

Conceptual Spaces in ViCoS

Claus Zinn

Max Planck Institute for Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
Claus.Zinn@mpi.nl

Abstract. We describe ViCoS, a tool for constructing and visualising conceptual spaces in the area of language documentation. ViCoS allows users to enrich existing lexical information about the words of a language with conceptual knowledge. Their work towards language-based, informal ontology building must be supported by easy-to-use workflows and supporting software, which we will demonstrate.

1 Research Background and Application Context

Language documentation aims at the creation of a representative and long lasting, multipurpose record of natural languages. Such documentation contributes to maintain, consolidate or revitalize endangered languages, and inherently, also contributes to the description of cultural elements of a language community [1]. Our aim is to increase this cultural aspect by allowing users to complement linguistic information — defined mainly by lexica and annotated media recordings — with ontological information. Our approach is centered around the notion of knowledge spaces, where users model and manipulate a world of concepts and their interrelations rather than just lexical entries.

In language documentation, both linguistic and ontological resources are targeted for *human consumption*. By and large, there are two main user groups: scientists such as linguists and anthropologists, and members of the language community. Scientists may contribute to and exploit resources to study the language and culture of a community, or to compare them to the ones of other communities. This user group requires technology for intelligent search, potentially across lexica with their diverse structure and metadata, and thus, a machinery that can cope with data heterogeneity. The focus for the second user group is a different one; the empowerment of community members to actively participate in describing their language 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, house building, 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. This user group requires effective technology for the construction of knowledge spaces, a sort of *informal* ontology of fuzzily-defined concepts and relationships, that enables community members to anchor the words of a linguistic resource, or the objects of a multimedia archive, to such associations.

2 Related Work

Given existing lexical and multimedia resources, users shall be empowered to represent, communicate, and thus preserve complex facts about their language and culture. Resources may help eliciting their tacit knowledge, but tools are required that help transform it into tangible representations that can be viewed and manipulated by others. Resulting conceptual structures may then facilitate the revival of a shared understanding within the community. Mind mapping and concept mapping software are a good starting point to support this task.

Mind maps form a tree structure as there is only a single central concept to which other concepts can be related to. Usually, the nature of these relations is left underspecified. *Concept maps* are more expressive. First, there is no single central concept so that the hierarchy of concepts form a graph structure. Moreover, it is possible to name the relations that hold between concepts, and links between concepts can be directed. There is, however, no formal definition of node types (classes, instances), or arc types (defining the relation types). *Ontologies* can be seen as a formalisation of concept maps as they are based on using formal languages, and thus, concept and relation types have a well-defined semantics.

Existing tools for ontology engineering (*e.g.*, Protégé) have only limited relevance for a community-driven knowledge engineering effort as their proper and effective use requires a considerable amount of expertise in knowledge representation formalisms. Moreover, the use of existing ontologies, in addition to the need to know their content and structure, is problematic as they induce a significant (and usually Westernized) bias of how the world should be modelled. Existing tools for constructing mind/concept maps are much easier to use but, for our purpose, lack integration with lexical resources and multimedia archives.

Our ViCoS tool is based on the notion of *conceptual spaces* (CS), where a concept must be anchored to its lexical resource(s), and where it can be illustrated with archived material (images, sounds, videos). In contrast to concept maps, we require users to define relation types before their first use. Their definition is semi-formal but help addressing some interoperability issues when doing studies across lexica and conceptual spaces. In contrast to ontology builders, there is, thus, no formal semantics of concepts or relations. In this sense, conceptual spaces are closer to topic maps [6] rather than RDF/OWL-based approaches. Conceptual Spaces also bear much resemblance to the representational aspect of ConceptNet [7], a semantic network where concept nodes (semi-structured English fragments) are interrelated by twenty rather pragmatic semantic relations.

3 ViCoS

ViCoS is a simple but effective tool that aims at empowering a broad base of users to describe and link together those words (and the concepts they denote), which they feel are important in their language and culture. ViCoS is coupled with Lexus [2], our tool for building and maintaining lexical resources, and has access to an archive that hosts multimedia objects and their annotations [3]. ViCoS is a web-based application requiring a Javascript-enabled web browser.

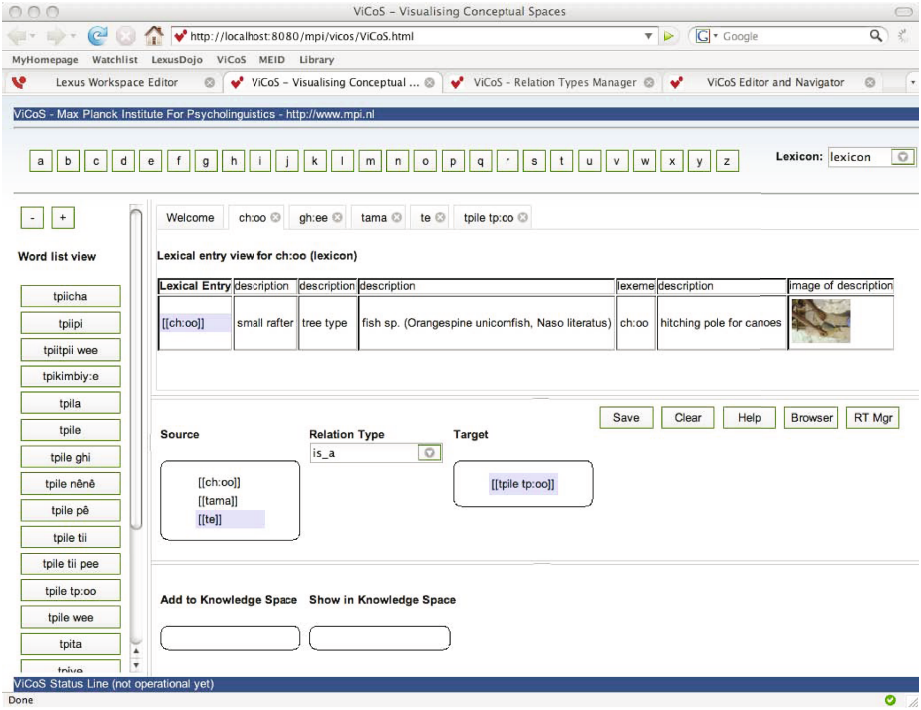


Fig. 1. ViCoS Main Window, showing the Yéli Dnye example lexicon)

Fig. 1 depicts ViCoS’ main window consisting of: the Lexicon Selection and Start Letter Selection frame; the Word List frame where lexical entries are represented by their lexeme; the Lexical Entry frame where selected lexical entries are displayed; and the Conceptual Workspace frame where relations can be created.

In the screenshot, we have selected the Yéli Dnye demo lexicon, and clicked on the letter “t” to obtain (from Lexus) all lexemes that start with this letter. We then selected “te” (Engl. “fish”) from the Word List frame to display its lexical entry. We then copied the complete lexical entry “[te]” (via drag&drop) to become the target of a “is-a” relation. In a similar way, we obtained the two fish species “ch:oo” and “ghee” as sources. Likewise, any relation between any pair of lexical entries can be created. With a tab mechanism that maintains access to prior selections of lexical entries, links can also be created across lexica.

ViCoS offers a few standard pre-defined relation types such as hypnomy, meronymy, holonymy, synonymy, and antonymy. Users may also define new relation types. To view the conceptual space for a lexical entry, the user drags this lexical entry from the Lexical Entry frame into the “Show in Knowledge Space” option of the Conceptual Space frame. The lexical entry’s corresponding concept will take center stage in the conceptual space browser, surrounded by all concepts that are directly related to it (see Fig. 2).

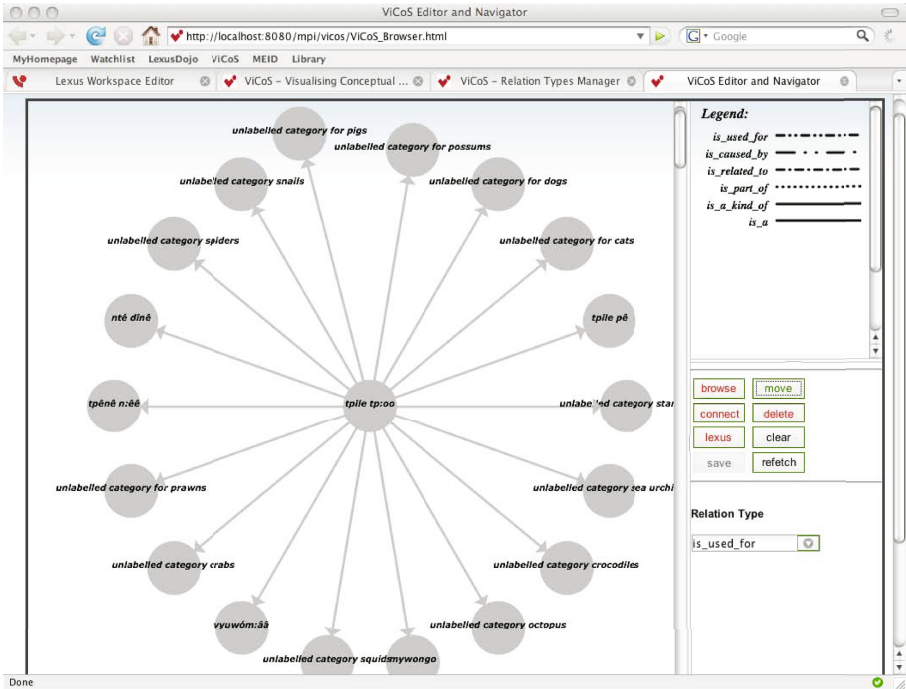


Fig. 2. ViCoS Browser Window, showing a few classes and their interrelations

The graphical editor and browser offers various modes: in Browse mode, when clicking on a concept node, additional information that stems from the lexical entry is displayed; and in Connect mode, new relations between concepts can be drawn. A double-click on a node will show a CS with the chosen node as centre, together with all concepts it is directly related to. There is also functionality for moving around nodes, or deleting concepts and relations. In Lexus mode, clicking on the node with open Lexus to show the corresponding lexical entry in linguistic context, thus strongly linking together lexical and ontological spaces.

On the back-end, ViCoS uses the OWL knowledge representation language [4]. Conceptual spaces are stored within the JENA framework, which provides a programmatic environment for OWL constructs as well as a rule-based inference engine for retrieving implicit information [5]. Once a query user interface is in place, users can then ask ViCoS, for instance, whether a given speech community categorises a dolphin as a type of fish, or whether dolphin meat is used as ingredient for food recipes. The reasoning engine will exploit class-subclass hierarchies and the characteristics of relation types, for instance, their domain and range, and whether they are symmetric, transitive, or functional.

ViCoS has entered a test phase with end-users. First feedback from two documentation teams has been positive. We are currently implementing new functionality that aims at supporting users in the creation of conceptual spaces: an interface to Lexus' search engine where search results can be easily entered into

the CS space; and a mechanism that suggests, for each noun, a CS that contains its corresponding concept as well as candidate concepts to which it may be related to. These candidates will stem from the lexicon, exploiting, i.e., “semantic” data categories, or from word co-occurrences within example sentences.

4 Discussion

Language documentation projects aim at preserving endangered cultural heritage for future generations, ideally within the communities themselves. Thus an active involvement of community members is crucial for such projects to succeed. To overcome the limitations of a purely linguistic approach to language documentation, we use knowledge engineering methods that allow members of indigenous communities to play an active role in the documentation process. This emphasises that a language is so much more than a list of lexical entries and their scientific description in linguistic parlour. Our approach turns words into culturally relevant concepts and places them in relation to other concepts. It attempts to engage and inspire community members to explore and to extend the resulting knowledge space. Because our design preserves the relationship between lexical and ontological space, users can browse them more or less simultaneously and can thus gain a richer experience of the language and culture being documented. In a way, our approach bridges scientific resources (lexica constructed by linguists; and multimedia assets annotated by experts) with indigenous knowledge resources (knowledge spaces constructed by community members).

Ensuring that emerging CSs stay manageable instead of becoming chaotic and hard to interpret will be a main challenge. One could hope that the users themselves organise the spaces they build. In particular, relation types will be an issue as it is unclear whether the built-in relation types will be accepted and properly used by community members. It is also to be seen how their set grows, and how well they will be documented and reused; regular interventions of moderators may prove necessary to clarify their semantics.

Acknowledgements. I would like to thank Jacquelijin Ringersma (MPI) and Gaby Cablitz (University of Kiel) for testing ViCoS, and for reporting their feedback.

References

1. Documentation of Endangered Languages (DoBeS), <http://www.mpi.nl/dobes>
2. Lexus, <http://www.lat-mpi.eu/tools/lexus>
3. Archive for linguistic resources, <http://corpus1.mpi.nl>
4. OWL Web Ontology Language Overview. W3C Recommendation from (February 10, 2004), <http://www.w3.org/TR/owl-features>
5. JENA, <http://jena.sourceforge.net/index.html>
6. Topic Maps, <http://www.ontopia.net/topicmaps/materials/tao.html>
7. ConceptNet, <http://web.media.mit.edu/~hugo/conceptnet>