

УДК 81'42

DOI 10.47388/2072-3490/lunn2021-53-1-82-95

СРАВНИТЕЛЬНО-СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ КОРПУСНЫХ ИНСТРУМЕНТОВ (на примере работы с корпусами кинодискурса)

Н. А. Шамова

Ивановский государственный университет, Иваново, Россия

Широкая сфера применения корпусных технологий указывает на их значимость в прикладной лингвистике. Целью статьи является сравнение и сопоставление основных корпусных инструментов программ *Sketch Engine*, *AntConc*, *WordSmith Tools* на материале текстов специализированных периодических изданий по кинематографу *Total Film* и *American Cinematographer* 2019–2020 годов. Сравнение проводится с целью представления рекомендаций о выборе инструментов и программ для получения определённого вида информации. Общий объём текстов составляет более 900 000 слов. Сравнительно-сопоставительный метод и метод компьютерного анализа являются основными в исследовании. Функции «конкорданс», «список слов», «коллокации» + «описание слов», «N-граммы», «ключевые слова» находятся в программах *Sketch Engine* и *AntConc*. Программа *WordSmith Tools* имеет только функции «конкорданс», «список слов» и «ключевые слова». Данные о наличии инструментов в корпусах представлены в таблице. Описываемые в статье функции разных программ выполняют одинаковые задачи, но наблюдаются некоторые отличия в способах представления данных. Платформа *Sketch Engine* даёт больше возможностей для выбора персональных настроек среди рассматриваемых программ. Конкорданс показывает слово в контексте, список слов включает все слова с фиксацией частотности их употребления в корпусе. Функция «коллокации» (или «описание слов») определяет устойчивые выражения, «N-граммы» находит словосочетания, состоящие из определённого количества элементов. Функция «ключевые слова» позволяет определять слова, характерные для определённой предметной области. На основе полученных сведений из корпусов возможно обновление англоязычных *LSP* словарей и глоссариев по кинематографу. Теоретическая значимость результатов заключается в систематизации материала о корпусных инструментах. Практическая значимость настоящего исследования состоит в изучении кинодискурса при помощи инструментов трёх корпусных программ. Под кинодискурсом в данном исследовании понимается язык, употребляемый сообществом кинозрителей и кинодеятелей про кинематограф в периодических специализированных изданиях *Total Film* и *American Cinematographer*.

Ключевые слова: корпусная лингвистика; специализированный корпус; компьютерный анализ; корпусные инструменты; кинодискурс.

A Comparative-Contrastive Analysis of Corpus Tools (Case Study: Working with Corpora of Cinematic Discourse)

Nadezhda A. Shamova
Ivanovo State University, Ivanovo, Russia

A broad scope of application of corpus technologies indicates their importance in applied linguistics. Employing the comparative-contrastive method and the method of computer analysis, the author seeks to compare and contrast the main corpus tools of the programs Sketch Engine, AntConc, and WordSmith Tools, focusing on texts from specialized periodicals about cinematography *'Total Film'* and *'American Cinematographer'* for 2019–2020. The primary goal of this comparison is to provide recommendations for optimal choice of tools and programs for obtaining certain types of information. The author processed the total volume of texts that contained over 900,000 words, using the functions “concordance”, “word list”, “collocations” + “word sketch”, “N-grams”, “keywords” in Sketch Engine and AntConc (WordSmith Tools has only “concordance”, “wordlist,” and “keywords”). Information about specific tools available in various corpora is collected and presented in a specially developed table. Different software programs described in the article have functions that perform the same tasks, but there are some differences in how data is presented. Among the software programs featured in this case study, the Sketch Engine platform gives the most options for choosing personal settings. The “concordance” function shows the word in context, “Wordlist” shows all the words on a given list with a record of their frequency in the corpus. The “collocation” function (or “word sketch”) recognizes fixed expressions, “N-grams” finds phrases that comprise a certain number of elements, while the “keywords” function allows users to identify words that are specific to a particular subject area. Information thus obtained from the corpora may be helpful in updating English LSP dictionaries and glossaries of cinematography. The theoretical significance of the present study lies in systematizing the material about existing corpus tools, while its practical value is in using the tools of three corpus programs for the study of cinematic discourse, understood here as language used by the community of movie goers and filmmakers in their discussions of cinematography in specialized periodicals *'Total Film'* and *'American Cinematographer'*.

Key words: corpus linguistics; specialized corpus; computer analysis; corpus tools; cinema discourse.

1. Введение

В эпоху научно-технического прогресса и разнообразных информационных технологий увеличивается значимость корпусных инструментов, необходимость применения которых связана с потребностью в проведении многоаспектных и междисциплинарных исследований различных видов дискурса, в частности кинодискурса.

В. А. Плунгян подчеркивает, что «корпус — это не только мощный инструмент исследования языка, но и новая идеология, ориентирующая исследователя на текст как главный объект теоретической рефлексии» (Плунгян 2008: 14). В. Тойберт полагает, что корпусная лингвистика расширяет знания о языке путём объединения трёх различных подходов: рас-

познавание языковой информации (определение её по категориям), поиск взаимосвязи языковой информации с помощью статистических методов и, наконец, представление результатов (Teubert 2001: 128–129). Проведение более качественного и точного лингвистического анализа становится доступным, в том числе за счёт автоматизации процессов и богатого функционала корпусов.

К корпусным инструментам активно обращаются, например, при рассмотрении способов анализа текстов и лексических единиц (Павловская, Горина 2017), проведении критического дискурс-анализа (O'Halloran 2009), составлении частотного списка слов (Toriida 2016). Корпусные технологии применяются при анализе синонимов-глаголов (Hu, Yang 2015), синонимичных английских прилагательных (Басинская 2019), при выявлении ключевых слов в определённой предметной области (Skorczynska, Carrió-Pastor 2015), терминологии (Гросс 2017), модальных частиц (Абдуллаев 2015), метафор (Солопова, Валиева, Ворошилова 2017), а также лексических средств передачи отрицательной оценки (Ковязина, Ильющеня, Хабибуллина 2018). Ими пользуются при установлении особенностей кинодискурса (Шамова 2016) и способов словообразования терминологии кинодискурса (Шамова 2019), а также гендерных стереотипов в языке фильма (Busso, Vignozzi 2017).

Целью работы является проведение сравнительно-сопоставительного анализа основных корпусных инструментов: «конкорданс» (*concordance*), «список слов» (*word list*), «коллокации» (*collocations*) + «описание слов» (*word sketch*), «N-граммы» (*N-grams*), «ключевые слова» (*key words*) — на примере таких программ, как: *Sketch Engine*, *AntConc*, *WordSmith Tools*. Выбор данных программ обуславливается возможностью загружать данные и создавать свои исследовательские корпусы на их основе. Подчеркнём, что в данной работе рассматриваются возможности базовой версии программы *Sketch Engine*. Задачи включают в себя создание исследовательских корпусов для анализа материала, а также анализ доступных инструментов в корпусах. Учитывая многообразие исследований, проводимых на платформе *Sketch Engine*, ожидается, что она обладает наиболее высокой функциональностью. Описание и сравнение современных корпусных инструментов определяет теоретическую значимость работы. На основе полученных результатов сопоставления представлены выводы о наиболее оптимальных инструментах для анализа лингвистического материала. Включение наглядных примеров из кинодискурса, сопровождающих перечисление технических возможностей корпусных программ, свидетельствует о практической значимости настоящей работы и её новизне. Актуальность определяется увеличивающимся количеством корпусных исследова-

ний и возникающей необходимостью в выборе наиболее оптимальных электронных инструментов для работы с лингвистическим материалом, исходя их поставленных целей.

2. Характеристика материала и методов исследования

Настоящее исследование проводилось на материале авторского корпуса, в основу которого положены выпуски специализированных журналов по кинематографу *Total Film* и *American Cinematographer* за 2019–2020 гг. Выбранные выпуски отражают современный язык предметной области «кинематограф» и содержат реальные примеры языкового использования. Общий объём исследовательского корпуса составляет более 900000 слов. При обзоре корпусных инструментов использовался сравнительно-сопоставительный метод. Для получения достоверных результатов при обработке информации из корпуса применялся метод компьютерного анализа. Отметим, что некоторые иллюстративные примеры в данной работе сопровождаются числовым обозначением в скобках, которое указывает на частоту встречаемости, выраженную в абсолютных единицах.

Объектом проводимого исследования являются программы *AntConc*, *WordSmith Tools* и *Sketch Engine*. В свою очередь предметом исследования предстают функции *concordance*, *word list*, *collocations* + *word sketch*, *N-grams*, *key words*, которые выполняют основные задачи, необходимые для автоматизированного качественного и количественного анализа лингвистического материала. Описание функционала сопровождается наглядными примерами, характеризующими современное состояние англоязычного кинодискурса. Важно подчеркнуть, что под кинодискурсом в данной работе понимается язык, употребляемый сообществом кинодеятелей и кинозрителей про кинематограф в периодических специализированных изданиях *Total Film* и *American Cinematographer*.

3. Результаты исследования и их обсуждение

Корпусы XXI века оснащены разными видами инструментов, удовлетворяющими потребности исследователя в получении определённого вида упорядоченной информации. Кроме того, возможность выбора вариантов её представления, исходя из индивидуальных предпочтений пользователя, свидетельствует о том, что данные технологии отличаются антропоцентричным характером. Сравнение и сопоставление инструментов (*concordance*, *word list*, *collocations* + *word sketch*, *N-grams*, *key words*) осуществляется на основе корпусных программ *Sketch Engine*, *AntConc*, *WordSmith Tools* (см. Рис. 1).

| Название программы Название функции | Sketch Engine | AntConc | WordSmith Tools |
|--|---------------|---------|-----------------|
| Concordance | + | + | + |
| Word list | + | + | + |
| Collocations + word sketch | + | + | - |
| N-grams | + | + | - |
| Key words | + | + | + |

Рис. 1. Сравнение функционала корпусных программ

В настоящее время при изучении языка активно используются корпусные менеджеры. «Специализированные программные средства — корпусные менеджеры предоставляют разнообразные возможности по получению из корпуса необходимой информации» (Захаров 2005: 10).

Конкорданс (*concordance* или *key word in context*) позволяет непосредственно увидеть функционирование исходной единицы в контексте и особенности её употребления. Исследуемая единица, как правило, располагается в центре строки и окружается с двух сторон контекстом. Объём контекста может варьироваться и устанавливаться исследователем в зависимости от поставленной цели. П. Бэйкер также отмечает, что в ходе анализа конкорданса можно проследить семантику предпочтительного выбора (*semantic preference*) (Baker 2006: 86). Её определяют как отношение не между отдельными словами, а между леммами или словарными формами и группой семантически связанных слов (Stubbs 2001: 65). Исследователи подчеркивают, что корпусные технологии устанавливают грамматическое окружение слова или фразы (*colligation*) (Cheng 2012: 82).

При работе с функцией конкорданс (*concordance*) на платформе *Sketch Engine* можно выбрать формат *KWIC* (*key word in context* — «ключевое слово в контексте») или опцию *Sentence* (предложение). В формате *KWIC* устанавливаются ограничения по количеству слов справа и слева от искомого слова. Опция *Sentence* позволяет видеть искомое слово в качестве части целого предложения, а расширенный контекст при необходимости увеличивается. Программа *WordSmith Tools* также предоставляет расширенный контекст. В свою очередь в функции «конкорданс» корпусного менеджера *AntConc* возможно как установление ограничений по количеству слов справа и слева от искомого слова, так и одновременно их цветовое выделение. Вместе с тем работа с расширенным контекстом не предполагает ограничения его по объёму. При двукратном нажатии на интересующее слово пользователь автоматически переходит в функцию *file view*, ко-

торая отображает весь текстовый файл, в котором встречается данное слово. Отметим, что, несмотря на небольшие отличия в способах представления данных, цель инструмента *concordance* одинакова во всех рассмотренных программах: работа с контекстом.

Рассмотрим несколько примеров из конкорданса со словом *action*, полученных при работе с корпусами. В предложениях 1 и 2 термин *action* обозначает жанр — «драма»:

(1) *...the filmmakers shot in the style of an action film* (WordSmith Tools).

(2) *We wanted to go wider than Hollywood action films...* (AntConc).

В примере 3 речь идёт о новом гибридном жанре *action crime-comedy*:

(3) *...crashing into cinemas later this year, he has action crime-comedy...* (Sketch Engine).

Интересной представляется также лексическая единица *Action!*, которая имеет значение «мотор». Акцентируем внимание на том, что в данном случае рассматриваемое слово сопровождается восклицательным знаком, который выполняет интонационно-экспрессивную функцию.

Благодаря данному инструменту удалось установить клише, существующие в кинодискурсе, на основе анализа грамматического окружения слова. Так, при работе с функцией «конкорданс» в *AntConc* со словом *script* были получены следующие клише: *script was supervised, script was written, script was quickly snapped up*; со словом *plot*: *the plot lacks, the plot unfolds, the plot settles into, the plot follows, the plot uses* (AntConc). Отметим, что со словом *script* глагол чаще употребляется в пассивном залоге, а со словом *plot*, напротив, в активном.

Важно подчеркнуть, что в *AntConc* есть дополнительная опция *concordance plot*, позволяющая узнавать частотность встречаемости слова в каждом загружаемом в программу документе. Сведения предстают в виде штрих-кода, на котором вертикальные черные полосы указывают на место расположения слова. Данная опция позволяет анализировать особенности положения любого слова в каждом отдельном файле, загруженном в систему.

Похожей на функцию *concordance* является функция «списки заглавных слов» (*key words, keyword list*), но в отличие от неё лексическая единица представляется вне контекста. У пользователя есть возможность выделить часто встречаемые слова в корпусе в сравнении со справочным корпусом (*reference corpus*). Т. В. Толстова обращает внимание на то, что списки ключевых слов (*key-word lists*) «включают элементы, обладающие статистически значимой высокой частотностью — положительные ключевые

слова (*positive key words*) или низкой частотностью — отрицательные ключевые слова (*negative key words*)» (Толстова 2018: 135–136).

Функция *key words* на платформе *Sketch Engine* находит односложные слова и слова, состоящие из нескольких компонентов. Программа *WordSmith Tools* при выделении «ключевых слов» позволяет получить *key words plot*. Информация предстаёт в виде изображения, которое похоже на штрих-код, получаемый в рамках опции *concordance plot* в *AntConc*, на котором вертикальные чёрные полосы обозначают место нахождения слова в выпуске журнала. В свою очередь программа *AntConc* в рамках опции *key word list* находит односложные слова, но представляет их в виде списка. Разные способы представления данных указывают на различные технические возможности программ. Подчеркнём, что программы *AntConc* и *WordSmith Tools* не имеют встроенных корпусов текстов, в отличие от программы *Sketch Engine*, которая в качестве справочного корпуса может, например, обращаться к корпусу *English Web 2015 (enTenTen15)*, содержащему 13 млрд слов, что указывает на его репрезентативность и представительность.

При работе со *Sketch Engine* были получены следующие ключевые слова: *film* (4902), *cinematographer* (2251), *camera* (2166), *shot* (1955), *cinematography* (593), *filmmaker* (458), *cinema* (443), *Hollywood* (417), *stunt* (127), *noir* (67) и т. д. (*Sketch Engine*). Кроме того, были выявлены словосочетания (*multi-words*): *camera operator* (70), *film school* (64), *movie genius* (50), *aspect ratio* (50), *running time* (48), *camera assistant* (37), *key light* (27), *spoiler alert* (26) и т. д. (*Sketch Engine*). Эти сведения могут быть полезны при компиляции учебного глоссария по киноиндустрии, цель которого сформировать общее представление о сфере кино.

Вместе с тем список слов (*word list*) — это список всех слов в корпусе с фиксацией их частотности и процентного соотношения, которое каждое слово вносит в корпус (Baker 2006: 51). Частотные списки (*frequency list*) указывают на часто встречаемые единицы (Толстова 2018: 133). Учёные используют термин *raw frequency*, когда говорят об общей частотности слова в корпусе (Biber, Conrad, Reppen 1998: 263). Исследователи полагают, что нормализованная частотность подразумевает количество употреблений слова на 1 млн слов (Biber, Conrad, Reppen 1998: 34).

На платформе *Sketch Engine*, в корпусной программе *AntConc* и в программе *WordSmith Tools* функция, составляющая список слов, называется *word list*. В программе *AntConc* есть дополнительные опции, заключающиеся в возможности выбрать способ представления результатов. Автоматически составляемый список слов может быть сформирован: по частотности (*sort by freq*), по алфавиту, ориентируясь на первую букву в слове

(*sort by word*), по алфавиту, ранжируя слова по последней букве в слове (*sort by word end*). Данная возможность указывает на антропоцентричность программы, её ориентированность на запросы пользователя. В программе *WordSmith Tools* упорядочивание слов осуществляется по частотности (*frequency*) и по алфавитному принципу (*alphabetical*) без дополнительного деления данной опции (как в *AntConc*). В *Sketch Engine* можно выбрать элементы (буквы или буквосочетания), которые исследователь желает видеть в формируемом списке слов, что также демонстрирует стремление удовлетворить потребности пользователя в получении определённого типа данных. Например, в *Sketch Engine* был получен список терминов, связанных с монтажом при вводе слова *cut*: *cutter*, *cutting*, *cutscene*; корня слова *assemble*: *assembling*, *assembler*, *assemblage* и т. д. (*Sketch Engine*). Тематические списки (списки однокоренных слов) могут найти применение при компиляции современного англоязычного глоссария по кинематографу. При работе с функцией *word list* был разработан частотный список глаголов, характерных для кинодискурса, например: *shoot* (2007), *create* (785), *play* (761), *lead* (455), *capture* (357) и т. д. (*Sketch Engine*).

Благодаря ранжированию слов по алфавитному принципу в *AntConc*, удалось сделать интересное наблюдение на основе анализа списка слов, относящихся к кинематографическим профессиям. В корпусе зафиксирована профессия «каскадер(ы)» (соответственно *stuntman*, *stuntmen* в английском языке) (*AntConc*). Кроме того, была зарегистрирована форма *stuntpeople*. (*AntConc*). Представленный пример указывает на то, что в кинематографии каскадерами наравне с мужчинами могут быть и женщины.

Отметим наличие функции *word sketch* на платформе *Sketch Engine*, которая позволяет автоматически получать сводные сведения о грамматическом или коллокационном поведении слов (Kilgariff, Rychlý, Smrž, Tugwell 2004). М. В. Хохлова определяет *word sketches* как определённые списки, в которых содержится информация о «лингвистическом поведении» — сочетаемости с другими словами с количественным указанием силы связи, которая рассчитывается на основе известных мер ассоциации» (Хохлова 2012). Функция коллокации (*collocations*) также выполняет задачу, связанную с определением устойчивых выражений. Т. Ю. Павельева акцентирует внимание на том, что «корпус текстов представляет неограниченные возможности для изучения и выявления коллокаций» (Павельева 2016: 58). О. Г. Горина подчеркивает, что «коллокации являются ключевыми концепциями корпусных исследований» (Горина 2011: 33).

В *Sketch Engine* функция, которая предлагает пользователю узнать коллокационные особенности слов, называется *word sketch*. Полученная информация систематизируется по колонкам в соответствии с классифи-

цирующим признаком. Например, в один столбец могут входить слова перед анализируемым прилагательным (*modifiers of...*), а в другой войдут существительные после прилагательного (*nouns modified by...*) и т. д. Безусловно, данные сведения полезны для лингвистического анализа. В свою очередь в программе *AntConc* есть функция, выполняющая такую же задачу, но имеющая другое название, а именно *collocates*, в рамках которой представляется общий список коллокаций. Возможно установление минимальной частотности встречаемости слов при работе с данной функцией. Со словом *camera* были зафиксированы следующие наименования: *handheld camera*, *virtual camera*, *crane-mounted camera*, *pocket cinema camera*, *compact camera*, *dollying camera*, *multiple camera*, *dual camera* и т. д. (*AntConc*).

На основе работы с функциями *word sketch* и *collocates* можно сделать вывод, что терминология киноиндустрии заметно расширяется, в том числе за счёт вхождения разнообразных терминов для наименования светового освещения, что указывает на расширение технического оснащения в кинематографе. Например, при работе со *Sketch Engine* были выявлены словосочетания: *ambient light*, *main light*, *overhead light*, *white light*, *hard light*, *soft light*, *practical light* и т. д. (*Sketch Engine*). Как правило, перед существительным *light* в данных случаях используется прилагательное. Интересным представляется следующий пример. Прилагательное *legendary* часто сопровождается существительными, обозначающими профессии: *cinematographer*, *director*, *stunt coordinator*, *production designer* (*Sketch Engine*). Таким образом, опция *word sketch* в *Sketch Engine* позволяет распределять результаты по столбцам на основе классифицирующего признака, а в *AntConc* инструмент *collocates* предлагает общий список устойчивых выражений. В программе *WordSmith Tools* функция *word sketch* или *collocates* отсутствует.

В современных корпусах доступно получение списков словосочетаний (*N-grams*), которые В. Ченг называет *clusters*, *bundles*, *chunks* (Cheng 2012: 72). Т. В. Толстова определяет N-граммы как «непрерывные последовательности элементов (количество элементов обозначается буквой N), встречающиеся в тексте или корпусе» (Толстова 2018: 138).

На платформе *Sketch Engine* пользователь может установить желаемое количество элементов, входящих в словосочетание (от 2 до 6) при работе с функцией *N-grams*. В свою очередь в программе *AntConc* такая же функция позволяет выбрать минимальное и максимальное количество элементов, причём максимальное количество является неограниченным. Таким образом, *AntConc* (в сравнении со *Sketch Engine*) не ставит ограничений перед пользователями. Подчеркнём, что в *WordSmith Tools* данная

функция отсутствует. При работе с функцией *N-grams* в *Sketch Engine* было установлено ограничение в 4 слова и были зафиксированы выражения, включающие превосходную степень, например, *The world's most trusted* (134), *one of the most* (58) (*Sketch Engine*). Превосходная степень характерна для кинодискурса и используется для привлечения внимания к киноиндустрии (в данном случае — читателей журналов).

При описании функциональности корпусных инструментов особое внимание следует уделить их антропоцентричному характеру. Платформа *Sketch Engine* предлагает наибольший выбор в способах представления данных (среди рассматриваемых корпусных программ) в рамках опций *change view options*, *view options* и т. д. Например, по желанию пользователя данные могут быть представлены в одной или нескольких колонках, визуализация может осуществляться с помощью разных графических средств и т. д.

4. Выводы

В программах *Sketch Engine* и *AntConc* представлены функции *concordance*, *word list*, *collocations + word sketch*, *N-grams*, *key words*. В *WordSmith Tools* представлены только *concordance*, *word list* и *key words*. Корпусные инструменты помогают успешно обрабатывать, упорядочивать и систематизировать информацию. Так, они позволяют анализировать поведение слова в контексте, устанавливать частотность определённых слов в корпусе, выявлять особенности их грамматического или коллокационного поведения, получать списки словосочетаний и часто встречаемых слов в корпусе в сравнении со справочным корпусом, который является более масштабным и, следовательно, репрезентативным.

Сравнив функционал программ *AntConc*, *WordSmith Tools* и *Sketch Engine* на примере работы с корпусом кинодискурса, приходим к выводу, что, несмотря на наличие в разных программах функций, выполняющих одинаковое назначение, существует определённая специфика, связанная с их техническими возможностями и способами представления данных.

Результаты сравнения инструментов позволяют выбирать наиболее оптимальные из них при анализе кинодискурса. Так, для получения «ключевых слов» использовалась платформа *Sketch Engine*, предлагающая встроенный репрезентативный справочный корпус. При выявлении прилагательных, согласующихся с существительным, целесообразнее обратиться к функции *word sketch*, которая систематизирует результаты по колонкам в зависимости от классифицирующего признака и которая представлена также в *Sketch Engine*. При формировании списка слов, содержащего определённые элементы, подходит функция *word list* в *Sketch Engine*. При поис-

ке словосочетаний, содержащих более 6 слов, рекомендуется использовать опцию *N-grams* в *AntConc*. Работа с контекстом в рамках инструмента *concordance* может осуществляться при помощи любой из рассмотренных программ. Кроме того, было установлено, что платформа *Sketch Engine* представляет собой не только источник, позволяющий работать с корпусными инструментами, но и ресурс, дающий выбор способа представления данных, что удовлетворяет потребности современного пользователя.

При корпусном анализе материала стало возможным выявить некоторые особенности кинодискурса, а именно: установить значение многозначной лексической единицы, исходя из контекста, клише, наиболее частотные глаголы, устойчивые выражения, словосочетания с превосходной степенью, наименования для кинопрофессий, камер, светового освещения, а также однокоренные слова и наиболее характерные слова для англоязычного кинодискурса. Полученные результаты могут использоваться при обновлении специализированных словарей и глоссариев по кинематографу.

Безусловно, для объективного и всестороннего рассмотрения предметной области «киноиндустрия» (с лингвистической точки зрения) необходимо, на наш взгляд, совместное использование различных корпусных программ и опций в рамках нескольких исследований.

Список литературы / References

- Абдуллаев Р. Ф. Модальные частицы в речи мужчин и женщин // Известия Волгоградского государственного педагогического университета. 2015. Вып. 9–10 (104). С. 161–166. [Abdullaev, Rojal F. (2015) Modal'nye chasticy v rechi muzhchin i zhenshhin (Modal Particles in Speech of Men and Women). *Izvestia of the Volgograd State Pedagogical University*, 9–10 (104), 161–166. (In Russian)].
- Басинская М. В. Корпусный анализ семантики синонимов (на материале английских прилагательных) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2019. № 6 (822). С. 66–78. [Basinskaja, Marija V. (2019) Korpusnyj analiz semantiki sinonimov (na materiale anglijskih prilagatel'nyh) (Corpus-based Analysis of Synonyms (a Case Study of English Adjectives). *Vestnik of Moscow State Linguistic University. Humanities*, 6 (822), 66–78. (In Russian)].
- Горина О. Г. Применение методов корпусной лингвистики для определения контекстно-специфических слов и коллокаций // Вестник Ленинградского государственного университета им. А.С. Пушкина. 2011. № 3. Т. 7. С. 27–36. [Gorina, Ol'ga G. (2011) Primenenie metodov korpusnoj lingvistiki dlja opredelenija kontekstnospecificeskikh slov i kollokacij (Implementation of Corpus Linguistic Techniques to Identify a Text or Genre-specific Key Words and Collocations). *Pushkin Leningrad State University Journal*, 3, v. 7, 27–36. (In Russian)].
- Гросс М. А. Автоматизированное выделение терминологии из текстов договоров на английском языке при помощи средств корпусной лингвистики // Омский научный вестник. Серия «Общество. История. Современность». 2017. № 1. С. 33–36.

- [Gross, Milena A. (2017) Avtomatizirovannoe vydelenie terminologii iz tekstov dogovorov na anglijskom jazyke pri pomoshhi sredstv korpusnoj lingvistiki (Automated Selection of English Contract Terminology by Means of Corpus Linguistics). *Omsk Scientific Bulletin. Series "Society. History. Modernity"*, 1, 33–36. (In Russian)].
- Захаров В. П. Корпусная лингвистика: Учебно-методическое пособие. СПб: СПб. гос. ун-т, 2005. [Zaharov, Viktor P. (2005) *Korpusnaja lingvistika: Uchebno-metodicheskoe posobie*. (Corpus Linguistics: Textbook). Saint Petersburg State University. (In Russian)].
- Ковязина М. А., Ильющеня Т. А., Хабибуллина С. Б. Средства выражения отрицательной оценки при описании санкционной политики западных стран в отношении России и российских ответных санкций в текстах британских сетевых СМИ: корпусное исследование // Политическая лингвистика. 2018. № 2 (68). С. 61–67. [Kovjazina, Marina A., Il'jushhenja, Tat'jana A. & Habibullina, Sajda B. (2018) Sredstva vyrazhenija otricatel'noj ocenki pri opisanii sankcionnoj politiki zapadnyh stran v otnoshenii Rossii i rossijskih otvetnyh sankcij v tekstah britanskih setevykh SMI: korpusnoe issledovanie (Negative Evaluative Language in Description of Anti-Russian Sanctions Policy and of Russian Counter-sanctions in British Online Media Discourse: a Corpus Research). *Political Linguistics*, 2 (68), 61–67. (In Russian)].
- Павельева Т. Ю. Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2016. Вып. 3–4 (155–156). Т. 21. С. 56–61. [Pavel'eva, Tat'jana Y. (2016) Izuchenie kollokacij na osnove lingvisticheskikh korpusov tekstov (Collocations Study Based on the Text Corpora). *Tambov University Review. Series: Humanities*, 3–4 (155–156), v. 21, 56–61. (In Russian)]. DOI: 10.20310/1810-0201-2016-21-3/4(155/156)-56-61.
- Павловская И. Ю., Горина О. Г. Корпусно-когнитивные методы изучения лексической стороны речи на иностранном языке // Вестник Череповецкого государственного университета. 2017. № 1 (76). С. 132–138. [Pavlovskaja, Irina Y. & Gorina, Ol'ga G. (2017) Korpusno-kognitivnye metody izuchenija leksicheskoy storony rechi na inostrannom jazyke (Corpus-based Cognitive Oriented Methods of Lexical Analysis in Foreign Language Studies). *Cherepovets State University Bulletin*, 1 (76), 132–138. (In Russian)]. DOI: 10.23859/1994-0637-2017-1-76-19.
- 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. [Plungjan, Vladimir A. (2008) Korpus kak instrument i kak ideologija: o nekotoryh urokah sovremennoj korpusnoj lingvistiki (Corpus as a Tool and as an Ideology: about some Lessons of Modern Corpus Linguistics). *Russian Language and Linguistic Theory*, 2 (16), 7–20. (In Russian)].
- Солопова О. А., Валиева Г. Р., Ворошилова М. Б. Корпусное исследование метафор со сферой-мишенью «Россия» (на материале современного американского дискурса) // Вестник Южно-Уральского государственного университета. Серия: Лингвистика, 2017. № 2. Т. 14. С. 37–43. [Solopova, Ol'ga A., Valieva, Gul'naz R. & Voroshilova, Marija B. (2017) Korpusnoe issledovanie metafor so sferoj-mishen'ju «Rossija» (na materiale sovremennogo amerikanskogo diskursa) (Corpus Analysis of Metaphors with the Target Domain «Russia» (Based on the Material of Modern American Discourse). *Bulletin of the South Ural State University. Series: Linguistics*, 2, v. 14, 37–43. (In Russian)]. DOI: 10.14529/ling170205.

- Толстова Т. В. Жанр и корпус: современные подходы к изучению и преподаванию языка. Самара: Издательство Самарского университета, 2018. [Tolstova, Tat'jana V. (2018) *Zhanr i korpus: sovremennye podhody k izucheniju i prepodavaniju jazyka* (Genre and Corpus: Modern Approaches to the Study and Teaching of the Language). Samara: Izatel'stvo Samarskogo universiteta. (In Russian)].
- Хохлова М. В. Лексико-синтаксические шаблоны как инструмент выявления специальной лексики предметной области [Электронный ресурс] // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной международной конференции «Диалог». 2012. URL: <http://www.dialog-21.ru/media/1392/хохловамв.pdf> (дата обращения 31.08.2020). [Hohlova, Marija B. (2012) *Leksiko-sintaksicheskie shablony kak instrument vyjavlenija special'noj leksiki predmetnoj oblasti* (Lexico-syntactic Patterns as a Tool for Extracting Lexis of a Specialized Knowledge Domain). *Komp'juternaja lingvistika i intellektual'nye tehnologii: materialy ezhegodnoj mezhdunarodnoj konferencii «Dialog»* (Computational Linguistics and Intelligent Technologies: Proceedings of the Annual International Conference "Dialogue") Retrieved from <http://www.dialog-21.ru/media/1392/хохловамв.pdf> (In Russian)].
- Шамова Н. А. Корпусные технологии при анализе кинодискурса (на материале авторского электронного корпуса кино) // Ученые записки Крымского федерального университета имени В. И. Вернадского. Филологические науки. 2016. № 4. Т. 2 (68). С. 109–115. [Shamova, Nadezhda A. (2016) *Korpusnye tehnologii pri analize kinodiskursa (na materiale avtorskogo jelektronnogo korpusa kino)* (Corpus Technologies in the Cinema Discourse Analysis (Based on the Author's Corpus). *Scientific Notes of V.I. Vernadsky Crimean Federal University. Philological sciences*, 4, v. 2 (68), 109–115. (In Russian)].
- Шамова Н. А. Особенности словообразования в сфере терминологии современного английского кинодискурса // Теоретическая и прикладная лингвистика. 2019. № 4. Т. 5. С. 152–162. [Shamova, Nadezhda A. (2019) *Osobennosti slovoobrazovanija v sfere terminologii sovremennogo anglijskogo kinodiskursa* (Word Formation Patterns in Contemporary English Cinema Discourse Terminology). *Theoretical and Applied Linguistics*, 4, v. 5, 152–162. (In Russian)]. DOI: 10.22250/2410-7190_2019_5_4_152_162.
- Baker, Paul. (2006) *Using Corpora in Discourse Analysis*. London: Continuum.
- Biber, Douglas, Conrad, Susan, Reppen, Randi. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Busso, Lucia, Vignozzi, Gianmarco. (2017) Gender Stereotypes in Film Language: a Corpus-Assisted Analysis. In Basili, Roberto, Nissim, Malvina, & Satta, Giorgio. (eds.) *Proceedings of the Fourth Italian Conference on Computational Linguistics CLIC-IT 2017*, 11–12 December 2017, Rome, 71–76. Retrieved from <http://ceur-ws.org/Vol-2006/paper017.pdf> doi: 10.4000/BOOKS.AACCADEMIA.2367.
- Cheng, Winnie (2012). *Exploring Corpus Linguistics: Language in Action*. Routledge.
- Hu, Chunyu, Yang, Bei. (2015) Using Sketch Engine to Investigate Synonymous Verbs. In Zhou, Peng. (eds.) *International Journal of English Linguistics*, 4, 5, 29–41. DOI: 10.5539/ijel.v5n4p29.
- Kilgariff, Adam, Rychly, Pavel, Smrz, Pavel, Tugwell, David. (2004) The Sketch Engine. In Williams, Geoffrey, Vessier, Sandra. (eds.) *Proceedings of Euralex*, 6-10 July 2004, Lorient, France, 105–115. Retrieved from

- https://euralex.org/elx_proceedings/Euralex2004/011_2004_V1_Adam%20KILGARRIFF,%20Pavel%20RYCHLY,%20Pavel%20SMRZ,%20David%20TUGWELL_The%20%20Sketch%20Engine.pdf.
- O'Halloran, Kieran A. (2009) Inferencing and Cultural Reproduction: a Corpus-based Critical Discourse Analysis. In Sarangi, Srikant. (eds.), *Text and Talk*, 1, 29, 21–52. DOI: 10.1515/TEXT.2009.002.
- Skorczynska, Hanna, Carrió-Pastor, María Luisa. (2015) Variation in General Meaning Keywords in Press Releases from British and Spanish Companies: Gaining Deeper Insights into Corporate Discourse. *Procedia — Social and Behavioral Sciences*, 198, 451–458. DOI: 10.1016/j.sbspro.2015.07.465.
- Stubbs, Michael. (2001) Words and Phrases: Corpus Studies of Lexical Semantics. *Blackwell*.
- Teubert, Wolfgang. (2001) Corpus Linguistics and Lexicography. In Mahlberg, Michaela (eds.) *International Journal of Corpus Linguistics*, 3, 6, 125–153. DOI: 10.1075/ijcl.6.3.11teu.
- Toriida, Marie-Claude. (2016) Steps for Creating a Specialized Corpus and Developing an Annotated Frequency-Based Vocabulary List. In Rossiter, Marian. (eds.) *TESL CANADA JOURNAL*, 1, 34, 87–105. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1145032.pdf>. DOI: 1018806/tesl.v34i1.1255.

Источники языкового материала / Language material resources

- AntConc (2020, September 01). Retrieved from <https://laurenceanthony.net/software/antconc>.
- Sketch Engine (2020, September 01). Retrieved from <https://www.sketchengine.eu>.
- WordSmith Tools (2020, September 01). Retrieved from <https://lexically.net/wordsmith/>.