A comparison of three digital systems for dialect geography

SHEILA EMBLETON
DORIN URITESCU
ERIC S. WHEELER
York University

Dialectology benefits from large digital data sets when answering questions such as 1) what is the relationship between geography and dialect variation? 2) where and what are the dialects in a given region? or 3) is a given hypothesis valid over an entire region?

There are several systems available to access and analyze large digital dialect data sets. We look at the R statistics package, Gabmap, and a generalized version of the Romanian Online Dialect Atlas (RODA) called the General Online Dialect Atlas (GODA), and try several different tasks. We show what it takes to do each task, using each of the systems.

The overall result we find is that each system has its strengths, which depend on 1) what the user wants to do, 2) what the user knows already, and 3) the size and nature of the data available. There is no one best system. Through the functions that allow for the exchange of data, the best approach may be to use more than one system to get meaningful answers to one's research questions.