Canonical Typology: a prerequisite to substantive interoperability

Dunstan Brown, Greville Corbett and Anna Kibort
Surrey Morphology Group, University of Surrey

For linguistics we can envisage two kinds of interoperability. The first kind facilitates data sharing. The second, more ambitious, kind provides the basis for comparison and checking of theoretical claims about languages. Unfortunately, interoperability of the first kind does not guarantee interoperability of the second kind, because the intuitions of linguists, even when they have used an interoperable format to describe a language, may not be recoverable.

If one wishes to take the more ambitious route and develop interoperability of the second kind, a major problem faced is that linguists often do not agree about the category or type to which a particular phenomenon belongs. The reasons for this are numerous, but often it has to do with choosing a particular characteristic or property as fundamental or defining. To illustrate with a concrete example from Russian, emphasis can be placed on a number of different properties when defining the notion 'agreement', including that the controller and target of agreement mark the same feature value. But this might lead one to exclude examples such as the following, when similar constructions in the language would fit:

(1) a. *ja pisa-l*
       1SG write-PST[SG.M]
    b. *ja pisa-l-a*
       1SG write-PST-SG.F
    'I was writing.' (man) 'I was writing.' (woman)

The point is that constructions such as these share some, but not all, of the properties we might associate with the notion 'agreement'. And as the definition of this notion has far-reaching consequences for theory, making a decision as to what counts or does not count is significant for the rest of the system. As a second example, a potential source of confusion in talking about gender systems is that they are both systems of nominal classification and systems of agreement; if linguists take one notion rather than the other as criterial, they may be talking about essentially different things.

This is why we believe that a key step in the development of ontologies should be to follow a canonical approach to typology (Corbett 2005, 2006; Seifart 2005;

Suthar 2006). In this approach a range of criteria are applied to determine whether a particular instance is a more or less canonical instance of a particular phenomenon (such as agreement or gender). The key insight for ontology developers is that emphasis needs to be placed on the criteria used to define linguistic entities as canonical or less canonical instances. Only then can we be sure whether linguists are talking about the same or similar things.

References

Corbett, Greville G. 2005. The canonical approach in typology. In: Zygmunt Frajzyngier, Adam Hodges and David S. Rood (eds) *Linguistic Diversity and Language Theories*. Amsterdam: John Benjamins, 25-49.

Corbett, Greville G. 2006. *Agreement.* Cambridge: Cambridge University Press.

Seifart, Frank. 2005. *The Structure and Use of Shape-based Noun Classes in Miraña (North West Amazon).* PhD thesis, Radboud University, Nijmegen.

Suthar, Babubhai Kohyabhai. 2006. *Agreement in Gujarati*. PhD Dissertation, University of Pennsylvania.