Белорусская лексикографическая традиция: история и современность / В. К. Щербин // Славянская лексикография; отв. ред. М. И. Чернышева. – М. : Изд. центр «Азбуковник», 2013. – С. 21–61.
6. *Щербин, В. К.* Функции научных цитат в терминологии и специальной концептографии / В. К. Щербин // Термінологічний вісник. – 2013. – № 2. – С. 162–170.
7. Тематический русско-белорусский словарь языка научных отчетов / Физ.-мат. науки; сост. В. К. Щербин. – Минск : Ротапринт УД АН Беларуси, 1992. – Вып. 1. – 248 с.
8. Тематический русско-белорусский словарь языка научных отчетов / Физ.-техн. науки; сост. В. К. Щербин. – Минск : Ротапринт УД АН Беларуси, 1993. – Вып. 2. – 293 с.
9. *Добролюбов, А. И.* Словарь типа INSITU английского языка / А. И. Добролюбов // Словарь, грамматический справочник и сборник текстов для изучающих английский язык. – Минск : Амалфея, 1999. – 367 с.
10. Патент Российской Федерации № 2054200 на изобретение «Метод «INSITU» для создания словарей с целью перевода на иностранные языки». – М., 1996.
11. *Кимягарова, Р. С.* Словарь языка басен И. А. Крылова / Р. С. Кимягарова // Крылов И. А. Избранные басни. – М. : Оникс, 2006. – 928 с.
12. Дыялектная мова Гродзеншчыны: хрэстаматыя; аўт.-склад.: М. А. Даніловіч, Н. К. Памецька. – Гродна : ГрДУ, 2007. – 260 с.
13. *Щербин, В. К.* Словарные истоки практики нелинейного письма / В. К. Щербин // Весн. Бел. дзярж. ун-та. Сер. 4. – 2003. – № 2. – С. 28–33.
14. *Зубов, А. В.* О возможностях использования параллельных учебных текстов в лексикографии / А. В. Зубов // Новое в теории и практике лексикографии: синхронный и диахронный подходы : матер. VIII Междунар. шк.-семина., Иваново, 10–12 сент. 2009 г. – Иваново : Иван. гос. ун-т, 2009. – С. 44–45.
15. *Беляева, Л. Н.* Корпус научных текстов как источник извлечения лексикографической информации / Л. Н. Беляева // Новое в теории и практике лексикографии: синхронный и диахронный подходы : матер. VIII Междунар. шк.-семина., Иваново, 10–12 сент. 2009 г. – Иваново : Иван. гос. ун-т, 2009. – С. 38–41.
16. *Зубов, А. В.* Корпусная лингвистика и ее возможности для проведения лингвистических исследований / А. В. Зубов // Сучасныя праблемы беларускай лексікалогіі і лексікаграфіі : матэр. Міжнар. навук. канф., Мінск, 19–20 ліст. 2005 г. – Мінск : Права і эканоміка, 2006. – С. 129–140.
17. *Зубов, А. В.* О целях издания и принципах организации параллельного тэгірованнаго русско-белорусского корпуса учебных текстов / А. В. Зубов // Язык и дискурс в статике и динамике. – Минск : МГЛУ, 2008. – С. 37–38.
18. *Щербин, В. К.* Семиотические особенности словарно-справочных текстов / В. К. Щербин // Актуальные проблемы теоретической и прикладной лингвистики : матер. Междунар. науч. конф., посвящ. памяти проф. Р. Г. Пиотровского, Минск, 15–16 июня 2010 г. : в 2 ч.; отв. ред. А. В. Зубов. – Минск : МГЛУ, 2010. – Ч. 2. – С. 16–21.
19. *Зубов, А. В.* Использование параллельных текстов для определения переводных эквивалентов многозначных слов / А. В. Зубов // Матер. Междунар. конф. «Корпусная лингвистика и лингвистические базы данных». – СПб., 2002.
20. *Зубов, А. В.* Алгоритм перевода терминологических словосочетаний с использованием параллельных текстов / А. В. Зубов // Теория и практика перевода. – 2005. – № 1. – С. 64–67.
21. *Зубов, А. В.* Использование параллельного корпуса текстов для перевода несвободных словосочетаний / А. В. Зубов // Вестн. Моск. гос. обл. ун-та. Сер. Лингвистика. – 2011. – № 6. – С. 79–83.