Mark Liberman

The Philology of the Future: Audio as Text for Language Documentation

Abstract

Language documentation requires (the equivalent of) tens of millions of words of diverse text, but endangered languages have little or no culture of writing, and few people who are capable of transcribing speech accurately. Projects like BOLD in Papua New Guinea show that it's not hard to collect the hundreds of hours of recorded speech that correspond to millions of words of text -- but then what? This talk will discuss how future language-documentation methods will get past this barrier, by turning digital recordings into (something even better than) texts. I will also we can do today to bring this future closer; and suggest efficient methods for documenting languages today that will give the philologists of the future the evidence that they need.