

Создание методов автоматизированного анализа ритма текста

Научный руководитель: Бойчук Елена Игоревна, доктор филологических наук, доцент, доцент кафедры романских языков

elena-boychouk@rambler.ru

Основной научной задачей, на решение которой направлена работа научной школы, является создание новых методов в компьютерной лингвистике для автоматизированной обработки текстов на различных языках с целью их авторизации, анализа и построения ритмичных текстов на этих языках. Данная задача интегрирует компьютерную лингвистику и информационные технологии с такими областями филологического знания, как лингвистика текста, текстология, стилистика текста, риторика, психолингвистика, культура речи, а также с количественной лингвистикой, суггестивной лингвистикой, лингвокультурологией. основополагающим параметром создания данных методов является ритм текста.

Актуальность данной задачи состоит в том, что большое количество существующих теоретических исследований в области ритма, который требует статистического подхода, применения методов количественного анализа, автоматизированной обработки текста, не имеет выхода к осуществлению такого рода анализа. Анализ ритма осуществляется вручную, что снижает эффективность текстовой обработки и не всегда приводит к точным результатам.

В основу предложенных исследователями современных программных приложений для русского языка положены такие факторы определения ритмичности текста, как размер предложений и среднеарифметический показатель безударных слогов [Белоусов, К.И., Дусакова Г.Ф., Леонов Д.В., 2017; Кишалова Л.В., 2017; Андрусенко Т.В., 2017]. Для английского и французского языков существуют приложения, позволяющие анализировать текст с точки зрения его лексического состава, размера предложений и n-грамм [Text Analyzer, Sketch Engine, Stylo, Word2vec и др.].

Анализ ритма в рамках работы школы направлен на определение авторства текстов, сопоставление переводов художественных произведений, а также на анализ и составление суггестивных текстов на основе определения их ритмических характеристик, складывающихся из употребления средств ритмизации, проявляющихся на фонетическом, лексико-грамматическом и структурно-композиционном уровнях. В состав данных средств входят стилистические средства выразительности речи, в основу выделения которых заложена повторяемость элемента (рифма, ассонанс, аллитерация, оноματοпея, анаграмма, параномазия, таутацизм, деривация, полиптотон, анафора, эпифора, симплока, хиазм, однородные члены, эпаналепсис, анадиплосис, редупликация и др.).

В задачи школы входит создание автоматизированных алгоритмов, позволяющих проводить сравнительно-сопоставительное исследование текстов с точки зрения ритмических характеристик; определение статистических характеристик ритма текста и создание банка данных, включающего тексты с выявленными характеристиками ритма; разработка алгоритмов анализа и составления ритмичного текста с целью активного суггестивного воздействия на реципиента; создание метода атрибуции на основе ритмических средств текста и программного прототипа, позволяющего хранить и обрабатывать тексты, проводить сравнительно-сопоставительный анализ ритма в переводе и осуществлять атрибуцию.

В ходе работы предлагается использование перечисленных выше ритмических средств в качестве основы создания автоматизированного метода атрибуции текстов. Разработанные в настоящее время алгоритмы составления ритмичных текстов основаны на абзацном членении и размере предложения. Предлагаемый метод позволит расширить возможности суггестивного воздействия на реципиента при помощи представленных ритмических характеристик. Новизна предлагаемого метода заключается также в том, что он позволяет осуществлять ритмический анализ текстов, оценивать адекватность перевода с точки зрения ритмических характеристик, осуществлять атрибуцию текстов на трех языках: русском, английском и французском. В разработке задействованы новые методы: ритмический анализ текста, сравнительно-сопоставительный ритмический анализ, которые позволят получить более точные результаты обработки ритма текста.

Автоматизированная обработка текста с точки зрения ритмических характеристик позволит более успешно реализовывать задачу авторизации. Областью практического применения автоматизированной обработки текста с позиций реализации ритмических характеристик может являться публицистика, в рамках которой эффективным с точки зрения воздействия на реципиента является создание суггестивных текстов.

Основным перспективным результатом работы должно стать создание программного прототипа, включающего банк текстов на иностранных языках (корпус электронных художественных текстов для осуществления анализа ритма и корпус художественных текстов с выявленными статистическими характеристиками ритма), позволяющего осуществлять автоматизированный анализ ритма, авторизацию текста и проводить сравнительно-сопоставительный анализ переводов текстов.