

**Steven Bird**

*Linguistic Data Management with the Natural Language Toolkit*

Recent technological advances are having a profound impact on the accessibility, quality and quantity of digital language documentation. However, linguists wanting to reorganise, restructure, and reuse this data are often frustrated by the proliferation of incompatible special-purpose tools and formats. NLTK is a suite of Python program modules, data sets, tutorials and exercises, covering symbolic and statistical natural language processing. I will show how NLTK can be used by linguists to write simple programs for manipulating and analyzing linguistic data. I will provide examples from a variety of documentation types, supporting descriptive research in such areas as phonology, syntax, lexicography and stylistics.