

Documenting endangered languages using LEXUS and ViCoS



The Language Archiving Technology (LAT) is meant to contribute to the sort of infrastructure that will be required in eHumanities. Its design will finally help to boost research in the humanities and to attract indigenous communities and the interested public to use the rich information in the language archive. It focuses on open accessibility of language resources; it supports dynamic and continuously enriched collections according to the *Live Archives* ideas; it stresses the need for long-term archiving of our digital collections covering unique material about languages that will probably become extinct in a few decades and it follows the trend towards service oriented architectures. **LEXUS** and **ViCoS** are part of the LAT suite.

LEXUS

LEXUS is a tool for the creation and exploration of multi-media online lexica in the context of language description and documentation.

With LEXUS researchers may create lexica from scratch or import lexica from other formats, like Toolbox or XML. LEXUS is not just another lexicon tool: LEXUS offers the possibility to visualize language through the import of multimedia (images, videos and audio), which can be linked to the lexical entries. LEXUS also interacts with media stored in the digital archive for linguistic resources housed at the institute.

Users create their lexica in private workspaces and can assign other users access rights to the lexica in these workspaces (either read-only or by assigning edit rights). This collaborative environment permits to share lexica amongst linguistic researchers and speech community members.

Lexicography in general is a domain where uniformity and interoperability have never been the operative words: a variety of formats, structures and concept are being adopted, depending on linguistic theories or purpose of the lexicon. This makes cross lexica search, merging, linking and comparison an extremely difficult task. LEXUS contributes to the effort of resolving this problem. Being based on the Lexical Markup Framework (LMF), an abstract model for the creation of customized lexicons defined following the recommendations of the ISO/TC 37/SC 4 group on the standardization of linguistic terminology. LEXUS allows researchers to create purpose-specific and tailor-made lexica, and, provided that the offered standards are used, it assures the interoperability with other resources.

LEXUS is freely available: <http://www.lat-mpi.eu/tools/lexus>



ViCoS

ViCoS (Visualising Conceptual Spaces) aims at providing a simple interface between the lexical space (giving access to lexical resources) and the ontological space (where users relate concepts with each other).

In conventional dictionaries relations between words are only cross-referenced within lexical entry articles, whereas ontological resources make the semantic network between headwords visible beyond lexical entry articles; the headword serves as a key to a multi-layered network of semantic relations that combines the linguistic properties of the word with the cultural meaning and usage of the concept.

ViCoS is designed to add value to LEXUS by allowing users to create arbitrary relations between lexical entries or their parts, such as images or example sentences. The resulting network of semantic relations between entries form a kind of conceptual space, where words denote increasingly complex concepts. While each concept is anchored in the lexical space, its meaning becomes culturally relevant in the context of the concepts it is connected to.

The main requirement for the construction of knowledge spaces is thus to enable community members to anchor the words of a linguistic resource according to their classifications, hence creating a kind of "ethno-ontology". It seems clear that existing ontologies have only limited relevance here. Their proper and effective use requires considerable expertise or training, and they also induce a significant and usually Westernized bias of how the world should be modeled.

ViCoS can be accessed from the LEXUS workspace



Yéli Dnye is a Papuan language spoken on Rossel Island in the Louisiade Archipelago, Papua New Guinea. Language documentation has been a continuous process since 1995 within the project *Pioneers of Island Melanesia* (Stephen C. Levinson). Different linguistic and cultural topics are recorded on video, audio and images.

For Yéli Dnye a toolbox dictionary has been created containing over 6000 lexical entries. A large part of these entries are nouns representing objects in the natural world. The original dictionaries are imported in LEXUS and enriched with multimedia fragments. These fragments represent the meaning of the words in their indigenous contexts.

One specific aim of the project is to document the objects in the natural world and the specify the relations that exists between them. ViCoS supports this project aim: the researcher can, together with the members of the speech community, fuzzily define concepts and relation types and create a conceptual space of the natural world which closely relates to the vision that the members of the speech community have.

Members of the speech community can actively participate in this process: the ViCoS user interface is easy to use and the browser is easy to use for the exploration of the conceptual spaces created.

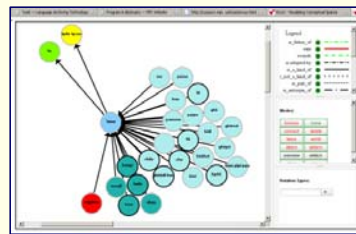
From the ViCoS browser users can return to the full content of the lexical item in LEXUS.



Rossel Island



Yéli Dnye lexicon in LEXUS



Conceptual space for 'nmo'

Major Goals Language documentation aims at the creation of representative and long lasting, multi-purpose record of natural languages. It contributes to maintain, consolidate or revitalize endangered languages and thus also contributes to the description of cultural elements of a language community. Our aim is to increase this cultural aspect by allowing users to complement linguistic information with ontological information.

Two groups of users are supported:

- Scientists such as linguists and anthropologists may use LEXUS and ViCoS to contribute to and use resources to study the language and culture of the community, or to compare them to the ones of other communities. ViCoS complements LEXUS by making available easy-to-use technology to impose a conceptual view over the lexical information in LEXUS. Moreover ViCoS permits to define structures that are orthogonal to the ones defined in the standardized LEXUS-LMF lexicon schema.

- Speech community members may use LEXUS to browse the lexicon by word list. In addition ViCoS can be used to describe their languages and culture and to learn from such resources. For community members, words are keys to access and describe relevant parts of their life and cultural traditions, such as food preparation, medicine, ceremonies, legends etc. For them, words are of foremost practical rather than theoretical nature; and their understanding of words is best described by the various associations they evoke rather than in terms of any formal and abstract theory of meaning.



LEXUS and ViCoS have been developed in close collaboration with the end users (researchers and members of the speech community)

Iwaidja is a language spoken by approximately 200 speakers on the Cobourg Peninsula, Northern Territory, Australia. The language has been documented within the framework of the DoBeS endangered languages project (Bruce Birch, Nick Evans and others).

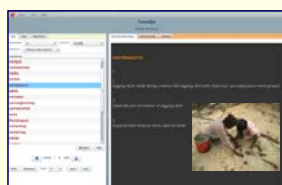
As a follow up on the documentation project the Iwaidja team is aiming at the creation of an accessible and attractive online dictionary, with multimedia capability and cross-referenced conceptual spaces will contribute to conservation, revitalization and empowerment of the language and community. As a pilot we have imported the current Iwaidja lexicon into LEXUS. The lexicon consists of 3500 entries and was originally created in Toolbox. The lexicon has been formatted and initial enrichment with multi media has started.

In ViCoS we have created a conceptual space of the concept 'rarlwa' (generic term for oyster). The conceptual space has been enriched with links, not only to the lexical entries in LEXUS, but also to information sources on the world wide web. From the ViCoS browser, objects can be linked to external resources, like for instance Wikipedia.

Another functionality of ViCoS which is explored in the Iwaidja pilot is the option of viewing the whole conceptual space in one view. Although the view is a rather dense network of relations between objects, within one glance the user of the tool can see which concepts or objects are most relevant in the created conceptual space.



Languages of the Cobourg Peninsula



Iwaidja lexicon



Iwaidja overview of CS

Max Planck Institute for Psycholinguistics
Technical Group

Jacqueline Ringersma
Marc Kemps-Snijders
Claus Zinn
Huib Verweij

