LEXUS: a web based tool for the creation of multi media encyclopedic dictionaries and lexica

ViCoS: a web based tool for constructing and visualizing conceptual spaces. Enrichment of lexical information with conceptual knowledge
Some history:

2006: Do we (&you) need a lexicon tool?

Yes:
(1) standardisation for interoperability between lexica, data exchange and archive exploitation.
(2) community needs to participate in the construction of lexica
LEXUS & ViCoS:
From lexical to conceptual spaces

Long and winding road

2005: First version of LEXUS developed with some input from DoBeS
2006: Construction of Yeli Dnye multi media lexicon (Steve Levinson)

Multimedia lexicon

Typed relations within the lexicon
2007: Start of Marquesan and Tuamotuan dictionary project 
(Gaby Cablitz, Edgar Tetahiotupa, Jean Kape, others)

* Many improvements on the LEXUS functionalities
* Many improvements on interoperability with Toolbox and XML
* Initial development of ViCoS, based on ideas from Gaby and the 
  Tuamotuan participants in the project
* Input on the new LEXUS user interface from the Marquesan 
  participants in the project
LEXUS & ViCoS: From lexical to conceptual spaces

Long and winding road

2007-2008: Input from other DoBeS projects to LEXUS
LEXUS & ViCoS:
From lexical to conceptual spaces

Long and winding road

2008-2009: Further refinements of ViCoS
New LEXUS user interface
Long and winding road

2008-2009: Further refinements of ViCoS
New LEXUS user interface
Long and winding road

2009: Launch of the tools, Workshop April 2, 2009

2006→2009: Do we (&you) still need these tools?

**Christina Lai Truong, Lilian Garcez:** Participatory Methods for Language Documentation and Conservation: Building Community Awareness and Engagement

“many language documentation programs, even when done in collaboration with a few members of the target community, do not engage with larger segments of the community”

**Christopher Doty:** Bridging the Gap between Linguistics and Community: Producing materials for language maintenance
Long and winding road

2009: Launch of the tools, Workshop April 2, 2009

2006→2009: Do we (&you) still need these tools?

**Tania Granadillo:** Speaker-driven approach to language documentation

“Without the collaboration of the speakers it is impossible to document a language”

**Ulrike Mosel:** Turning the linguist’s lexical database into a community dictionary

“the differences between a lexical data base as it typically results from a language documentation project and the kind of dictionaries the speech community wants”
LEXUS

Based on two ISO TC 37 standards for linguistic resources

LMF: Linguistic Markup Framework (lexicon structure)
DCR: set of standardized data categories to be used as a reference for the definition of linguistic annotation schemes or any other formats used in the area of language resources (concept naming)

LMF/DCR:
• A modular structure for content interoperability between lexical resources.
• Archiving exploitation framework, XML based

Web based (!):
• Users work in LEXUS workspaces
• Workspaces can be shared by users with different rights (read-only, editor)
• Workspaces can contain multiple lexica
• Creation of lexica from scratch, import lexica from other formats (Toolbox, XML, Chat)
• User definable views of word list and lexical entries

<table>
<thead>
<tr>
<th>Word list</th>
<th>Lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>pa:ckesii</td>
<td>ti:ctawi:qi acs hacki:pa:ckesiih</td>
</tr>
</tbody>
</table>

*It is hard for me to hear*  
ta'i uc tawi:qi acs hacki:pa:ckesiih  
pres pocacg previal bedifficult hang be good qiv oneruh heur subp ixc preq  

<table>
<thead>
<tr>
<th></th>
<th>Lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>ti:ckwhak:i:hi:qi acs hacki:pa:ckesiih</td>
<td></td>
</tr>
</tbody>
</table>

*It is hard for me to hear*  
ta'i uc whakich hi:hi:qi acs hacki:pa:ckesiih  
pres pocacg previal acrual behard bedifficult ixc be good qiv oneruh heur subp ixc preq
LEXUS: the lexical space

- Creation of lexica from scratch, import lexica from other formats (Toolbox, XML, Chat)
- User definable views of word list and lexical entries
- Linking multi media fragments to lexical entries

Audio
Video
Image
Link to archived files
LEXUS: the lexical space

- Creation of lexica from scratch, import lexica from other formats (Toolbox, XML, Chat)
- User definable views of word list and lexical entries
- Linking multi media fragments to lexical entries
- Setting sort orders for each data category
LEXUS: the lexical spaces

- Creation of lexica from scratch, import lexica from other formats (Toolbox, XML, Chat)
- User definable views of word list and lexical entries
- Linking multimedia fragments to lexical entries
- Setting sort orders for each data category
- Using different character sets (UTF-8 is required)
Towards a multi-media dictionary of the Marquesan and Tuamotuan languages of French Polynesia

- Building a digital multi-media encyclopedic dictionary with LEXUS
- Improving basic LEXUS functionalities
- Conceptual spaces
- Improved User Interface

Project team:
- Linguist team (Gablitz, Mosel)
- Technical coordination, testing and implementation (Ringersma)
- Developers (Kemps, Zinn)
- Speech community (Kape, Guillome, Tetahiotupa, Tahia, Mataiki, Bruneau Pati)
Issues that came up:

User Interface

Conceptual spaces in multi media encyclopedia
Conceptual spaces in multi media encyclopedia

Conventional (paper) dictionaries: network of meanings less visible

Paper dictionaries limited usefulness in language maintenance and language revival (Manning et al., 2000)

Members of speech community prefer following semantic links of different semantic types (synonyms, antonyms, lexical, taxonomies)
From LEXUS to ViCoS

Conceptual spaces in multi media encyclopedia
Existing lexical resource is starting point
  Words offer key access

ViCoS is technology to link words, images etc. – and the
associations they evoke – to other words, images etc.

Ease-of-use to empower community members to
  actively describe their language & culture
  and to learn from such resources
  resources targeted for human consumption

Resulting Conceptual Spaces = sort of informal ontology of
fuzzily-defined concepts and relationships
  but where concept nodes are anchored in corresponding formal
  resources (lexica)
ghee (N)

child with its mother

A kpám ghee kní yapwo tědde dě leec dmi. My wife and children have gone to the garden.

ghee (N)

fish sp (parrotfish, Chlororus sp, or Hipposcarus longiceps)
User definable relation types
Default ‘Universal’ relation types

- **is_a_kind_of**
- **is_not_a_kind_of**
- **is_part_of**
- **is_antonym_of**
- **is_synonym_of**
- **is_related_to**
- **is_caused_by**

**Relation Type Definition for: is_a_kind_of**
To delete, press here: [delete]

**Name of relation type**
is_a_kind_of

**Import relation type from DCR - ISO 12820**

**Characteristics**
- Functional
- Symmetric
- Transitive

**Linetype**
(solid)

**Colour**
These ideas originate from the DoBeS documentation team (Beaver language): Carolina Pasamonik, Dagmar Jung, Gabriele Müller.

The ideas are still young and fresh, need further thinking

‘User definable relation types’ also allows to define kinship relations:
Additional requirements:

differentiate the shape and colour of the ‘concept’ link from arrow to object
Two dictionaries for the KSDP DoBeS project (Michael Riessler):

**SaRuDic:**
Based on: Electronic Saami dictionary of Jur’jev (2003-2007)
Kildin Saami – Russian
Rich structure and content – including derivations and example sentences
Including multi-media fragments

**RuSaDic:**
Russian-Kildin wordlist
Preliminary stage (905 entries)
Limited structure
Limited content (no multi-media)
One of the project aims is revitalization of the Saami language. Primary language for most speakers is Russian.

**RuSaDic** serves as an index for **SaRuDic**. Pointers from RuSaDic to SaRuDic are created using ViCoS and visible in LEXUS.
Additional information on the Saami entry comes from external data bases:

Using ViCoS it is possible to create a link between a LEXUS object and any other online object (like e.g the Algu dictionary or Wikipedia):
Web based:
Advantages
Share lexica between users (read rights)
Multiple user input to the creation of lexica (write rights)
Access from (all) over the world

Disadvantages (technical)
Not everyone has (fast) Internet access (yet)
Browser dependencies
Display of characters (even when in UTF-8)

Discussion (non-technical)
Should there be a chief-editor of the lexicon, or can it work like Wikipedia?

Maybe more in Gaby Cablitz presentation?
Display of characters is browser dependent. It depends on the supported UTF-8 set of the browser. Particularly IE has problems there. In new UI (Flex based) this will/should be the same for all browsers

Marc Kemps-Snijders, 3/4/2009
Launch workshop:

April 2, 2009 in Nijmegen (The Netherlands)

On the program:
Gaby Cablitz (Marquesan and Tuamotuan)
Carolina Pasamonik (Beaver)
Presentation of the Iwaidja lexicon and conceptual space
Tutorials

Creation of a user group to input in the further development of the two tools

Want to participate in the workshop or in the user group? 

jacquelijn.ringersma@mpi.nl