Language Documentation: the Nahuatl Grammar

Mike Maxwell,[1] Jonathan D. Amith[2]

[1] Linguistic Data Consortium
maxwell@ldc.upenn.edu

[2] Gettysburg College
jonathan.amith@yale.edu

We describe an on-going documentation project for Nahuatl, an indigenous language of Mexico. While we follow standard recommendations for documenting text corpora and for the dictionary, the usual recommendations are not explicit concerning the grammar. Since Nahuatl is an agglutinating language, the morphological component of the grammar is highly complex. Accordingly, we consider it essential to not only provide static information about the language, such as a lexicon and parsed text, but dynamic documentation in the form of a working morphological grammar. When compiled into a finite state transducer, this grammar provides parses for arbitrary inflected forms, including many not in the corpus, as well as the generation of the partial or full inflectional paradigms. In keeping with the archival goals of language documentation, we argue that this grammar should be simultaneously human readable and computer processable, so that it will be re-implementable in future computational tools. The notion of literate computing provides the appropriate paradigm for these dual goals.

Language Description and Documentation

"It is to be lamented… that we have suffered so many of the Indian tribes already to extinguish, without our having previously collected and deposited in the records of literature, the general rudiments at least of the languages they spoke. Were vocabularies formed of all the languages spoken in North and South America, preserving their appellations of the most common objects in nature, of those which must be present to every nation barbarous or civilized, with the inflections of their nouns and verbs, their principles of regimen and concord, and these deposited in all the public libraries, it would furnish opportunities to those skilled in the languages of the old world to compare them with these, now or at a future time, and hence to construct the best evidence of the derivation of this part of the human race." –Thomas Jefferson (1781-1782) *Notes on the State of Virginia*

There are over 6000 languages in the world today [1]. The diversity of these languages has the potential to provide us with a window into the mind, an