

ПРЕДСТАВЛЕНИЕ ОБЪЕКТНЫХ СЛОВСОЧЕТАНИЙ В СЛОВАРЯХ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ С КОЛЛОКАЦИЯМИ НА МАТЕРИАЛЕ НКРЯ И СНАТGPT¹

Корпусы текстов стали незаменимы при проведении лингвистических исследований, однако становится очевидно, что большие данные, доступ к которым опосредованно предоставляется системами искусственного интеллекта (ИИ), могут оказывать влияние на корпусные исследования и приводят к новому витку развития в области корпусных технологий. В ряде работ обсуждается, могут ли системы генеративного искусственного интеллекта (такие как, например, ChatGPT) в определенном смысле стать заменой корпусу. Создание конкордансов и частотных списков анализируется в статье [1]. В работе приводятся промпты (специальных запросов к системе ИИ), которые могут быть использованы для построения конкордансов. Для решения данной задачи необходимо задать точную роль или идентичность при работе с системой, а также привести пример желаемого результата и перечислить требования к выходному формату.

Нами разработана база данных коллокаций русского языка [2; 3], в которой представлены устойчивые словосочетания, извлеченные из словарей и корпусов текстов русского языка при помощи автоматических методов. Данные были получены из заромбовой части толковых словарей и из словарей сочетаемости (в них приводятся списки соответствующих единиц). В настоящее время в ней содержится 40 тыс. единиц, которые были зафиксированы в 9 толковых словарях и словарях устойчивых словосочетаний и представлены следующими моделями: «прилагательное + существительное» (*насущенный интерес*), «причастие + существительное» (*наболевший вопрос*), «наречие + прилагательное» (*невыносимо скучный*) и «глагол + существительное» (*извлечь выгоду*).

В докладе будут рассмотрены объектные словосочетания, обладающие высоким словарным индексом в базе данных коллокаций (то есть те единицы, которые зафиксированы в 4 или 5 источниках,

¹ Исследование выполнено за счет гранта Российского научного фонда № 22-18-00189, <https://rscf.ru/project/22-18-00189/>

например, нести ответственность, одержать победу, принять меры и др.) и их представленность в Национальном корпусе русского языка (НКРЯ) [4], а также произведено сравнение с результатами, полученными при использовании ChatGPT в качестве конкорданса [1]. Подкорпус основного корпуса был ограничен текстами с 1950 по 2023 гг. (общий объем 197 159 944 слов). Цель исследования заключается в том, чтобы проанализировать, как представлены в корпусе словарные коллокации, а также исследовать возможности применения системы ИИ для решения лексикографических задач. Предварительные результаты показывают, что данная система ИИ выдает устойчивые словосочетания, зафиксированные в словарях и корпусе, однако ранжирование данных единиц отличается от ранжирования по частоте в основном корпусе НКРЯ.

Примечания

1. Lin P. ChatGPT: Friend or foe (to corpus linguists)? // Applied Corpus Linguistics, 3 (3), 2003. P. 1–5.
2. Хохлова М. В. Атрибутивные коллокации в золотом стандарте сочетаемости русского языка и их представление в словарях и корпусах текстов // Вопр. лексикографии. 2021. № 21. С. 33–68.
3. Хохлова М. В., Мамаев И. Д. Разработка базы данных коллокаций: обзор золотого стандарта на примере атрибутивных словосочетаний // Труды международной конференции «Корпусная лингвистика–2021». СПб.: Скифия-принт, 2021. С. 370–379.
4. Национальный корпус русского языка. URL: <http://ruscorpora.ru> (дата обращения: 20.08.2024).

© Е. М. Черкасова

Уральский федеральный университет,
г. Екатеринбург, Россия

ПРОЦЕСС АКУСТИЧЕСКОГО ВОСПРИЯТИЯ: СПЕЦИФИКА ДЕНОТАТИВНО-ИДЕОГРАФИЧЕСКОГО КОНСТРУИРОВАНИЯ¹

Взаимосвязанные процессы звучания и восприятия звука объединены в синопсисе Уральской семантической школы (о работе над синопсисом пишет Л. Г. Бабенко, см., например, [1]) в один объемный раздел «Акустическое восприятие», который включает в себя следующие денотативно-идеографические группы (ДИГ) [2]: «Процесс акустического восприятия», «Звуки, их разновидности и источник» и «Характеристика звука». Дальнейшей рубрикации в

¹ Исследование выполнено за счет гранта Российского научного фонда, проект № 22-18-00352, <https://rscf.ru/project/22-18-00352/>