

Master's Thesis Abstract

Student ID Number: 22GH201

Name:

The Graduate School of Humanities and Social Sciences
Major Program / Course :
Modern Inter-Cultural Studies Course

Mikoto Ono

Title

Creation of Tsugaru Dialect Dictionary Database Based on Tsugaru Dialect Corpus and Proposal of Usage as Educational Tool

This dissertation describes the creation of a dialect dictionary database linked to a corpus of Tsugaru dialect folktale data. Considerations are also given for uses of this dictionary database and corpus as tools to help with language learning.

As of June 2022, while there are 2,961,969 foreigners living in Japan, in Aomori prefecture the number is especially low at 6,306 foreigners. This low number makes providing dialect language support difficult. Language, especially dialect, can be a big barrier to communication for foreign students and workers. One reason for this is that existing collections of dialects and dialect dictionaries are difficult for Japanese learners to understand, and there are few studies that focus on the vocabulary of the Tsugaru dialect itself, particularly those that create a dialect dictionary database based on a corpus.

The dialect corpus used for this project, The Matsunoki Treebank, is a collection of recorded Japanese folktales. The recordings are from performances by the "Wa no Mukashi Ko" group, storytellers who are active in Hirosaki City and the surrounding areas. The folktales are transcribed using a revised version of Hepburn Romanization. This transcription method aims to reflect the pronunciation of the Tsugaru dialect in a way that is easier for non-Japanese native speakers to understand.

Sense definition entries for a headword of the dialect dictionary database follow a template headed by a "key" consisting of English words to capture the rough meaning of the headword and numbers to individuate the entry definition. Entries also contain simplified part-of-speech information, headword definitions, pronunciation variations, synonyms, and links to attested examples from the corpus. Also, there can be additional information depending on the part of speech: pictures and entity information for nouns, thesaurus frame IDs for verbs, etc. Sense definitions follow the COBUILD (advanced learner dictionary) method of embedding headwords in English sentences that both define and illustrate word use.

In parallel with the creation of the dictionary database, the dialect corpus was made more accessible to speakers of Japanese by adding Japanese script. Also, English translations were added that were crafted to follow the analysis of the parsed corpus. These translations serve as both an aid for language learners and as meta-level comments to help with understanding the parse analysis of the Tsugaru dialect data.

In the future, there are plans to add additional headwords and their definitions, and information on regional and chronological differences in the Tsugaru dialect.

It is believed that Japanese learners will be able to obtain information on the Tsugaru dialect more easily by using the content created in this study and now they can access this information on the internet.