Is linguistics an empirical science?

Bruce Morén-Duolljá, University of Tromsø  
bruce.moren@uit.no

In this talk, I will try to address the question of whether or not we should trust the published data on which we base our analyses and theories. There are several themes that I will touch on, including:

- E-language versus I-language
- Cross-linguistic phenomenon typology versus whole-language analysis
- Traditional description versus instrumental examination
- Dead language data versus living language data
- Data verifiability, completeness and variability

It seems that we are often opportunistic in the data we use to support our latest theoretical contributions. It also seems that we often assume much about the details of data without either verifying them or arguing that they are correct for the given language/phenomenon in question. That is, we take as a given what our predecessors have said about data that they have collected, reconstructed and/or analyzed. But, how do we know that the data are correct? How do we know that our theories are adequate if we do not/cannot verify the data?

Those who look carefully at the details of particular languages are sometimes frustrated by the way those languages are portrayed in the literature. Not only do languages/phenomena often not fit into a typological niche as claimed, but analyses of particular phenomena in given languages sometimes do not seem compatible with analyses needed elsewhere in those same languages. In examining the details, we sometimes discover discrepancies between language fact and language description that are not trivial. Yet, how many of us look at detailed primary data from explicit individuals or sets of individuals? How many of us ensure that our analysis of one phenomenon is compatible with the rest of the grammar?

I will use two examples to illustrate some of the problems with using traditional descriptions. First, I will look at the tonal system of standard Thai. Although traditionally described as a “simple” five-tone system (high, mid, low, falling and rising), there are important distributional restrictions and phonetic facts suggesting that the tones are not so simple after all. Further, data that are often overlooked are actually vital in providing a satisfying analysis (Morén and Zsiga 2006).

The second example comes from Lule Saami, a highly endangered language of the Saamic branch of Finno-Ugric. In many ways, this language seems to be a typological outlier – at least from the perspective of how phonetics, phonology and morphology interact. It is claimed to have a constellation of properties that are exceedingly rare, defy analysis (and often characterization), and are claimed by some popular theories to not be possible. These include unusual segment inventories, a bias for large syllable rimes, three degrees of linguistically significant consonant duration, laryngeal contrasts only in “weak” positions, unusual morpho-phonological consonant mutations, etc. While it is tempting to take the position that we should
disregard Lule Saami and other “languages of rather unusual structure” (Anderson 1985:125), is this actually scientifically justifiable or responsible?

Given recent advances in both technology and theory, it seems that linguists, like other scientists, should occasionally reassess our priorities and assumptions. At the very least, we should make sure that the data we use are actually adequate and verifiable.

References