

## Visualizing conceptual spaces (ViCoS) in LEXUS

Jacqueline Ringersma and Claus Zinn

Max Planck Institute for Psycholinguistics  
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands  
[jacqueline.ringersma@mpi.nl](mailto:jacqueline.ringersma@mpi.nl)

### Introduction

The language documentation projects of the DOBES program [1] are entering a new stage. In the first years of the program the projects were mainly occupied with the collection of video and audio recordings of a broad variety of spoken text genres in various contexts and socio-cultural interactions. Most of the material is described by metadata and stored in the archive for linguistic resources housed at the MPI [2], and part of the documented material is transcribed using ELAN [8]. For 45% of the documented languages lexica have been created, mostly using SIL's Toolbox for linguists [3]. Entering the new stage, documentation teams are facing the challenge of data exploitation. Given the increasingly urgent purpose to help saving endangered languages from extinction, documentation could contribute to maintain and possibly revitalize a documented language, and consequently, to present and preserve the cultural heritage of its speech community. This educational purpose complements the community-external purposes of linguists, ethnologists, and anthropologists.

We are describing LEXUS [4], a tool that can offer the documentation projects computer support for exploiting documentation resources. LEXUS is a web-based lexicon tool to create and view multimedia encyclopedic lexica. LEXUS interacts with archived documentation material through IMDI for archive browsing and metadata search [5] and ANNEX for searching and viewing annotated multi media resources [6]. Furthermore LEXUS is a new component in archiving, since it has possibilities of linking together linguistic as well as cultural concepts, thus creating semantic networks of ontological knowledge that complement existing lexical knowledge. ViCoS [7] is an extension of LEXUS, allowing users to link words to the concepts they denote, to create relations of user-defined types between concepts, to visualize the resulting conceptual spaces, and to navigate from the conceptual space to the lexicon and *vice versa*.

Two language documentation projects are intensively involved in the development of LEXUS and ViCoS:

- Towards a multimedia dictionary of the Marquesan and Tuamotuan languages of French Polynesia [9]
- Creation of a dictionary and semantic network for the Papua New Guinea language Yélf Dnye [10]

Other DOBES language documentation projects are also contributing ideas and suggestions for the development of LEXUS and ViCoS, namely, Semang, Iwaidja, Yurakaré, Kuikuro.

### LEXUS: basic functionalities

Lexica in LEXUS consist of a lexicon schema, defining the structure of the lexical entries, and the lexical entries that are conform to the schema. A schema consists of data categories (*e.g.*, 'part of speech') and data category groups (*e.g.*, 'descriptionGroup', containing the data categories, say, 'description' and 'note'). Schemata, thus, can be highly structured. LEXUS users can create lexica from scratch, or import lexica that were created in other formats. LEXUS offers import functionality for structured and unstructured Toolbox lexica and lexica created in Chat and XML. For the export of lexica, XML as well as the Toolbox format are supported.

In LEXUS, the lexical entries are presented in a word list. The formatting of the word list view is user definable; it usually defines information units (*e.g.*, the data category 'lexeme', followed by, say 'English definition') that, when displayed, facilitate quick browsing through the lexicon. The order of the word list's character set is also user-definable, allowing users, for instance, to order the character ['i] under [i], or to view it as a separate character. Thus, sort orders that deviate from the defined Unicode character ordering can be displayed taking into account linguists' configuration. Selecting a lexical entry from the word list displays the lexical entry view, a more elaborate (and user-definable) view of the lexical entry. The lexical entry view allows displaying (and playing, when applicable) multimedia items like images, sound files and video fragments (Fig. 1). A full view of the data for one lexical entry can be obtained by selecting the edit option in the word list. This will display the values of the data categories for the selected lexical entry and an edit mode where existing values can be changed or new ones entered.

### ViCoS: from lexical spaces to conceptual spaces

The LEXUS extension ViCoS aims at empowering users to complement linguistic information with ontological information. Users may manipulate a world of concepts and their (inter)relations rather than just lexical entries. Links can be created originating from entries within a lexicon or across lexica. The source and target of a relation can be a complete lexical entry, or any part (*i.e.*, data category) thereof.

Users can make use of predefined relation types or define new ones (see below). The creation of such relations, and the definition and use of new relation types, may result in multi-dimensional concept networks representing indigenous knowledge on a specific theme or subject.

ViCoS offers the option to browse conceptual spaces (navigating between concepts) as well as freely moving between conceptual spaces and lexical spaces (LEXUS) and *vice versa*. The functionality opens the opportunity of actively involving people from the language communities in adding rich semantic/encyclopedic information, complementing the work of linguists. ViCoS thus needs to take into account: (i) the various user groups and their computer literacy levels (linguists, anthropologists and members of the speech community); (ii) the broad range of interest and expertise of the different user groups (iii) the connection between the lexicon and the ontological space; and (iv) a sound technical framework for the representation of ontological knowledge.

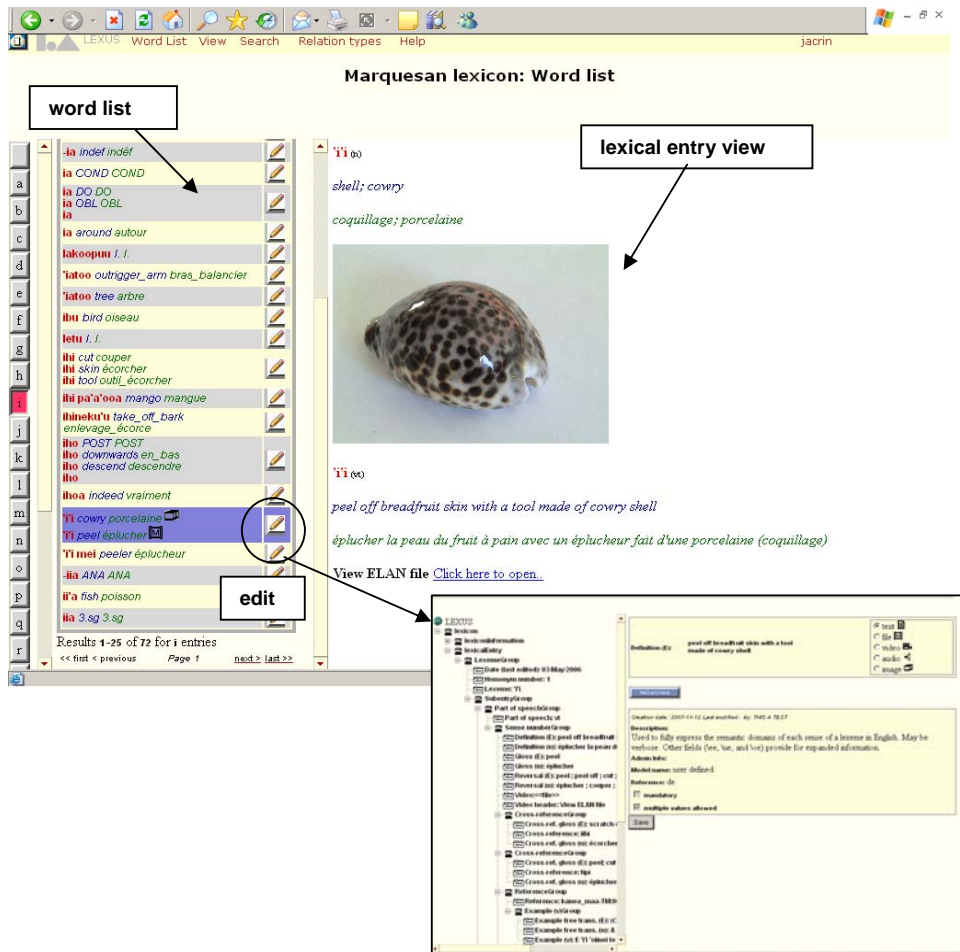


Figure 1: LEXUS: word list, lexical entry and edit view for lexical entry 'i'i' (from Marquesan lexicon [9])

### ViCoS in action

The Yéfi Dnye team [10] aims at creating a conceptual space that represents Russell Island's natural world from an ethno-biological perspective. Although there are no existing biological taxa, the content of the Yéfi Dnye lexicon hints at a great richness of traditional knowledge about the biological environment. The lexical entries, thus, provide an ideal starting point for such ontological enterprise.

ViCoS provides an easy but powerful interface for the creation of conceptual spaces (Fig. 2). The interface consists of three window frames: (1) the **Word List frame** where lexical entries are displayed with a minimal view for easy browsing through the lexicon; (2) the **Lexical Entry frame** where selected lexical entries are displayed, mirroring LEXUS's lexical entry view; and (3) the **Conceptual Workspace frame** where relations can be created. The **alphabet pane** at the top allows users to make display all words (in the word list frame) that start with the character selected. The alphabet displayed follows the sort order information as defined in LEXUS.

In Fig. 2 we have selected four lexical entries: *te* (fish) and *ghee*, *koo* and *tootpi* (three fish species) in the **Lexical Entry frame**. A relation between the four entries has been created by dragging the entries representing the fish species in 'Source' and the *te* entry into the 'Target' options of the **Conceptual Space frame** (3), and by selecting the relation type 'is-a' to connect them. Likewise, any relation between any pair of lexical entries can be created. ViCoS offers pre-defined relation types: *hyponymy (is-a-kind-of)*, *meronymy (is-a-part-of)*, *holonymy (has-part)*, *is-an-instance-of*, *is-synonym-of* and *is-antonym-of*. Users may also define new relation types. To view the conceptual space for a lexical entry, the user should drag that lexical entry from the **Lexical Entry frame** (2) into the 'Show in Knowledge Space' option of the **Conceptual Space frame** (3). The corresponding concept will take center stage in the conceptual space browser, surrounded by all concepts that are directly related to it (Fig. 2). The graphical editor and browser offers the following modes when clicking on a concept node: browse (show additional information, stemming from the corresponding lexical entry); move (changing the position of nodes); connect (creating new relations between concepts, see below); delete (removing concepts or relations); clear (deleting all items); refetch (reloading the conceptual space of the center node from a server); save (storing a conceptual space) and LEXUS (open the lexical entry in LEXUS, Fig. 2). The conceptual space can also be used to create relations: In the ViCoS user interface the user can drag&drop a lexical entry from the **Lexical Entry frame** into the 'Add to knowledge space' option of the **Conceptual Space frame**. In the graphical editor, when in 'connect' mode, users can simple draw edges between all nodes (part of the conceptual space, or dropped there) to create relations between them.

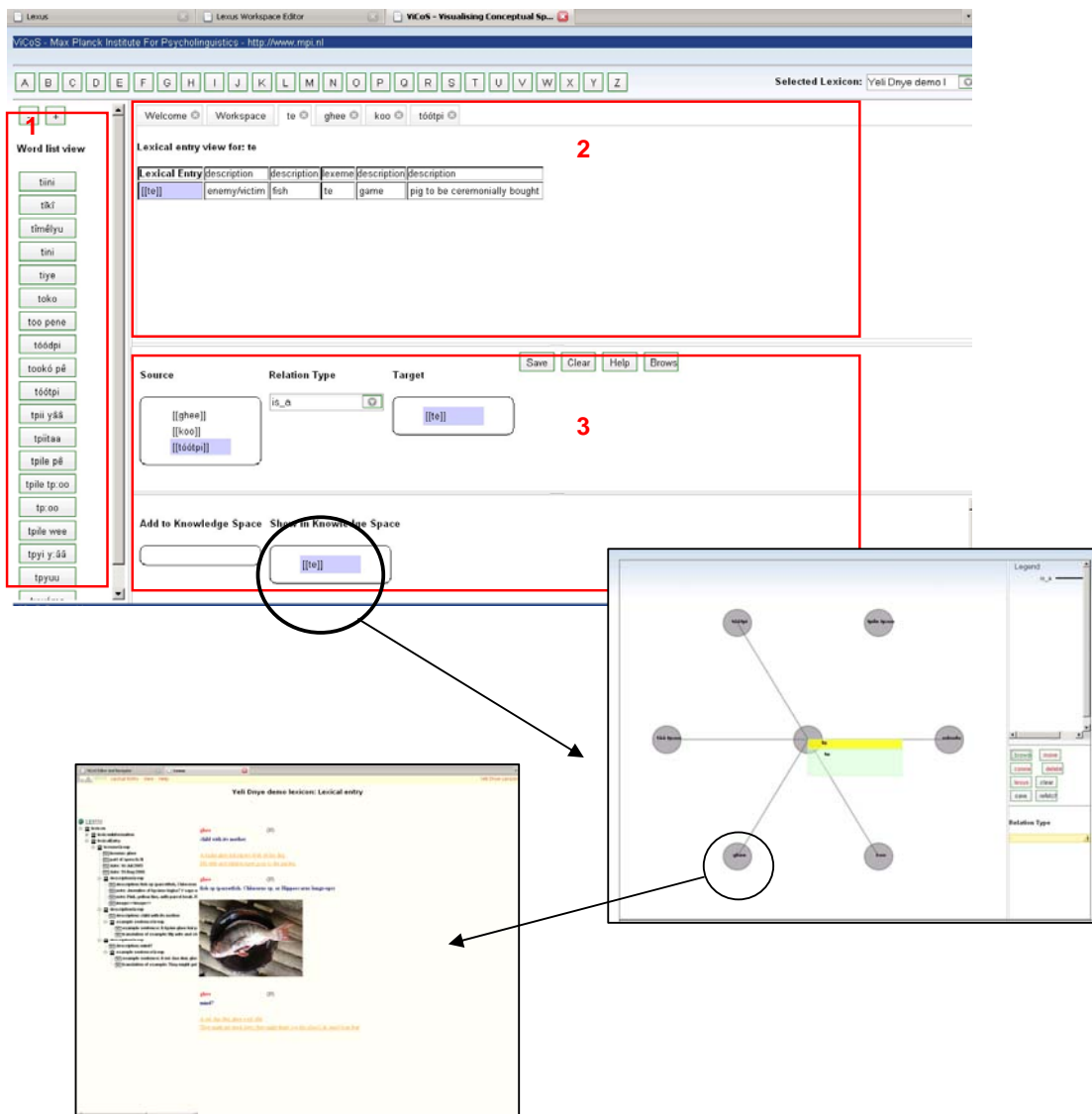


Figure 2: ViCoS user interface and conceptual space for lexical entry *te* (fish) from Yéli Dnye lexicon

## Conclusion

LEXUS and ViCoS are tools under ongoing development. Whereas LEXUS is already available for use by the documentation teams (10 lexica from the DOBES program have been imported in LEXUS), ViCoS is still in a test phase and only available for a selected number of pilot users, who are asked to cooperate in the development. Currently there are four documented languages with lexica and knowledge spaces: Marquesan, Tuamotuan and Semang from the DOBES program and Yélî Dnye (MPI project). First feedback from these teams acknowledges the need for the ViCoS module and is positive on the usability. We are currently implementing additional functionality that aims at supporting users in the creation of conceptual spaces: (i) an interface to LEXUS' search engine where search results can be easily entered into the conceptual space; and (ii) a mechanism that suggests, for each noun, a conceptual space that contains its corresponding concept as well as candidate concepts to which it may be related to. These candidates will stem from the lexical entry, exploiting, *i.e.*, "semantic" data categories, or from word co-occurrences within example sentences (*e.g.*, fish – dolphin – bait).

**Acknowledgement** We would especially like to thank Gaby Cablitz (University of Kiel, Germany) for her contributions to the development of LEXUS and ViCoS.

**Note** We would like to invite documentation teams to discuss their options of using LEXUS and ViCoS in the exploitation phase of the projects.

## Links

- [1] DoBeS: <http://www.mpi.nl/dobes>
- [2] Archive for linguistic resources: <http://corpus1.mpi.nl>
- [3] SIL Toolbox: <http://www.sil.org/computing/toolbox/>
- [4] LEXUS: <http://www.lat-mpi.eu/tools/lexus/>
- [5] IMDI: <http://www.lat-mpi.eu/tools/imdi/>
- [6] ANNEX : <http://www.lat-mpi.eu/tools/annex/>
- [7] ViCoS: <http://www.lat-mpi.eu/tools/vicos>
- [8] ELAN: <http://www.lat-mpi.eu/tools/elan>
- [9] Marquesan project: <http://www.mpi.nl/DOBES/projects/marquesan>
- [10] Yélî Dnye project: Pioneers of Islands Melanesia <http://www.eastpapuan.ling.su.se>