

Jacquelijn Ringersma

Claus Zinn

Marc Kemps-Snijders



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



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





Long and winding road

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



Long and winding road

2007-2008: Input from other DoBeS projects to LEXUS

	LEXUS Word List View Se	arch R	elation types Help				bru	icebirch		
			LEXUS Word List View Se	earch Re	latio	on types Help			niclas	
						Malayall:	Word	list		
	esults 1-25 of 3468 for ng entries									
i a	<pre>< first < previous Page 1 next Lexical Entry</pre>		Results 1-25 of 205 for all entries		<u> </u>	animal				
j	ng	a	<< first < previous Page 1 next	<u>t > last >></u>		Batek deq kuala koh	'?aj			
k	ng	b	all	2		Batek teh pasir linggi	'Dai			
n	- Iga-	с	animal			Details tell proper in i				
la	nga-	d	ashes			Datek ten sungai taku	'7aj			
T n	zabi	e	back (of person)			Jahai sungai mangga	'?aj		Click here to open.	
ng	nga Idajanat	f	bark (of tree)			Jahai sungai rual	pa'ŋẽn			
E	gadbun-	B h	belly			Kensiw lubok legong	'?aj			
	ngadburnduj		big			Kensiw to	'ha'p0?			
n	ngadburrung	j	bird			Kentag bukit asu	'?ai			
th N	Ngadikili	k	bite			I anah kartai				
u _ n	ngajabiny	1	blood				7aj			
		m	blow (of wind)			Menriq kuala lah	'?aj			
		n		абрсэ		Monrie annesi cuel				
				абрсэлл	19			TRANSLATION		
				абрсэлн	э			дождь м		



Long and winding road

2008-2009: Further refinements of ViCoS New LEXUS user interface

Lexus View Help		
	Yeli_Dnye	
List Tree Selections	Lexical Entry View Lexical entry Search	
StartLetter: a V Lexicon: Yeli_Dnye V	lexicalEntry	
Selection:		
List items	lexemearoup	
alamga	lexeme	alamga
	part of speech	Ν
	date	19/Aug/2006
	descriptionGroup	
	description	fish sp. (Anthinae spp)
	note	Anthia family (Atnhiinae), red/pink, e.g. Pseudoanthias bimaculatus
Acsult 1 of 373		
First Previous Page 1 V Next Last	Multimedia Properties Notes	-



Long and winding road

2008-2009: Further refinements of ViCoS New LEXUS user interface

ical Entry View	Lexical entry	Search	
Searc	h Quick search		
Lexio	on: Yeli_Dnye		
wher	lexeme	▼)
ch:oo		search	
		contains	
		begins with	
		ends with	1
•	1	ш	•



Long and winding road

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

2006 \rightarrow 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 \rightarrow 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

LEXUS Word List View	Search Relat	on types Help	wichita lexicon				
Wichita corrected glottal stop: Word list							
Word list pa::wasi pa:: pa::piha:: pa::pipa:: pa::pipa:: paciya:: pachir paciya:: paciya: paciya: packepe pa::kepes packepespi pa::kepesi pa::kepesi pa::khari pa::kharip Results: 4-25 of 2577 for all entrice		 Lexical entry 7a:cke?esi v <i>hear</i> ti:ctawi:ji acs hackija:cl it is hard for me to hear tta i2 uc tawi:ji tawi acs ha-ki t parckepesi pres pfocagt prevdat bedifficult difficult ha ti:ckwakhárhi:ji acs hack ti:s hard for me to hear ta i2 uc wakhahr hi:ji hi: ji acs ha-ki t parc pres pfocagt prevdat activpat behard bedifficult 	képesih h ng be good sjtv onesub hear subp xx prevq kipa: cképesih kepesi h ñcult xx be good sjtv onesub prevq hear subp				
	LEXUS Word List View VVord list ya::wasi ya::ya::wasi ya::pa::pa::pa::pa::pa::pa::pa::pa::pa::	LEXUS Word List View Search Relative Wich Word list pac.	LEXUS Word List View Search Relation types Help Wichita corrected glottal stop: N Vord list Image: Construction of the construle of the construction of the construction of the cons				



- 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





- 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 Lexicon Structure View Search	Relation types Help jacrin
2	Marquesan lexicon: Lexicon structure
W LEXUS	Grand Log 1
E-	Exeme C Hatural
금 🛱 LexemeGroup Em Lexeme	• Other:
Image: Construction of the second	a(['a][-a]['A][-'a]AaÀàÁáÂâÃãÄäÄäÅå)b(Bb)c(CcÇç)d(DdĐð)e(['e][-e]['E]EeÈèÉéÊêËë)f(Ff)g(Gg)h(Hh)i Description: The Record marke speakers. Admin info: to be filled out D mandetary
Importantion (ii) Importantion (iii) Importantion (v)	✓ multiple values aloved Sort Order: ○ Itatural ○ Other: a(['a][-a]['A][-a]AaÀàÁáÂâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅâÅ



- 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
- 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)





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)







User definable relation types

ViCoS - Max Planck Institute For Psycholinguis	ics - http://www.mpi.nl
	Lexicon: Dico général - te 🔻
Relation Types	Welcome Help is_product_of S is_material_for S
New Relation Type	Relation Type Definition for: is_material_for. To delete, press here: delete!
is_product_of	Name of relation type
is_variety_of is_image_of	is_material_for is_material_for building).
is_medicine_for	Import relation type from DCR - ISO 12620
	import!
	Characteristics
	Functional Symmetric Transitive
	Linetype
	(dot) -
	Colour



Default 'Universal' relation types

		Lexicoli: Universal
	Welcome Help is_product_of S	is_material_for 🛽 is_a_kind_of 🕲
Relation Types		
is_a_kind_of	Relation Type Definition for: is_a_kind_o	of. To delete, press here: delete!
is_not_a_kind_of		
is_part_of	Nome of relation type	
is_antonym_of	is a kind of	
is_synonym_of	modify!	The relation A is a kind of B specifies that A has as hypernym B , or that B is a
is_related_to		
is_caused_by	Import relation type from DCR - ISO	Example: car is_a_kind_of vehicle
	12020	This relation is a universal relation and can thus be applied across lexical resour
	im port!	
	Characteristics	
	Functional Symmetric	
	Linetype	
	(solid) 💌	
	Colour	





 MyHomepage Watchlist LexusDojo Lexus ViCoS MEID Library JSON Lexus Workspace Editor ViCoS - Visualising Conc ViCoS - Visualising Conc 	ngManager.html
MyHomepage Watchlist LexusDojo Lexus ViCoS MEID Library JSON ✓ Lexus Workspace Editor ✓ ViCoS - Visualising Conc ✓ ✓	ONLInt ViCoS Image Manager 💿 <table-cell> ViCoS Editor and Naviga S <table-cell> ViCoS - Relation Types S</table-cell></table-cell>
ViCoS - Visualising Conc	ViCoS Image Manager 💿 ¥ ViCoS Editor and Naviga 🕲 ¥ ViCoS - Relation Types 🕲
Image Map Shapes:	
add delete save	
Coords: 247, 142, 29 Href: www.tree.com A	Alt: tree Target: self
oricle	Alt: rainbox Target: self





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:

ViCoS - Max Planck Institute For Psycholinguistic	s - http://www.mpi.nl				
		L			
Relation Types	Welcome Help child of Stepsibling of S	ibling of 🛽 spouse 2 of 🖉 fathe			
New Relation Type	Relation Type Definition for: father of. To delete, press he	re: delete!			
father of					
mother of spouse of	father of modify!	n 🗠 😽 🛱 📋 🖪 👔			
spouse 2 of	Import relation type from DCR - ISO 12620 [No description supplied yet.]				
sibling of stepsibling of	im port!				
stepchild of	Characteristics				
stepfather of	Functional Symmetric Transitive				
stepmother of	Linetype				
	(dash) 💌				
	Colour				



ViCoS functionalities – ideas for 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 revilalization 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 visable 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?

MKS3 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