

ЛИНГВИСТИЧЕСКИЙ КОРПУС КАК ИНСТРУМЕНТ ГЕНЕРАЦИИ УЧЕБНЫХ МАТЕРИАЛОВ ДЛЯ ФОРМИРОВАНИЯ ПЕРЕВОДЧЕСКИХ КОМПЕТЕНЦИЙ

Рассматривается способ автоматической генерации различных учебных материалов по иностранным языкам. Эти материалы могут быть использованы как отдельно, так и для составления учебников и учебных пособий, целью которых является формирование переводческих компетенций. Инструментом исследования является оригинальное авторское программное обеспечение. В качестве исходного «материала» применяются тексты немецкоязычных электронных СМИ. Делается вывод о целесообразности предлагаемого способа.

Ключевые слова: лингвистический корпус; обучение письменному переводу; специальное программное обеспечение; немецкий язык.

A method for automatic generation of various learning materials in foreign languages is considered. These materials can be used either separately or for compiling textbooks and exercise books, the purpose of which is to develop translation competencies. The research tool is original software. Texts from German-language digital media are used as source material. A conclusion is drawn about the feasibility of the proposed method.

Key words: linguistic corpus; translation training; special software; German language.

Корпусные исследования являются стабильно востребованными со стороны научного сообщества в России и за рубежом, поскольку предоставляют ученым надежный языковой материал уже более полувека.

Лингвистические корпуса применяются в настоящее время в контексте лексикографии [1], фразеологии [2; 3; 4], различных диахронических исследований [5; 6; 7], анализа произведений художественной литературы [8; 9; 10].

Немало исследователей работают на стыке корпусной лингвистики и педагогики. Так, В. В. Изюмская, анализируя научные публикации за 2019-2023 гг., отмечает, что теме «Корпусная лингвистика» внутри направления «Цифровизация образования» посвящены 15 % статей на русском и 20 % английском языках, причем подтема «Корпусы в преподавании перевода» разрабатывается в 20 % русскоязычных публикаций, когда как в англоязычных она не затронута вовсе [11, с. 31].

Цель нашей работы – продемонстрировать возможности оригинального программного комплекса для генерации и оперирования лингвистическим корпусом с формальной морфологической разметкой, созданным по правилам библиотеки обработки естественного языка spaCy [12; 13], для быстрого составления качественных учебных материалов (упражнений) для обучения письменному переводу, в нашем случае – с немецкого языка на русский язык.

Принципы работы генератора и корпусного менеджера были описаны нами в ряде предыдущих публикаций [9; 14], поэтому мы остановимся далее только на тех функциях программного комплекса, которые необходимы в рамках настоящего исследования.

Созданный на материале массива текстов немецкоязычных СМИ лингвистический корпус имеет объем 13120 токенов и способен выводить в результате поискового запроса предложения, содержащие части речи с теми или иными грамматическими характеристиками. Например, глаголы в конъюнктиве или в императиве или прилагательные в превосходной форме.

Поиск возможно также осуществлять по одной или нескольким леммам, добавляя к этому критерию еще и требования к их грамматическим признакам. Другой полезной опцией является поиск по последовательностям, которые задаются формулами, например, «артикль + прилагательное + существительное» или «существительное + N любых слов + глагол». При этом для каждого компонента формулы можно задать параметры (падеж, число, род, временная форма и пр.).

Таким образом, параметры поиска могут складываться, образуя большое сочетание и вариативность поисковых запросов.

Приведем примеры двух типов упражнений, которые в рабочем порядке назовем упражнениями *для формирования мгновенной реакции* и упражнениями *с грамматической доминантой*. Они применяются на очном занятии в режиме «преподаватель-группа».

Первое упражнение заключается в том, что преподаватель называет отдельные существительные, а студенты по очереди должны мгновенно назвать перевод (как правило, первым должен придти в голову самый частый вариант перевода этого существительного). Упражнение выполняется в очень высоком темпе.

Сам характер упражнения может подвергнуться объективной критике со стороны приверженцев того или иного метода обучения, но здесь мы сосредоточимся на процедуре подготовки языкового материала для упражнения. Проблема заключается в том, что необходимо собрать большое количество существительных, отделив их от прочего текста. С помощью корпусного менеджера, осуществив поиск по части речи «существительное», мы получили частотный список всех существительных корпуса в следующем виде:

1. *Jahr* : 45
2. *Euro* : 27
3. *Planet* : 25
4. *Erde* : 24
5. *Prozent* : 22
6. *Wasser* : 18
7. *Satellit* : 18
- ...
1357. *Bord* : 1

(1. 'Город, 2. Евро, 3. Планета, 4. Земля, 5. Процент, 6. Вода, 7. Спутник ... 1357. Борт' – *здесь и далее перевод наш – А.Г., И. Г.*)

Далее посредством небольшой компьютерной программы на Python список «очищается» от цифр, а двоеточия заменяются на запятые; результат выводится в одну строку. На базе этого вывода был сформирован комплект из 44-х упражнений. Приведем ниже первое из них:

Jahr, Euro, Planet, Erde, Prozent, Wasser, Satellit, See, Land, Kilometer, Weiche, Meter, Hai, Tier, Leben, Welt, Stern, Problem, Tag, Elektroautos, Architekt, Fläche, Landschaft, Stadt, Million, Frage, Musik, Mensch, Stunde, Strom, Stelle, Eurozone, Loch, Boden, All, Publikum, Zukunft, Hälfte, Auto, Studierende, Folge, Bagger, Rekultivierung, Gesteinsplanet. ‘Год, евро, планета, Земля, процент, вода, спутник, море, земля, километр, стрелка, метр, акула, зверь, жизнь, мир, звезда, проблема, день, электромобили, архитектор, поверхность, ландшафт, город, миллион, вопрос, музыка, человек, час, поток, место, Еврозона, дыра, почва, вселенная, публика, будущее, половина, машина, обучающийся, последствие, экскаватор, рекультивация, твердокаменная планета’.

Для демонстрации процесса получения упражнений второго рода выберем грамматическое явление «придаточные предложения с подчинительными союзами». Для получения необходимых образцов воспользуемся поиском подчинительных союзов, в результате которого мы получим список искомых предложений, которые мы выделили в 16 блоков по десять единиц. Приведем в качестве примера часть блока № 3:

2. *Wenn man doch ins Wasser geht und einen Hai sieht, hat man gute Chancen, unversehrt davon zu kommen, verspricht Kock.* 6. *Solange der Hai nicht selbst angreift, sollte man das auch nicht selbst tun.* 7. *Wenn der Raubfisch zu nahe kommt, hilft oft ein beherzter Schlag, am besten auf die Augen oder die Kiemen.* 8. *„Wenn man nicht weiß, wo Augen oder Kiemen sind, einfach zuschlagen, bevor man lange sucht.“* 9. *Diesen Plan hatten wir genau so lange im Kopf, bis wir Reichweiten recherchierten: 60 Kilometer, 140 Kilometer, 200 Kilometer – das sind die üblichen Zahlen.* 10. *Falls es ein reines Elektroauto ist, bleibt es danach stehen, ein Hybridauto schaltet auf Benzin um.* 2. ‘Если вы все-таки зайдете в воду и увидите акулу, у вас есть хорошие шансы выбраться невредимым, – заверяет Кок. 6. Если сама акула не нападает, не стоит делать этого и самостоятельно. 7. Если хищник подплывает слишком близко, часто помогает смелый удар, желательно по глазам или жабрам. 8. «Если вы не знаете, где находятся глаза или жабры, просто бейте раньше, чем найдете». 9. Этот план был у нас в запасе, пока мы не исследовали максимальную дальность: 60 километров, 140 километров, 200 километры – это обычные цифры. 10. Если это чисто электрический автомобиль, то он потом останавливается, гибридный автомобиль переходит на бензин’.

Упражнение также выполняется в группе. Стандартно осуществляется перевод, комментируются применяемые трансформации и иные переводческие трудности.

Мы кратко осветили возможности авторского программного комплекса и показали на двух примерах некоторые из них. Описанный метод, безусловно, имеет свои ограничения, но в то же время и позволяет быстро собрать материал на основе заданного набора текстов. На наш взгляд, эффективным можно было бы назвать решение формировать лингвистический корпус для последующей генерации упражнений на основе текстов, которые уже изучались или будут изучаться обучающимися, реализуя тем самым педагогический принцип повторения.

СПИСОК ЦИТИРУЕМЫХ ИСТОЧНИКОВ

1. Балканов И. В. Корпусная лингвистика: новые возможности лексикографии // Вестник Дагестанского государственного университета. Серия 2: Гуманитарные науки. 2022. Т. 37, № 4. С. 81–86. DOI 10.21779/2542-0313-2022-37-4-81-86. EDN CQINIT.
2. Айкина Т. Ю., Сыскина А. А. Модификация фразеологических единиц с семантическим компонентом “learning” в английском языке // Русский лингвистический бюллетень. 2023. № 6(42). DOI 10.18454/RULB.2023.42.17. EDN KFPJGL.
3. Султанова А. П. Использование корпусных данных в изучении английской фразеологии // Вестник НЦБЖД. 2022. № 4(54). С. 133–140. EDN YTQYZP.
4. Нурзет Ц. С. Б. Проблемы корпусного поиска немецких фразеологизмов в Мангеймском корпусе немецкого языка // Ученые записки Новгородского государственного университета. 2022. № 5(44). С. 575–578. DOI 10.34680/2411-7951.2022.5(44).575-578. EDN CXOMPA.
5. Жолобов О. Ф., Баранов В. А. Трансформации лексического ряда животъ – жизнь – житие: опыт лингвостатистического описания // Вопросы языкознания. 2022. № 2. С. 65–101. DOI 10.31857/0373-658X.2022.2.65-101. EDN LZPBHS.
6. Сарайкин И. В. Статистический подход к описанию семантической эволюции наречных предлогов // Верхневолжский филологический вестник. 2023. № 1(32). С. 75–84. DOI 10.20323/2499_9679_2023_1_32_75. EDN TZCBQD.
7. Kochetova L. A., Ilyinova E. Yu. Diachronic and dialect variation of English intensifying adverbs in the film dialogue discourse: corpus-based study // Science Journal of Volgograd State University. Linguistics. 2022. Vol. 21, No. 5. P. 95–107. DOI 10.15688/jvol-su2.2022.5.9. EDN MUYIGN.
8. Снегирев И. А., Непомнящих Н. М. «Скальд» и «бард» в русской поэзии. Опыт корпусного анализа // Власть истории и история власти. 2023. Т. 9, № 3(45). С. 41–49. EDN LVQBGG.
9. Горожанов А. И., Шевцова В. А. Технология определения цветовой характеристики текста художественного произведения (на материале немецкого языка) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2023. № 6(874). С. 63–68. DOI 10.52070/2542-2197_2023_6_874_63. EDN VOUQHW.
10. Парамонова Д. А. Метрическая организация текстов Б. Б. Рыжего и ее художественная роль в формировании поэтики автора // Труды института русского языка им. В. В. Виноградова. 2022. № 3. С. 121–129. DOI 10.31912/pvrl-2022.3.7. EDN CSUDTX.
11. Изюмская В. В. Актуальное состояние корпусной лингвистики: тенденции и перспективы // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2023. № 8(876). С. 27–33. DOI 10.52070/2542-2197_2023_8_876_27.
12. Ayre K., Bittar A., Kam J., Verma S., Howard L. M., Dutta R. Developing a natural language processing tool to identify perinatal self-harm in electronic healthcare records // PLoS ONE. 2021. No. 16(8). DOI: 10.1371/journal.pone.0253809.
13. Okhapkin V. P., Okhapkina E. P., Iskhakova A. O., Iskhakov A. Y. Constructing of Semantically Dependent Patterns Based on SpaCy and StanfordNLP Libraries // Communications in Computer and Information Science (in books). 2021. Vol. 1395 CCIS. PP. 500–512. DOI: 10.1007/978-981-16-1480-4_45.
14. Горожанов А. И. Создание лингвистического корпуса на основе инструментов обработки естественного языка: планирование программных решений // Филологические науки. Вопросы теории и практики. 2023. Т. 16, № 5. С. 1616-1620. DOI 10.30853/phil20230252. EDN BHZCSE.