

Дарья Васильевна Дроздова, преподаватель
Минский государственный лингвистический
университет, Минск, Беларусь
э-почта: drozdova.darya.v@gmail.com

Darya Vasilievna Drozdova, Lecturer
Minsk State Linguistic University, Minsk,
Belarus
e-mail: drozdova.darya.v@gmail.com

АНАЛИЗ ПОСТОВ ТЕЛЕГРАМ-КАНАЛА МГЛУ С ИСПОЛЬЗОВАНИЕМ ANTCONC: ВОЗМОЖНОСТИ И РЕЗУЛЬТАТЫ

Выявлены особенности инструмента корпусного анализа текста AntConc. Раскрыты основные тематики постов в официальном телеграм-канале МГЛУ.

Ключевые слова: корпусная лингвистика; AntConc; базы данных; медиaprостранство вуза.

ANALYZING MSLU TELEGRAM CHANNEL POSTS USING ANTCONC: CAPABILITIES AND RESULTS

The features of AntConc text corpus analysis tool are revealed. The main topics of posts in the official telegram channel of MSLU are revealed.

Key words: corpus linguistics; AntConc; databases; university media space.

Целью данного исследования является выявление основных принципов работы с *AntConc* для анализа тематики постов официального телеграм-канала УО «Минский государственный лингвистический университет».

Antconc – это бесплатная, мультиплатформенная программа, представляющая собой инструмент для статистических исследований текстов. Она была разработана профессором Лоуренсом Антони, директором Центра обучения английскому языку в науке и технике Школы науки и техники университета Васеда (Япония) для лингвистов, только начинающих свое знакомство с корпусными технологиями.

Корпус-менеджер *AntConc* предназначен для обработки корпусов первого порядка. С помощью данной программы можно производить поиск и подсчет различных элементов текста, анализировать частотность и контекст употребления словоформ, словосочетаний и морфем, сравнивать употребительность словоформ в разных текстах. Отсутствие морфологического анализатора частично компенсируется возможностью подключения пользовательского списка лемм. Программа может быть использована для получения привязанных к заданной предметной области словарных минимумов, списков устойчивых сочетаний (в том числе терминологических), выборок к тематическим группам слов.

Медиакоммуникационное пространство высших учебных заведений в настоящее время активно расширяется за счет различных интернет-платформ и сервисов. Мессенджер «Телеграм» – это веб-приложение для обмена мгновенными сообщениями, широко распространённое среди современной интернет-аудитории и отличающееся большой скоростью отправки сообщений и высоким уровнем безопасности. Телеграм-каналы – это площадка для распространения информации и контента среди пользователей мессенджера «Телеграм». Исследование постов в мессенджере приобретают актуальность по нескольким причинам: во-первых, телеграм-каналы обеспечивают оперативный доступ и обновление информации; во-вторых, телеграм-каналы охватывают большую аудиторию, представленную различными демографическими группами; мессенджер предлагает также возможность внедрения в посты информации разных форматов.

Материалом нашего исследования являются посты из официального телеграм-канала МГЛУ «МГЛУ Минск» общим объемом 2105 единиц, собранные в датасет. Отобранный материал представляет собой не только текст самих постов, но и дополнительные сведения о них: ссылки, эмоджи и другие графические элементы, которые нежелательны для корпусных исследований, так как не являются знаменательными единицами. Следовательно, данные единицы необходимо убрать из собранного датасета при помощи регулярного выражения языка программирования *Python*. После «очистки» датасета от нежелательных элементов, файл датасета формата *csv* необходимо переформатировать в *txt*, так как *AntConc* создает и обрабатывает корпуса из файлов данного формата.

Для классификации тем постов в первую очередь необходимо выявить ключевые слова для корпуса постов при помощи функции *KWIC* корпусного менеджера *AntConc*. Ключевым словом в корпусной лингвистике является слово, которое встречается чаще в исследуемом корпусе, чем в контрольном (референтном) корпусе, обычно большем по размеру. Таким образом, для выявления ключевого слова в корпусе постов телеграм-канала МГЛУ нам было необходимо загрузить в корпусный менеджер *AntConc* референтный корпус, которым стал открытый корпус русского языка (*OpenCorpora*) объемом 1989538 токенов, численно превалирующий над корпусом постов объемом 73529 токенов.

Ключевыми словами для корпуса постов из телеграм-канала МГЛУ являются: «МГЛУ», «студентов», «преподавателей», «университета», «года», «языка», «факультета», «Беларусь».

Далее, при помощи функции *Concordance Tool* мы можем проанализировать результаты поиска ключевых слов в контексте, для выявления тематики контекстов постов телеграм-канала МГЛУ. Контекстами употребления ключевого токена «МГЛУ» являются участие студентов, преподавателей и ректората в республиканских и международных мероприятиях.

Контекстами употребления токена «студентов» вместе с токеном «преподавателей» являются анонсы мероприятий, проводимых в сотрудничестве, а также информационные объявления. Ключевым словом является также слово «языка», использованное в контексте наименования факультетов, центров МГЛУ.

Кластеры (*Clusters*) в *AntConc* позволяют находить группы слов или фраз, которые являются частотными для анализируемого корпуса текстов. Данная функция может быть использована для выявления устойчивых выражений и коллокаций. В ходе анализа кластеров с ключевым токеном «МГЛУ» были выявлены такие пары как «МГЛУ приглашает», «в МГЛУ прошла», «МГЛУ помнит», «МГЛУ международный», что свидетельствует об активном вовлечении студентов в научную, творческую жизнь университета, почитании традиций, ориентированности университета на взаимодействие с международным сообществом.

Следует отметить, что в корпусном менеджере *AntConc* также есть функция *Concordance Plot*, которая позволяет визуализировать распределение слов в текстах. Это помогает выявить тенденции использования определенных слов. Таким образом, нам удалось определить распределение частотного токена «факультета» – 0.872, что свидетельствует о его значительной употребляемости в анализируемых текстах. Данная информация может быть полезна для дальнейшего анализа, например, для выявления взаимосвязей с другими терминами или для понимания контекста, в котором используется слово «факультет».

Активное развитие новых медиаресурсов вузов приводит к необходимости их изучения не только на уровне социопрагматических, целевых, тематических, содержательных характеристик, но и на уровне лингвистических особенностей. Официальный телеграм-канал «МГЛУ Минск» является атрибутом имиджа УО «Минский государственный лингвистический университет», который себя позиционирует как открытый к международному сотрудничеству, сохраняющий память о важных датах и событиях, сосредоточенный на высококачественном образовании, активном взаимодействии между студентами и преподавателями. Университет стремится быть актуальным и привлекательным для студентов, подчеркивая свои сильные стороны в области лингвистического и языкового образования.

Главными темами постов являются участие студентов и преподавателей в общественной жизни университета и страны в целом, анонсы мероприятий, информационные объявления. Публикации канала «МГЛУ Минск» на платформе Телеграм по своему структурному и стилевому оформлению: имеют мультимедийный и текстовый блоки, отличаются выраженной тезисностью и дробностью текста, интерактивностью и гипертекстовостью.

Мультиплатформенная программа *AntConc* показала свою эффективность при обработке пользовательского корпуса текстов, предоставляя инструменты для анализа языковых данных. Корпусный менеджер позволяет исследовать частотность слов, коллокации, контексты употребления. В результате *AntConc* становится незаменимым инструментом для тех, кто занимается корпусной лингвистикой и языковыми исследованиями.