Morphological Aspects of Computer-Driven Elicitation of Knowledge about Any Language

Sergei Nirenburg
University of Maryland Baltimore County

This talk will describe morphological aspects of the Boas environment, which supports the acquisition of machine-tractable knowledge about any language from informants who need not be linguists. The discussion will focus on certain choice spaces available to developers based on (a) cross-linguistic research of language typology, (b) the current state of natural language technologies and (c) expectations regarding the capabilities and preferences of users. For example, when configuring the module to elicit productive paradigmatic inflectional morphology of so-called "open-class" meanings, developers have to make decisions and choices about at least the following issues:

a) whether paradigms or inventories of affixes and morphotactic processes are elicited;
b) what constitutes a valid paradigm for the purposes of a given learning algorithm;
c) how complex should the learning algorithm be, as complexity comes at a cost;
d) whether the learning algorithm is designed to be assisted by human input;
e) if so, what kind of input is preferable;
f) whether the system will generate inflectional forms as validation of its rule sets;
g) the balance between manual and rule-based acquisition;
h) the way multi-word inflectional forms are treated – as part of paradigms or separately;
i) the best pedagogical strategy for teaching non-linguists about word forms, roots, affixes, paradigms, etc.

Questions such as the above arise not only with respect to paradigmatic inflectional morphology but also with respect to agglutinating inflectional morphology, derivational morphology, and the morphological realizations of closed-class meanings.

This is joint work with Marjorie McShane.