Creating the Chinese online dialect atlas

ROBERT SANDERS
University of Auckland

ERIC WHEELER
York University

Access to dialect data is greatly enhanced by the use of digitized databases and information technology. The future of dialect studies depends crucially on having that expanded access.

A published dialect atlas, such as Cao et al. (2008), with 510 maps in the full set of data, each map containing 930 different sampling locations, has the capacity to show you just what the editors have selected to show (Sperleich and Sanders, 2010). With our digital system, based on a prototype developed for Romanian (Embleton, Uritescu and Wheeler 2010), we provide an almost unlimited set of views of the data, defined over user-chosen selections of linguistic features, traditional dialect classification labels and geographic regions. But further, the system provides tools for analyzing the selected data, from simple quantitative measures and previously-developed multidimensional scaling maps to new comparison charts.

With this system, a pilot study was carried out using ten grammatical interpretive maps from the above Chinese dialect atlas. Its findings support previous observations of Finnish and Romanian dialects (Embleton, Uritescu and Wheeler 2008, 2012) that generally speaking, 1) any given classification of dialect groupings will change as the features selected to define them change, and 2) a rough correspondence exists between geographical distance and linguistic distance.

The challenges, both technical and socio-political, of adapting the printed data to the online technology to be discussed in this presentation are worth noting because other researchers will face comparable challenges. However, the value of digitizing such data is undeniable for the future of dialect studies.