

Русакевич А.

АЛГОРИТМ РАССЧЕТА ФОНЕМАТИЧЕСКОГО СХОДСТВА ТЕКСТОВ НА КИТАЙСКОМ ЯЗЫКЕ

Система образования всегда стремится к наиболее точной оценке результатов работы обучающихся. Существуют методические указания на то, какую оценку ставить при каких ошибках, но их количество и его влияние на балл обычно являются безотносительной константой, одинаковой для текстов любых размеров, либо оценки выставляются преподавателем исходя из собственного ощущения неточности в произношении обучающегося. Подобные проблемы также присутствуют и в китайском языке.

С целью более объективной оценки произношения обучающихся нами был разработан алгоритм определения процента фонематического совпадения двух текстов на китайском языке, получивший название «Метод фонематической последовательности Левенштейна». Подразумевается, что для оценивания произношения алгоритм будет применяться для оценивания совпадения фонем в речи обучающегося с ожидаемыми. Алгоритм оценки совпадения включает в себя следующую последовательность действий.

1. Китайский текст записывается с помощью системы Пиньинь. Перед записью, чтобы впоследствии считать ошибку в эризации за одну, все иероглифы ㄝ, обозначающие эризацию, заменяются на букву «r».

2. Из полученного текста убираются знаки препинания и пробелы, все буквы становятся строчными, при этом перед буквами «а», «е» и «о» ставится апостроф. В алгоритме слова соединяются в одно, поскольку китайский язык не имеет такой же четкой границы между словами, как европейские, следовательно, в Пиньинь стоит проблема постановки пробелов, которые в оценке фонематического сходства являются несущественными. Полученный результат впоследствии будет упоминаться как «фонематическая последовательность Левенштейна».

3. Далее по алгоритму В. Левенштейна (Левенштейн, 1965) высчитывается расстояние между строками.

4. Полученное число отнимается от количества символов в самом длинном из фонематических последовательностей Левенштейна, а разность затем делится на это же количество.

Таким образом, процент фонематического сходства китайских текстов поддается объективному расчету, но, судя по количеству требуемых вычислений и шагов, алгоритм является слишком времязатратным для практического применения. Помимо вышесказанного, встает вопрос об оценке интонации и о субъективности человеческого восприятия в целом и слуха в частности при применении описанного нами метода для проверки китайского произношения. На эти вопросы призваны дать ответ программа shprote (Standardized Hanyu pronunciation tester), разработка которой активно ведется в репозитории Github (<https://github.com/alex-rusakevich/shprote>), а также виртуальный робот в программе мгновенного обмена текстовыми сообщениями Telegram (https://t.me/shprote_bot).