# Making a tagged dialect corpus using a computer: Verbal conjugation can be automated

Motoei Sawaki
*Shinshu University*
Chitsuko Fukushima
*University OF Niigata Prefecture*
Yumi Nakajima
*Hitotsubashi University*

We are completing a tagged dialect corpus utilizing a computer. There has never been such a corpus in Japan; thus, there is no similar example.

In Tokunoshima dialect of Japanese language, nouns, particles and verbs have more than one form. The basic form and the form realized in an actual sentence often do not coincide. Therefore, each morpheme in the corpus needs specific information to be realized in a sentence. For example, a verb needs information to be conjugated. If the information is tagged manually, it takes time and mistakes may occur.

In our attempt, conjugation rules are programmed to produce all possible verb conjugation forms and compare them with actually realized forms. By doing this, it is possible to presume what kind of forms the realized forms are. There are more than 400 conjugation forms because auxiliaries following a verb also conjugate.

In addition, by revising this program, we can verify whether the tagged information on conjugation is correct. The program produces a form based on the basic form and its tagged information on conjugation. If the program cannot produce the same form as in the text, then the information is wrong.

The program has been developed based on the characteristics of Tokunoshima dialect that auxiliaries follow verbs and that the order is fixed with regard to the meaning of auxiliaries. Since the other Ryukyu dialects and Japanese mainland dialects share these characteristics, the method valid in the analysis of Tokunoshima dialect is applicable to other dialects.

The description of verbal conjugation of Tokunoshima Dialect was transplanted into this program. If the program-produced form is not the same as the actually realized form, then the description was not complete. This is how the description is elaborated.