

УДК 82-34:81'33

ИСПОЛЬЗОВАНИЕ КОРПУСНОГО МЕТОДА ПРИ ИЗУЧЕНИИ ФУНКЦИОНИРОВАНИЯ ЛЕКСЕМЫ «ЛЮБОВЬ» В СКАЗКАХ И ФЭНТЕЗИ

В. В. Петрожицкая¹⁾, О. И. Колесникова²⁾

¹⁾ Вятский государственный университет,
ул. Московская, 36, 610000, г. Киров, Кировская обл., Россия, vikselyok@mail.ru

²⁾ Вятский государственный университет,
ул. Московская, 36, 610000, г. Киров, Кировская обл., Россия, usr08728@vyatsu.ru

Корпусные технологии произвели своего рода революцию в лингвистических исследованиях благодаря высокой скорости обработки большого объема данных. Одним из новых направлений их применения является изучение концептов через их лексические репрезентации в текстах литературы. В статье представлен обзор возможностей отдельных инструментов программы AntConc, использованных для выявления особенностей функционирования лексемы «любовь» как репрезентанта одноименного концепта в художественных текстах «сказочных» жанров.

Ключевые слова: корпусная лингвистика; корпусные технологии; программа AntConc; лексема «любовь».

THE USE OF THE CORPUS METHOD IN THE STUDY OF THE FUNCTIONING OF THE LEXEME “LOVE” IN FAIRY TALES AND FANTASY

V. V. Petrozhitskaya¹⁾, O. I. Kolesnikova²⁾

¹⁾ Vyatka State University, Moskovskaya str., 36, 610000, Kirov, Kirov region, Russia,
vikselyok@mail.ru

²⁾ Vyatka State University, Moskovskaya str., 36, 610000, Kirov, Kirov region, Russia,
usr08728@vyatsu.ru

Corpus technologies have revolutionised linguistic research due to the high processing speed of a large amount of data. One of the new directions of their application is the study of concepts through their lexical representations in literary texts. The article provides an overview of the capabilities of AntConc program tools used to identify the features of the functioning of the lexeme “love” as a representative of the same-name concept in fiction texts of fairy-tale novels as the genre.

Keywords: corpus linguistics; corpus technologies; AntConc program; lexeme «love».

Корпусная лингвистика представляет собой некую совокупность процедур и методов для исследования явлений в языке [1]. Суть данных методов заключается в компьютерном анализе обширной коллекции специально подобранных, соответствующих исследовательским целям письменных текстов в машиночитаемом формате, или корпусов. По мнению В. А. Плунгяна, широкое распространение корпусных исследований обусловил подход, ориентированный на узус, т. е. тексты. Ученый подчеркивает, что узус «является единственной подлинной реальностью науки о языке, т. е. объектом, доступным непосредственному наблюдению» [2, с. 16]. Кроме того, в методологическом отношении корпусная лингвистика, как отмечает А. А. Яковлев, имеет неоспоримое достоинство – быстрый охват большего объема информации (конкретных словоупотреблений), чем было возможно до применения корпусов [3, с. 67]. При этом основной проблемой корпусных исследований является то, что полученные исследователем результаты и выявленные закономерности применимы лишь для конкретного корпуса. Тем не менее корпусные технологии все чаще используются в лингвистических исследованиях по объективным причинам, связанным с прагматическими задачами охватить большой объем языковых данных, оперативно провести их анализ и получить достоверные результаты. В частности, К. С. Шилихина, использовавшая корпусный метод для отбора языковых фактов с целью описания когнитивных, семантических и прагматических свойств вербальной иронии, отмечает, что традиционный подход к выбору эмпирического материала, основанный на интуиции и опыте исследователя, не всегда эффективен из-за выбора тех фактов, которые в наибольшей степени подтверждают выдвигаемую гипотезу. В этом случае не принимается во внимание вариативность языка, игнорируются те контексты, которые не подтверждают теорию и, более того, приводят к повторяемости примеров в работах лингвистов [4].

Среди программ, позволяющих осуществлять автоматический анализ большого объема данных, можно выделить программу AntConc, разработанную Л. Энтони [5]. В связи с бесплатным доступом и понятностью интерфейса она неоднократно использовалась лингвистами. Например, Д. А. Палийчук характеризует перспективность применения инструментария AntConc при изучении коллокаций [6]. К достоинствам относятся бесплатный доступ, технические характеристики, способы представления данных. В то же время отмечены недостатки. По сравнению с программой SketchEngine AntConc имеет менее удобный пользовательский интерфейс и менее эффективно определяет коллокации. Ю. В. Богоявленская представила методику работы с программой в исследовании синтагматической сочетаемости. Среди достоинств указанного сервиса отмечаются скорость,

простота и удобство пользования, понятный интерфейс, свободный доступ, самостоятельный выбор исследуемого материала, различные функциональные возможности, статистический аппарат. Среди недостатков, которые являются не критичными, названы невозможность сохранения загруженных текстов, большое количество «шума» (бессмысленных коллокаций), невозможность свести выявленные словоформы к одной лемме и некоторые другие [7]. В работе Е. С. Луньковой проиллюстрированы возможности программы AntConc для лингвистического анализа художественного текста на примере повести А. С. Пушкина «Метель» и приведены некоторые количественные (частотность) и качественные (лексические и грамматические особенности произведения) результаты [8]. Именно программа AntConc, обладающая простым и понятным принципом работы, а также доступностью, была выбрана для нашего исследования, целью которого было выявление возможностей данного сервиса при изучении лексической репрезентации концепта «любовь» в текстах сказок и фэнтези. В качестве материала исследования использовались 11 текстов: сборник «Заветные русские сказки» А. Н. Афанасьева, а также произведения «новой прозы», созданные по мотивам ряда народных сказок: А. В. Рубанов «Финист – ясный сокол», Д. Морган «Иди через темный лес», «Вслед за змеями» и «Самое красное яблоко», Л. Арден «Невеста Ноября», К. Кей «Плащ для Красной Шапочки», У. Черкасова «Золотые земли. Сокол и Ворон», «Золотые земли. Птицы Великого леса» и «Золотые земли. Совиная башня», А. Чароит «Кощеева невеста».

Прикладной задачей исследования было практическое применение инструментария программы AntConc для: 1) автоматического составления конкорданса – списка контекстов с употреблением леммы «любовь»; 2) определения частоты встречаемости леммы «любовь» в загруженных текстах; 3) выявления словоформ, в которых она употребляется; 4) выделения кластеров и коллокатов; 5) исследования равномерности распределения леммы в материале. Все тексты в формате PDF были загружены в программу. При этом зафиксировано 1237273 токенов (слов), подлежащих анализу. Далее использовались возможности инструмента KWIC, который формирует конкорданс, т. е. позволяет искать конкретные слова / фразы и исследовать их употребление в контексте. Для составления конкорданса активизировались следующие параметры в поисковом окне: *case* – для поиска словоформ леммы «любовь» в верхнем и нижнем регистре и *regex* – для ввода регулярного выражения, позволяющего осуществлять поиск всех словоформ исследуемой леммы «любовь», т. е. «любовь», «любви», «любовью». Были установлены следующие параметры сортировки: 1 слово справа и 1 слово слева, в центре – интересующая нас лемма «любовь», а также упорядочивание по частоте. Программа

выдала 155 совпадений. Уже на этом этапе мы смогли увидеть определенные закономерности, характерные для анализируемых текстов: например, словоформе «любви» в 8-ми случаях предшествует предлог «в», при этом в 4-х случаях за словоформой «любви» следует союз «и», словоформа «любовью» в 5-ти случаях употребляется с предшествующим ей предлогом «с».

Чтобы выявить, в каких именно словоформах употребляется лексема «любовь», использовался инструмент *Word*. При запуске поискового запроса получены следующие результаты: словоформа «любви» характерна для всех текстов и встречается в них с частотой 71; словоформа «любовь» также характерна для всех текстов и встречается в них с частотой 70; для 5-ти текстов характерна словоформа «любовью», она имеет частоту 10; в 2-х текстах с частотой 2 встречается словоформа «нелюбви»; в одном тексте 1 раз употребляется словоформа «нелюбовь»; в одном тексте 1 раз встречается словоформа «нелюбовью». Таким образом, для исследуемого материала наиболее характерным и частотным является использование лексемы «любовь» в 2-х словоформах: «любви» и «любовь».

Следующим заслуживающим исследовательского внимания функционалом программы *AntConc* является инструмент анализа дисперсионных графиков *Plot*. Графики выглядят как штрихкод и показывают, насколько широко распространено поисковое слово/фраза в загруженных файлах. При поиске с помощью регулярного выражения, установив сортировку выводимых результатов по показателю дисперсии и также определив в настройках инструмента *Plot* показатель дисперсии *Juilland's D*, получаем таблицу, в которой представлена информация о названии файла, количестве токенов в нем, частоте встречаемости поискового слова в конкретном файле, показатель дисперсии, а также график дисперсии. Ориентируясь на этот график, а также на показатель дисперсии, приходим к выводу, что наиболее широко и равномерно словоформы лексемы «любовь» распределены в тексте Д. Морган «Самое красное яблоко». В файле с данным текстом программа определяет 77614 токенов, среди которых словоформы лексемы «любовь» встречаются с частотой 36.

Инструмент *Cluster* программы *AntConc* формирует кластеры определенного размера, заданного пользователем, ориентируясь на положение слева или справа от поискового слова. Например, используя поисковый запрос в виде регулярного выражения, выбирая кластер размером 2, сортируя результаты по частоте, учитывая контексты слева и справа, мы получаем сведения, что для исследуемых текстов характерны следующие кластеры: «любви и» (данный кластер встречается в 8 текстах с частотой 16), «любви к» (встречается в 6 текстах с частотой 9), «любовь и» (встречается в 5 текстах с частотой 8), «в любви» (встречается в 3 документах с

частотой 8) и т. д. Всего программа выделила 232 типа кластеров, учитывая даже те, которые встречаются 1 раз в одном документе. Чтобы уменьшить количество результатов, можно воспользоваться параметром *Min. Range*: например, если установить показатель 5, то программа будет выводить только те кластеры, которые встречаются не менее чем в 5 текстах.

Инструмент *Collocate* позволяет исследователю отследить, как используется запрашиваемое слово в контексте слева и справа. При этом программа находит те слова, которые формируют устойчивые словосочетания с запрашиваемым словом. В параметрах настройки можно устанавливать размер окна (*Window Span*), чтобы уточнить, на сколько слов справа и слева необходимо обращать внимание. Имеет значение и сортировка по различным показателям, например *Likelihood* или *Effect*. Показатель *Likelihood* – это мера вероятности, статистический показатель того, что слова появляются рядом статистически наиболее вероятно, нежели случайно. Показатель *Effect* указывает на силу словосочетания. Так, используя поиск с помощью регулярного выражения, учитывая контекст в размере 1 слово справа и 1 слово слева, сортируя результаты по показателю *Likelihood*, устанавливая настройки инструмента *Collocate* по умолчанию (это Log-likelihood (4-term), $p < 0,05$ (3,84 with Bonferroni), MI), получаем на рабочем экране результаты в виде таблицы. Так, наиболее вероятным коллокатом для словоформ лексемы «любовь» является предлог «к», примыкающий справа. Это характерно для 8 документов, в которых указанный предлог примыкает в 15 случаях, причем в 2 случаях речь идет о нелюбви: *Дело в Ноябре и его нелюбви к золотым оберегам; Сама Марья топала за мною следом – инстинктивное любопытство оказалось сильнее ее нелюбви к матери*. При этом можно обратить внимание и на то, что для всех анализируемых текстов характерно употребление союза «и», который примыкает к словоформам лексемы «любовь» 30 раз (5 раз справа, 25 раз слева). Сочинительной связью с понятием «любовь» в текстах объединены преимущественно позитивные лексемы «мир», «признание», «признательность», «радость», «согласие», «страсть», «тепло», «теплота», «уважение», «благодарность», «верность», «восторг», «поддержка», «рассудок», «добро», «вера», «память», «удача», «покой» (исключение – «горе»). Устанавливая сортировку по показателю *Effect*, мы обнаруживаем, что наибольшей силой обладают словосочетания с разнообразными коллокатами «безответная», «неразделенную», «воспевали», «несчастливой», «познав», «завоевала», «оставшейся», «материнскую», «загубленной», «символическую», «братской», «материнская», «традициями», «несчастлива», «тюрьмой», «сорили». При этом наибольшими показателями – *Likelihood* 11,963, *Effect* 30,403 – обладает коллокат «материнская», т. е. максимальной силой и статистической значимостью обладает словосочетание

«материнская любовь», например: *Мне хочется верить, что, зная она будущее, она никогда не рассталась бы с дочерью, но чего стоит материнская любовь, когда речь идет о деньгах?*

Таким образом, использование программы AntConc в рамках концептуального анализа позволило выявить частотность употребления репрезентантов концепта «любовь» в текстах сказок и фэнтези, зафиксировать конкретные словоформы, характерные для языкового материала, выявить широту и равномерность распределения в нем лексемы «любовь», определить кластеры и коллокаты. Инструментарий сервиса также позволил быстро сформировать конкорданс, который может применяться в автоматическом режиме по алгоритмам сервиса не только для обнаружения разного рода закономерностей, присущих материалу нашего исследования, но и для реализации других задач когнитивной лингвистики.

Библиографические ссылки

1. McEnery T. Corpus linguistics : method, theory and practice / Tony McEnery, Andrew Hardie. Cambridge University Press, 2012.
2. Плунгян В. А. Корпус как инструмент и как идеология : о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 2 (16). С. 7–20.
3. Яковлев А. А. Корпус как универсальный информант (об экспериментальном изучении семантики и языковой картины мира) // Вестник ТвГУ. Серия «Филология». 2017. № 2. С. 64–73.
4. Шилихина К. С. Дискурсивная практика иронии: когнитивный, семантический и прагматический аспекты : автореф. дис. ... д-ра филол. наук : 10.02.19 / Воронежский госуниверситет. Воронеж, 2014.
5. AntConc. [Электронный ресурс]. URL: <https://laurenceanthony.net/software/antconc/> (дата обращения: 08.02.2024).
6. Палийчук Д. А. Корпусные технологии в изучении коллокаций (на примере сервисов «AntConc» и «SketchEngine») // Studia Humanitatis. 2022. № 2. [Электронный ресурс]. URL: www.st-hum.ru (дата обращения: 08.02.2024).
7. Богоявленская Ю. В. Применение сервиса AntConc в исследовании синтагматической сочетаемости (на примере адъективных коллокаций с компонентами «coronavirus» и «COVID-19» во французском языке) // Вестник ПНИПУ. Проблемы языкознания и педагогики. 2022. № 1. С. 30–39.
8. Лунькова Е. С. Использование программы ANTCONC 3.2.4W для анализа художественного текста // Системы компьютерной математики и их приложения. 2014. № 15. С. 260–262.