Лещинская Е.Ф., магистрант, Лагунов А.Ю. к.п.н, доцент, Шестакова Т.Э., к.ф.н., доцент

ОСОБЕННОСТИ ОБУЧЕНИЯ РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ В УСЛОВИЯХ СОВРЕМЕННОЙ ИНФОРМАЦИОННО-ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ

Федеральное государственное автономное образовательное учреждение высшего образования «Северный (Арктический) федеральный университет имени М. В. Ломоносова», Архангельск, kaniskova@rambler.ru, a.lagunov@narfu.ru, t.shestakova@narfu.ru

Sofronova N. V., doctor of pedagogical Sciences, Professor

FEATURES OF LEARNING RUSSIAN AS A FOREIGN LANGUAGE IN THE CONDITIONS OF A MODERN INFORMATION AND EDUCATIONAL ENVIRONMENT

Northern (Arctic) Federal University named after M.V. Lomonosov, Arkhangelsk, kaniskova@rambler.ru, a.lagunov@narfu.ru, t.shestakova@narfu.ru

Аннотация. В статье авторы решают проблему малого количества обучающих текстов для обучения русскому языку как иностранному студентов технических специальностей, подробно раскрывают процесс формирования корпуса технического языка и частотного словаря учебных текстов, определяют лексический минимум по языку для обучения иностранных студентов технических специальностей.

Abstract. In the article, the authors solve the problem of a small number of teaching texts for teaching Russian as a foreign language to students of technical specialties, disclose in detail the process of forming the corpus of the technical language and the frequency dictionary of educational texts, determine the lexical minimum for the language for teaching international students of technical specialties.

Ключевые слова: обучение, информационные технологии, русский язык, иностранный язык, частотный анализ, корпус русского языка.

Key words: training, information technology, Russian language, foreign language, frequency analysis, Russian corpus.

Не существует решения без эмоций, которое будет иметь последствия в процессе обучения, так как существует связь между принятием решений и эмоциями [1]. «Мы чувствуем, что учимся», — это ключевой тезис к вовлечению учащихся в учебный процесс. Учащиеся учатся только в том случае, если они соотносят себя со школьными ценностями и правилами. Учащиеся должны знать, какой путь они должны пройти, чтобы стать частью общества. Без ценностей и связи с научным сообществом учащиеся не будут создавать чувство принадлежности, и они не будут интегрировать себя в него как его часть.

На лабораторных работах по физике и химии обучающиеся получают импульс в результате выполнения исследования. Данное действие кроме стремления к изучению технического предмета будет стимулировать обучающихся к изучению русского языка, чтобы лучше понимать физику, химию и другие технические предметы.

Лабораторные занятия улучшают развитие научных компетенций, а значит, в конечном итоге улучшают умение владением устной и письменной речью, но экспериментальные мероприятия, такие как выполнение заранее определенных шагов (инструкций), на которые ссылаются многие школьные учебники, часто не соответствуют ожиданиям учеников, которые, не одобряя их, сопротивляются их реализации. Данный факт может стать проблемой при обучении русскому языку, как иностранному студентов технических специальностей. К такому же эффекту приводят действия, когда учитель демонстрирует опыт, а ученик находится в пассивной позиции. Всё это может не способствовать активному вовлечению ученика, если такая деятельность не проводится в совместной среде обучения сверстников. Совместная деятельность предполагает общение, которое улучшает языковые компетенции. Таким образом, чтобы вовлечь учеников в лабораторные занятия, учитель должен предлагать такие мероприятия, которые не только мотивируют учеников, но и повышают их самооценку, заставляя их верить в себя.

Однако в существующей методической и филологической литературе имеют место проблемы, связанные с малым количеством современных технических текстов, которые были бы понятны обучающимся [2,3].

Частотный анализ является одним из относительно простых методов обработки текста на естественном языке (NLP). Его результатом является список наиболее часто встречающихся в тексте слов. Частотный анализ также дает представление о предмете и основных понятиях текста.

Для составления частотного словаря мы сформировали корпус текст-инструкций к лабораторным работам по физике, химии и основам исследовательской деятельности на основе учебных материалов Высшей школы естественных наук и технологий САФУ [4] и материалов электронной библиотеки сотрудников САФУ.

Проведение частотного анализа можно условно разделить на несколько этапов:

1. Загрузка и обзор данных.
2. Очистка и предварительная обработка текста.
3. Перевод слов в основную форму.
4. Удаление стоп-слов.
5. Подсчёт статистики встречаемости слов в тексте.

На 1 этапе мы произвели загрузку и предварительную подготовку текста. Все материалы представлены в библиотеке в формате PDF. Мы получили массив документов в 1549 страниц по физике, 443 страницы по химии и 752 страницы по основам исследовательской деятельности. Формат PDF очень удобен для обмена файлами с коллегами, но не совсем удобен для обработки компьютерными программами.

На 2 этапе мы удалили часть издательской информации, преобразовали тексты в формате PDF в формат TXT. Текстовый формат позволяет избежать проблем, связанных, например с тем, что в текстовом редакторе Word в файл добавляется дополнительная служебная информация, которая может исказить результаты исследования. Все документы были сохранены с использованием кодировки UTF-8. UTF-8 (от англ. Unicode Transformation Format, 8-bit — «формат преобразования Юникода, 8-бит») — распространённый стандарт кодирования символов, позволяющий более компактно хранить и передавать символы Юникода, используя переменное количество байт (от 1 до 4). Сохранение документов в данном формате позволяет работать с русским языком в полном объёме.

На 3 этапе для перевода слов в основную форму мы использовали программу MyStem [5]. Программа MyStem разработана в компании Yandex и производит морфологический анализ текста на русском языке. Программа имеет множество настроек, которые позволяют выполнить качественный анализ текста. Программа работает в среде DOS, что требует от исследователя некоторой предварительной подготовки.

Для запуска программы мы использовали команду:

mystem.exe input.txt output.txt -ld

Параметр -l выпишет только начальные формы слов (леммы), параметр -d попытается снять омонимию.

В качестве примера приведем часть исходного и переработанного текста.

Исходный текст: Способность металлических проводников хорошо проводить электрический ток, обусловлена наличием в них свободных электронов – «электронного газа».

Переработанный текст: {способность} {металлический} {проводник} {хорошо} {провожать} {электрический} {ток} {обусловливать} {наличие} {в} {они} {свободный} {электрон} {электронный} {газ}.

В результате мы получили удобный список слов в начальной форме в трёх файлах по физике, химии и основам исследовательский деятельности.

4 и 5 этап мы проводили, используя программу Antconc [6]. Программа AntConc представляет собой свободно распространяемое мультиплатформенное средство лингвостатистического анализа текста. Программа разработана профессором Лоуренсом Антони (Laurence Anthony), директором Центра обучения английскому языку в науке и технике Школы науки и техники университета Васеда (Япония).

В начале работы мы загружаем полученные на 3 этапе файлы в программу, используя меню File → Open File(s) (Рис.1).



Рисунок 1 — Окно программы Antconc.

Затем переходим на вкладку Word List и нажимаем внизу кнопку Start. В результате мы получаем список всех слов, которые встречаются в загруженных документах (Рис.2). Первый столбец (Rank) задает ранг слова в соответствии с частотой (второй столбец Freq), с которой слово (третий столбец Word) встречается в тексте. Всего в наших текстах используется 19237 слов. На первых местах, как ожидалось, располагаются стоп-слова или шумовые слова. К общим стоп-словам относятся предлоги, суффиксы, причастия, междометия, цифры, частицы и т. п. Для построения нашего частотного словаря данные слова не нужны, и мы их удалим.

Первое значащее слово в списке «работа», которое встречается в текстах 3022 раза (Рис.2).



Рисунок 2 — Список слов

Это может быть как физический термин в контексте «электрон совершает работу», так в контексте выполнения лабораторной работы. Все варианты использования конкретного слова можно наблюдать на вкладке Concordance (Рис.3).

На Рис.4 можно наблюдать частоту вхождения слова «работа» в словосочетания, такие как «работа по физике», «работа и обработка результата», «работа в библиотеке», «обучаться работе с результатом», «работа с различными источниками» и т.д.

Применение меню File-Save Output позволяет сохранить полученный частотный словарь. После обработки по удалению стоп-слов, которую необходимо выполнить вручную, мы получаем полный частотный словарь по технической дисциплине. В таблице 1 приложен фрагмент полученного частотного словаря.



Рисунок 3 — Контексты слова «работа» на вкладке Concordance.



Рисунок 4 — Частота входа слова «работа» в словосочетания на вкладке

|  |
| --- |
| Таблица 1 Частотный словарь |
| Rank | Freq | Word |
| 1 | 3022 | работа |
| 2 | 1932 | измерение |
| 3 | 1671 | значение |
| 4 | 1569 | раствор |
| 5 | 1484 | рисунок |
| 6 | 1417 | научный |
| 7 | 1333 | результат |
| 8 | 1293 | формула |
| 9 | 1180 | метод |
| 10 | 1151 | величина |
| 11 | 1137 | сила |
| 12 | 1119 | исследование |
| 13 | 1113 | получать |
| 14 | 1095 | определять |
| 15 | 1080 | таблица |
| 16 | 1072 | колебание |
| 17 | 999 | ток |
| 18 | 920 | температура |
| 19 | 898 | определение |
| 20 | 891 | система |

При формировании лексического минимума по техническим специальностям мы будем исходить из того факта, что «Анализ финального перечня с точки зрения лингводидактики показал, что для понимания 50% корпуса учебно-профессиональных текстов по дисциплине необходимо знание 1000 самых частотных знаменательных слов, а оптимальным количеством является интервал в 3500–4500 слов частотного списка, после освоения которых рост показателя покрытия замедляется, а понимание обеспечивается в большой степени средствами связности, логической структурой и стратегиями коммуникативного развертывания текста» [77].

На основе данных рекомендаций мы сформировали лексический минимум для студентов технических специальностей и список оптимальных слов, знание которых позволит студентам технических специальностей обучаться в вузе.

Литература

1. Damasio A. The neural basis of social behavior // Neuroethics. Mapping the Field, S. Marcus, New York: Dana Press. 2004. Pp. 100-107.
2. Аржанова И.А., Шичавина Е.В. Трудности перевода технических терминов в инструкции // Дневник науки. 2017. № 6 (6). С. 14.
3. Быстрова Л.В., Касаткина К.А. Специфика перевода текстов технических инструкций с английского языка на русский язык (на материале англоязычной инструкции bosch group) // Инновации. Наука. Образование. 2017. № 5 (7). С. 9.
4. Учебные материалы Высшей школы естественных наук и технологий. – URL: https://narfu.ru/hsnst/obuchenie/uchebnye-materialy/, свободный. – Загл. с экрана. – (Дата обращения: 07.05.2021).
5. MyStem. – URL: https://yandex.ru/dev/mystem/, свободный. – Загл. с экрана. – (Дата обращения: 07.05.2021).
6. AntConc. – URL: https://www.laurenceanthony.net/software/antconc/, свободный. – Загл. с экрана. – (Дата обращения: 07.05.2021).