Delivering the benefits of best practice data: Lexicon exploration in Kirrkirr

EMELD

Presented by: Christopher Manning, *Stanford University*
Project/Software Title: Kirrkirr
Project/Software URL: http://nlp.stanford.edu/kirrkirr/
**Access/Availability:** Free for download

Description:

Kirrkirr aims to provide an end-user tool (for both linguists and language community members) that showcases and exploits the benefits of having lexical data in best-practice formats. It simultaneously aims to highlight innovative methods of transforming and visualizing dictionary data and to support the user and augment a dictionary in various useful ways.

The EMELD project and others have stressed the importance of archival data formats, including structured formats for textual data, such as XML. However, despite increasingly good tools for creating such data, producing well-structured data is inevitably more work for the producer than gathering inconsistent or unstructured data. It is thus essential to show the creator the benefits that come from having well-structured archival data.
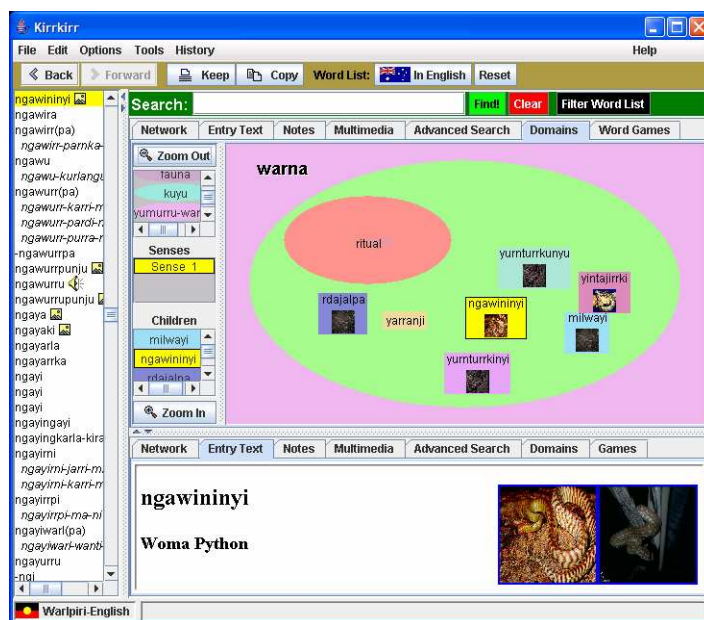
The central premise of Kirrkirr is that once you have well-structured dictionary data, then computer software can automatically (and cheaply) transform and visualize that data in ways that satisfy varying user needs, and in ways that go well beyond what online dictionaries typically offer.

Kirrkirr focuses on indigenous language lexicons. For native speakers or learners, it can provide pronunciations, easy lookup via spelling correction, maps of word associations, and automatically produced activities such as word games. For a linguist, the software provides flexible searching, visualization of semantic domains and lexical chaining, and a way to find dictionary gaps and inconsistencies.
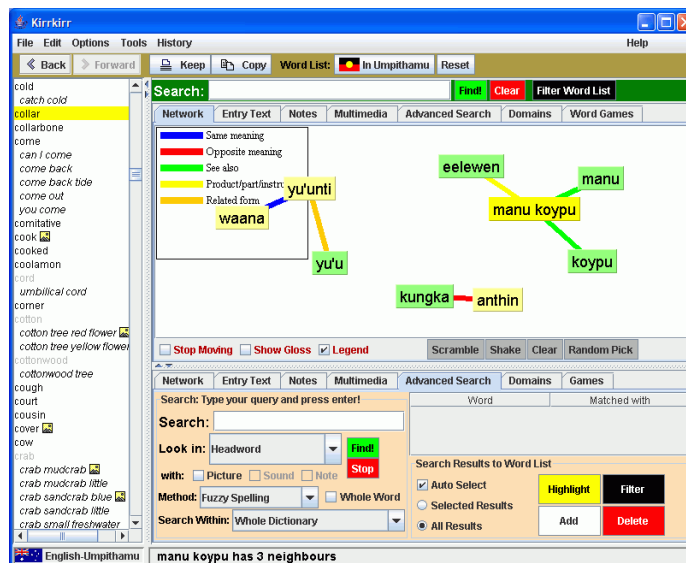
The initial design of Kirrkirr and its usability are discussed in Manning et al. (2001) and Corris et al. (2004). In the last few years, Kirrkirr has continued to be developed:

- Generalization of the data model and data access so that Kirrkirr can work directly with a broad range of XML lexicons, including lexicons encoded in the FIELD format or SIL's MDF (MultiDictionaryFormatter) – once this has been converted to XML by a tool such as Toolbox

- A new visualization panel that allows exploration of dictionary content via (hierarchical) semantic domain browsing

- A means of leveraging an L1-L2 dictionary to automatically produce a rough L2-L1 dictionary

- Automatic deployment of word games based on dictionary content.

Taken together, these tools not only transform a dictionary into a more captivating, dynamic environment, but they give the dictionary creator a clear sense of something new that well structured lexical data has enabled.

References:

Miriam Corris, Christopher Manning, Susan Poetsch, and Jane Simpson. 2004. How useful and usable are dictionaries for speakers of Australian Indigenous languages? *International Journal of Lexicography* 17:33-68.

Christopher D. Manning, Kevin Jansz, and Nitin Indurkhya. 2001 Kirrkirr: Software for browsing and visual exploration of a structured Warlpiri dictionary. *Literary and Linguistic Computing*, 16(2): 135-151.