

Developing linguistic Atlas of Japan database and advancing analysis of geographical distributions of dialects

YASUO KUMAGAI
NINJAL

The Linguistic Atlas of Japan (LAJ, 6 vols.) was published from 1966 to 1974. The total of 285 items (mainly lexical field) and 2400 localities were surveyed.

Various nationwide quantitative studies have been conducted on LAJ since its publication, such as dialect areas, geographical diffusion of standard forms and so on, using multivariate analyses and arithmetic techniques. However, most of them were based on prefecture-unit calculations (there are 47 prefectures in Japan). Since LAJ was a printed material, the lack of digital data has restricted the ways of quantitative studies. In 1999, we started constructing the Linguistic Atlas of Japan Database (LAJDB) with the aim of preserving the original survey materials (540,000 cards) and advancing various utilization of LAJ. At present, 119 items have been completed.

LAJDB consists of image database of original cards and coded data. The card images linked to each entry of coded data enable us to access the original phonetic recordings, informant's comments and the notes of the fieldworkers and the editors.

In order to explore the potential of LAJDB in advancing the quantitative approaches to LAJ, we made preliminary observations, such as geographical distributions of the frequency of (1) standard forms, (2) multiple answers, (3) informant's comments on standard forms, (4) geographical distributions of degrees of similarities, and (5) network representation of degrees of similarities. All of these showed clear patterns and structures.

We compared these patterns and structures with each other and with extra-linguistic features such as network of roads and so on to examine the relations among them. For example, the "high resolution" patterns obtained by LAJDB enable us to trace various diffusion routes, which former studies could not detect. We will show visual presentation of these results and discuss several implications for advancing quantitative analysis of LAJDB.