# Detecting typographic, particle and sentence formation errors in JLPT N4 sentences using a rules-based system

Ana Verulidze  Earlham College  Richmond, IN  averul19@earlham.edu

## 1) Motivation

NLP provides excellent computer assisted language learning resources, however it still remains somewhat underdeveloped for many languages including Japanese.

In order to improve student-learning experience, this research aims to explore the effectiveness of a rules based system for detecting typographic, particle and sentence formation errors in JLPT N5 sentences.

## 2) Methodology

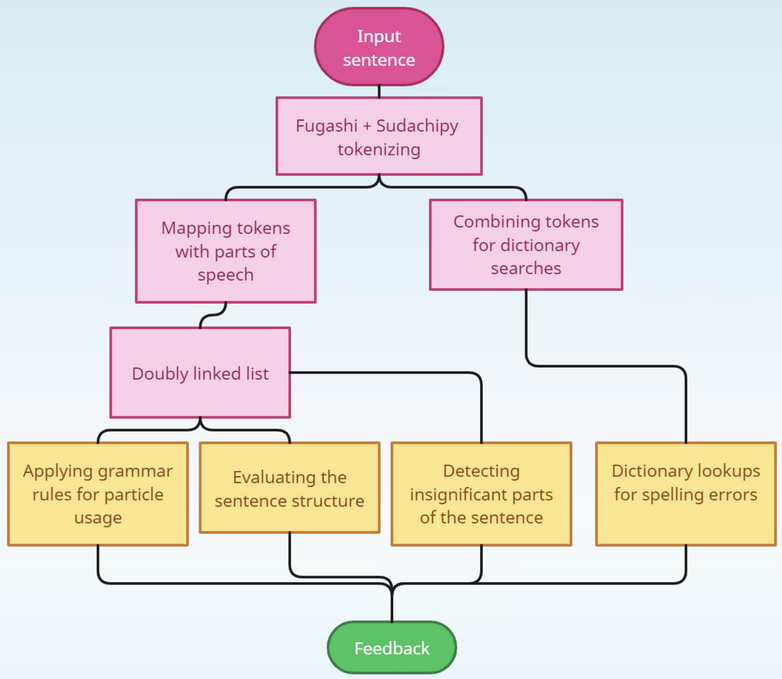
### Preprocessing Phase

Just like many other NLP tools, preprocessing starts with tokenization, which implies splitting text into meaningful parts of sentences. Tokenized output is transformed into a syntax tree for morphological analysis.

### Evaluation Phase

In this phase, we perform morphological analysis to evaluate sentence structure, and completeness, and put together a brief summary of errors detected, which will later be outputted to the user.

## 3) Data Architecture Diagram

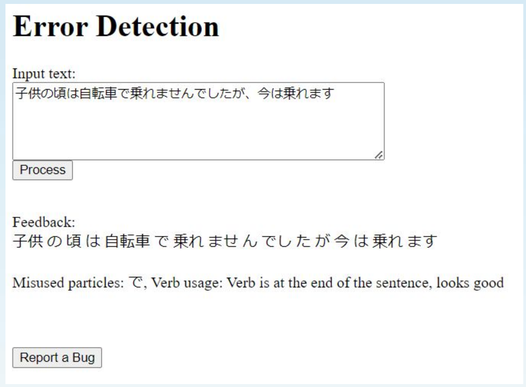



The feedback will provide proposed improvements, parts of speech and readings.

## 4) Testing

Accuracy testing is conducted with the help of native speakers, as well as Earlham's Japanese Language and Linguistics department.

## 5) Results




(1) Particle error was detected で
(3) Kanji readings were added
(4) Verb position was evaluated
(5) Parts of speech were identified

## 6) Future Work

Expanding the software to fit higher language proficiency levels.

Reducing kanji-reading errors via machine learning.

Training a machine learning model on a language corpus to detect spelling errors.

## Acknowledgement

I would like to express my gratitude for the unwavering guidance and constructive criticism provided by Dr. Charlie Peck and Dr. David Barbella.