

УДК 004.912

АВТОМАТИЧЕСКАЯ ПРЕДОБРАБОТКА АННОТИРОВАННОГО АНГЛИЙСКОГО ВЫСКАЗЫВАНИЯ ДЛЯ ПОСТРОЕНИЯ ТОНОГРАММ

Тугарев Н.А. (НИТУ МИСИС)

Научный руководитель – доцент кафедры иностранных языков и
коммуникативных технологий, к.пед.н. Толстых О.М.
(НИТУ МИСИС)

Введение. При анализе просодической составляющей речи используются тонограммы – графики, демонстрирующие относительные изменения интонации. Развитие технологий распознавания и синтеза речи, систем локализации и озвучки, а также проведение лингвистических исследований актуализируют задачу визуализации просодических характеристик текста. Для описания просодических явлений в лингвистике разработаны системы тонетической разметки (в частности, рассматриваемая в работе система O'Connor [1]), которые используют набор знаков для обозначения тонов, акцентов, пауз и других элементов. Интерпретация и визуализация такой разметки производится вручную, что требует от исследователя значительных временных затрат, особенно при работе с длинными фразами или при необходимости сопоставлять различные варианты разметки в исследовательских целях. Автоматизация этого процесса позволит ускорить анализ и может повысить его надёжность.

Цель данной работы – разработать автоматический метод, который позволяет определять структуру фразы и строить набор опорных точек (слог – его характеристики – относительная высота) для каждого слога по аннотированному высказыванию. Полученная таким образом структура может быть использована как для визуализации (построения тонограммы), так и для дальнейшего количественного анализа просодии.

Основная часть. Предлагаемый подход к автоматическому преобразованию аннотированной фразы на языке Python включает последовательную обработку входной фразы, аннотированной просодическими символами (например: «I ʌdon't 'want to 'go to the \cinema») [2].

На первом этапе требуется парсинг исходных данных, результатом которого является список слов, каждому из которых сопоставлен набор просодических меток.

Разбиение слов на слоги производится при помощи фонетического словаря CMU Pronouncing Dictionary, доступ к которому предоставляется посредством библиотеки nltk [3]. Количество слогов определяется по фонемной транскрипции, представляющей собой модифицированную систему ARPAbet [4]. Словарь CMUdict содержит более 134 тысяч записей, охватывающих большую часть общеупотребительной лексики. В словаре гласные отмечены маркерами ударения: 0 обозначает безударный гласный; 1 – первичный ударный гласный; 2 – вторичный ударный гласный. Подобная нотация позволяет соотнести фонетические слоги с метками перед ударными слогами в исходном тексте.

Чтобы разделить на слоги слова, отсутствующие в базе, возможно использовать принцип последовательности сонорных (Sonority Sequencing Principle, SSP), которому подчиняется слогообразование в английском языке [5]. Данный принцип реализован в классе `tokenize.sonority_sequencing` библиотеки nltk [6].

Далее на основе полученных данных выделяется структура фразы. Поиском справа налево находится ядерный слог. Первый ударный слог относится к началу такта; все предшествующие ему образуют предтактовую часть. Слоги между первым ударным и ядром относятся к такту, а следующие за ядром – к послетактовой части. На данном этапе также фиксируются тип шкалы или предъядерной части при наличии соответствующих символов во входных данных.

Высоты ударных слогов в такте определяются типом шкалы. При нисходящей шкале каждый последующий ударный слог располагается ниже предыдущего, при восходящей шкале – выше предыдущего. При скользящей шкале наблюдается чередование повышений и понижений. При скандирующей шкале происходит равномерное повышение от начала к концу такта. При уровневой шкале все ударные слоги находятся на одном уровне. На ядерном слоге при падении высота понижается, при подъёме – повышается. Слоги затакта продолжают тенденцию ядра с затуханием.

Для численной реализации указанных соотношений с целью последующей визуализации используется 10-балльная шкала, где предтактовой части соответствует минимальный уровень, а максимальный уровень достигается в такте в зависимости от его типа.

Выводы. Разработан метод автоматического построения набора опорных точек по аннотированному тексту, включающий парсинг входных данных, слоговую сегментацию, выделение структуры фразы (предтактовая часть, такт, ядро, затакт) и назначение относительных высот тона. Метод реализован в виде программного прототипа на языке Python с использованием библиотеки nltk. Тестирование на наборе из 50 аннотированных фраз, охватывающих основные типы просодических символов, при ручной проверке подтвердило корректность слогового разделения в 47 (94%) случаях, выделения структуры фразы – в 44 (88%), назначения относительных высот – в 41 (82%). Выходные данные представляют собой последовательность записей для каждого слога: текст слога, принадлежность к структурной части, показатель ударности, относительная высота тона. Разработанный метод может быть использован для автоматизации построения тонограмм в лингвистических исследованиях.

Список использованных источников:

1. O'Connor J. D., Arnold G. F. Intonation of colloquial English. – London : Longman, 1973. – 270 p.
2. Соколова М. А. и др. Практическая фонетика английского языка. – М. : Владос, 2001. – 383 с.
3. The CMU Pronouncing Dictionary [электронный ресурс]. – URL: <http://www.speech.cs.cmu.edu/cgi-bin/cmudict> (дата обращения: 20.02.2026).
4. Abdullah S. Detecting Syllables in a Word [электронный ресурс]. – URL: <https://www.baeldung.com/cs/syllabification-nltk-pyphen> (дата обращения: 20.02.2026).
5. Tits N. MUST&P-SRL: Multi-lingual and Unified Syllabification in Text and Phonetic Domains for Speech Representation Learning // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track. – 2023. – P. 74-82.
6. nltk.tokenize.sonority_sequencing [электронный ресурс]. – URL: https://www.nltk.org/api/nltk.tokenize.sonority_sequencing.html (дата обращения: 20.02.2026).

Автор _____ Тугарев Н.А.

Научный руководитель _____ Толстых О.М.