

РАЗРАБОТКА ПРИЛОЖЕНИЯ НА ОСНОВЕ МОДЕЛИ ЦЕПЕЙ МАРКОВА С ЦЕЛЬЮ УГЛУБЛЁННОГО ИЗУЧЕНИЯ ЛЕКСИКИ АУТЕНТИЧНЫХ ТЕКСТОВ

Успенский М.П., магистрант

Московский политехнический университет, Москва, Россия

Ларцина С.В., магистрант

Государственный институт русского языка им. А. С. Пушкина, Москва,
Россия

Аннотация. Цепи Маркова в перспективе могут выступать в качестве эффективного средства для изучения лексики аутентичных текстов. Предложена концепция приложения «WordChain», работа которого базировалась бы на этом алгоритме. Приложение выложено в сеть Интернет, для внедрения в учебный процесс требуется его существенная доработка.

Ключевые слова: лингводидактика, русский язык как иностранный, аутентичные тексты, цепи Маркова, приложение

Цепи Маркова – это частный случай марковского процесса, математическая модель, которая описывает последовательность случайных событий. В этой модели распределение вероятности наступления каждого события зависит только от текущего состояния (набора возможных событий).

На сегодняшний день теория марковских цепей, идеи которой впервые сформировались ещё в 1907 г. благодаря А.А. Маркову, получила широкое распространение и применяется в различных областях: в биологии, статистике, автодорожном строительстве и т.д. В лингвистике цепи Маркова применялись повсеместно в машинном переводе, при последовательном переводе фразы с одного языка на другой, а также при генерации текстов. Однако современные

технологии ушли далеко вперёд, и модель уступила место нейросетям и применению ансамблевых методов машинного обучения, но не утратила актуальности до конца. Изредка выходят исследования, задействующие генерацию текстов на основе цепей Маркова [5]. В научной среде в рамках вычислительной («цифровой») педагогики продолжают использовать понятие цепей Маркова для построения системы рекомендаций в электронной образовательной среде [2]. Тем не менее, этот алгоритм обладает потенциалом, который может быть интегрирован в учебный процесс для углублённого изучения лексики аутентичных текстов на русском языке.

Цель статьи – представить концепцию приложения, работа которого основана на цепях Маркова, а также возможные пути внедрения этого приложения для углублённого изучения лексики аутентичных текстов в рамках занятий по русскому языку как иностранному.

На сегодняшний день к одним из важнейших задач лингводидактики относится формирование коммуникативной компетенции. Приобретённые в ходе её развития речевые навыки задействуются во всех сферах общественной жизни. Значимая роль в этом плане отводится освоению лексики иностранного языка, поскольку с её помощью реализуется чтение, которое является «источником удовлетворения реальных жизненных потребностей читателя...» [1, с. 23]. В перспективе это способствует самостоятельному ознакомлению учащихся с новым материалом.

Как отмечают современные учёные, развитию коммуникативных навыков особенно способствует использование на занятиях аутентичных текстов [3; 4]. Методист М. Брин, рассматривая специфику работы с подобным материалом, отмечал, что существуют разные точки зрения на этот вопрос. По мнению некоторых педагогов, учебные «искусственные» тексты искажают «данные целевого языка» [6, с. 70], в то время как аутентичные тексты, наоборот, предоставляют возможность применить полученные в ходе занятий знания при работе с речевыми произведениями, написанными носителями языка, и оттого сохранившие культурные и национальные особенности. Сам М. Брин

предполагал, что учащиеся склонны искажать данные им тексты для самостоятельного освоения новых знаний, и тем самым поддерживал идею, что создание специальных адаптированных текстов не является первостепенной задачей педагогов.

Несмотря на успешную интеграцию информационных технологий во многие области методики и педагогики, не были обнаружены вспомогательные средства – компьютерные программы, которые могли бы помочь эффективнее изучать лексику аутентичных текстов. Чтобы восполнить этот пробел, было разработано приложение, суть которого заключается в помощи преподавателю в организации *после текстового* этапа работы с аутентичными текстами, заданными для домашнего чтения.

Написание программного кода осуществлялось на объектно-ориентированном языке программирования общего назначения C#. Программная платформа .NETFramework, частью которой является C#, подходит для разработки приложений различного формата: веб-приложений, мобильных и игровых приложений и т.д. К прочим его преимуществам относится *компилируемость*, которая обеспечивает более высокую производительность программы.

В этом плане *интерпретируемые* языки в значительной мере уступают компилируемым. К примеру, для языка Python уже существуют готовые к установке библиотеки, позволяющие реализовать алгоритм цепей Маркова (pipinstallmarkovify). Тем не менее, необходимость быстрой обработки крупных текстовых массивов обусловила выбор языка C#. Общая длина кода составила 105 строк.

При сборке кода в Microsoft Visual Studio было выведено окно с пользовательским интерфейсом (Рис. 2).

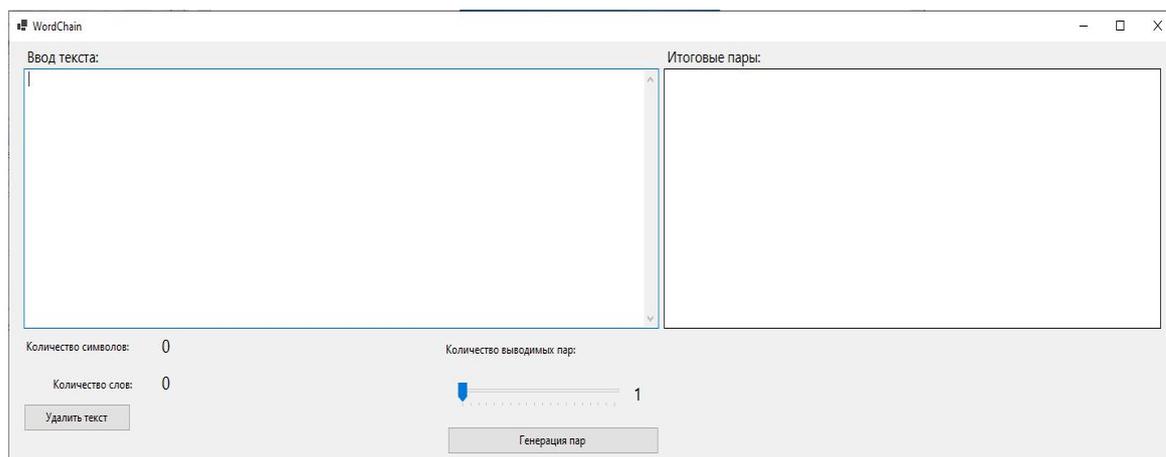


Рисунок 2 – Интерфейс приложения

Максимальный объём текста, который может быть загружен в программу для обработки составляет 30 000 знаков, что соответствует $\frac{3}{4}$ авторского листа (40 тыс. знаков). Это – приблизительный размер главы крупного произведения или небольшой рассказ целиком.

При загрузке текста и нажатии кнопки «Генерация пар» будет выведено выбранное количество словосочетаний.

Алгоритм работы с приложениям вариативен и можно свободно корректироваться преподавателем, но возможно выделить усреднённую последовательность действий: 1) преподаватель выбирает текст на изучаемом языке (предполагается, что это будет аутентичный текст); 2) учащиеся работают с текстом на дому в формате домашнего чтения; 3) студенты переводят текст, читают его, изучают незнакомую лексику самостоятельно; 4) на занятии преподаватель загружает текст в программу, после чего настраивает параметры вывода; 5) программа выводит цепочки слов, с которыми производится дальнейшая учебная работа.

Для демонстрации вывода программы в неё был загружен текст рассказа «Старуха Изергиль» М. Горького. Далее указывается, что необходимо вывести 15 пар. При нажатии на кнопку «Генерация пар» диалоговое окно приняло следующий вид (рис. 3).

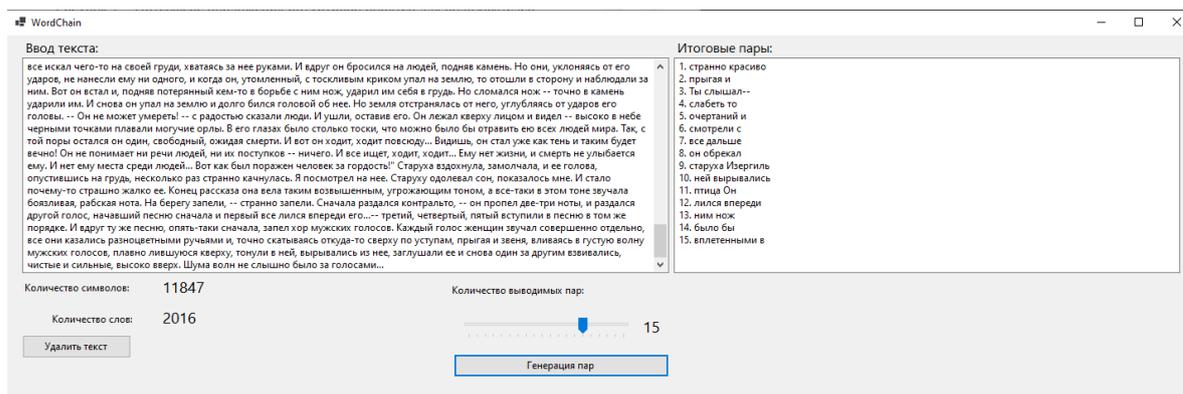


Рисунок 3– Вывод программы после загрузки рассказа «Старуха Изергиль» М. Горького

На данный момент с выводом программы связан ряд проблем, решение которых предполагается в дальнейшем:

1. Программа выводит хаотичный набор из двух словосочетаний, среди которых могут быть как бессмысленные сочетания слов, так и фразеологизмы, коллокации, фрагменты идиом.

2. Среди цепочек слов попадают бессмысленные сочетания служебных слов, чаще всего встречающиеся в позиции рядом.

3. Программа обладает широким спектром возможностей, который необходимо систематизировать и, вероятно, сузить их количество для решения конкретных задач с наивысшей эффективностью.

Помимо чисто технических вопросов, предполагается внесение ряда методических уточнений: 1) уточнить, для обучения на каком уровне целесообразно её применение; 2) определить, в рамках какого учебного профиля возможно применение «WordChain» как вспомогательного средства.

В имеющемся виде программа была заархивирована под названием «WordChain.rar» и выложена на онлайн-сервис [itch.io](https://theinspiredhedgehog.itch.io/wordchain) под следующим URL-адресом: <https://theinspiredhedgehog.itch.io/wordchain>. В описании указана инструкция с указанием пути к файлу: «WordChain ->MarkovLing ->MarkovLing ->bin ->Debug -> net6.0-windows -> MarkovLing.exe». Для запуска необходима последняя версия .NET Desktop Runtime (при первом запуске пользователю будет предложено автоматически скачать и установить нужную среду

разработки). Для работы с ней (после того, как программа будет в первый раз скачана и среда разработки установлена) не требуется подключение к сети Интернет.

На данный момент программа «WordChain» свободно распространяется и доступна для скачивания, но для эффективного внедрения в учебную среду требуется её существенная доработка.

Литература

1. Аболина Н.С. Процесс профессионального обучения и формирование коммуникативной компетенции. Инновационные проекты и программы в образовании. 2013. № 6. С. 21-28.

2. Геращенко Т.М. Алгоритм построения рекомендаций в электронных образовательных средах на основе стохастических моделей Маркова // Моделирование, оптимизация и информационные технологии. 2021. Т. 9, №4(35).

3. Казакова М.А., Евтюгина А.А. Аутентичные текстовые материалы в обучении иностранному языку // Вестник Бурятского государственного университета, 2016, № 4, с. 50-59.

4. Кондарин А.К. Аутентичные текстовые материалы в обучении иностранному языку будущих студентов-лингвистов // Теория и практика обучения иностранным языкам: традиции и перспективы развития : Материалы VII международной научно-практической онлайн-конференции студентов и молодых учёных, Мытищи-Гомель-Москва-Санкт-Петербург-Минск-Речица, 16 марта 2022 года. М.: Московский государственный областной университет, 2022. С. 129-133.

5. Павлов А.С. Метод обнаружения массово порожденных неестественных текстов на основе анализа тематической структуры // Вычислительные методы и программирование. 2011. Т. 12, № 3. С. 58-72.

6. Breen M.P. Authenticity in the Language Classroom / M.P. Breen // Applied Linguistics. 1985. Vol. 6, No. 1. P. 60-70.