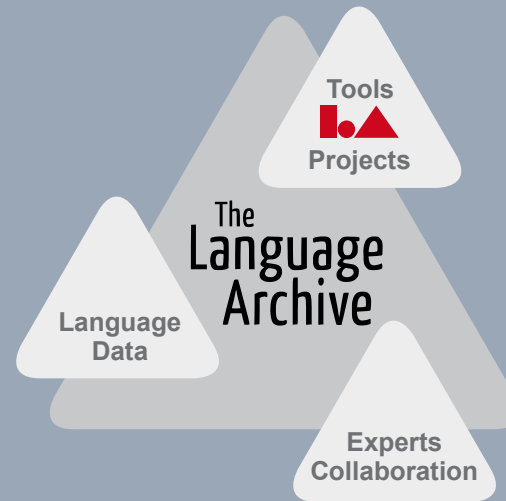


# CLARIN

Common Language Resources and Technology Infrastructure



MAX-PLANCK-GESELLSCHAFT



## LEXUS

Collaborative multimedia lexicon creation

Alexander König

The Language Archive - Max Planck Institute for Psycholinguistics  
Nijmegen, The Netherlands



# Lexica - What is out there?



MAX-PLANCK-GESELLSCHAFT

## Structure

Structure of lexicon depends on language, linguistic theory, purpose

## Conceptual

Large variation in linguistic concepts and value naming

## Format

Large variation in formats (XML, Shoebox, Chat, Word)

## Data interoperability problem

cross lexica searches, merging, comparison

## Archive requirements

Representation format (XML)



# LEXUS concepts



MAX-PLANCK-GESELLSCHAFT

## Structure

Structure of lexicon can be completely free (LMF or MDT templates to get you started will soon be available)

## Conceptual interoperability

Naming of linguistic concepts (fields) is free. Interoperability is ensured by pointing to a Data Category Registry (ISOcat)

## Format

LEXUS handles lexica internally in an XML database. Import from and export to other common formats is possible (XML, Toolbox, FLEX)

## Archive requirements

Exported LEXUS XML can be archived



# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

- Creation of lexica from scratch
- import lexica from other formats (Toolbox, XML)

The screenshot displays the LEXUS software interface. The title bar shows 'File Switch to Help Alexander Koenig'. The main window is titled 'NewLexicon: a new LEXUS lexicon'. On the left, a 'Lexicon structure' tree view shows a hierarchy: 'lexicon' (expanded) contains 'lexiconInformation', 'lexicalEntry', 'Lexeme', 'form', 'phonetic form', and 'sense' (selected). Under 'sense', there are 'definition (E)' and 'definition (H)'. On the right, the 'Schema element' view is active, showing 'List view' and 'Lexical entry view' tabs. The 'General information' section includes: 'Name: sense', 'Description: Contains attributes that describe meanings of a lexical entry', 'Admin Info:' (empty field), 'Mandatory: ', and 'Multiples allowed: '. Below this, there are 'Sort order' and 'Notes' buttons, a 'Sort order' dropdown menu set to '<No sortOrder present>', and a table with columns 'StartLetter' and 'Characters'.



# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

- Creation of lexica from scratch
- import lexica from other formats (Toolbox, XML)
- User definable views of word list and lexical entries

The screenshot displays the LEXUS web interface. At the top, there is a browser window with the address bar showing 'http://corpus1...selectedIndex=0'. Below the browser window, there is a navigation bar with 'File', 'Switch to', and 'Help' menus. The main content area is titled 'Galerie de photos: Link between lexemes, photo galleries, photo captions and corresponding video clips'. The interface is divided into two main sections: 'Lexicon' and 'Lexical entry'. The 'Lexicon' section on the left contains a list of lexical entries, each with its name, description, and a link to a photo gallery. The 'Lexical entry' section on the right shows a detailed view of the entry for 'apau kokomi', including a photo of a hand holding a piece of wood and a photo of a person preparing a drink.

Advanced search — Max Planck Institute | CallList - TG Wiki | http://corpus1...selectedIndex=0 | VICO5 - Visualising Conceptual Spaces

File Switch to Help marquesan

Galerie de photos: Link between lexemes, photo galleries, photo captions and corresponding video clips

Lexicon Filters

<switch lexicon> <grapheme> Lexical entry Lexical Entry View


Switch to structure view <filter>

Lexical entries

- 'apau kokomi *médicament ; purge pour des bébés* medicine ; purgative for babies galerie de photo
- faraoa *pain bread* galerie de photo
- haïka 'eka kira *médicament ; purge* medicine ; purgative galerie de photo
- haïka 'epa *médicament ; purge (femmes)* medicine ; purgative (women) galerie de photo
- haïka hati vaevae *médicament (jambe cassée)* medicine (broken leg) galerie de photo
- haïka hō 'enana motua *huile de massage* massage oil galerie de photo
- haïka hō toiki *huile de massage* massage oil galerie de photo
- haïka ho 'oi kivakiva *médicament (rein)* medicine (kidneys) galerie de photo
- haïka ko'i 'ē 'eli *médicament (bébé, fille)* medicine (baby girl) galerie de photo
- haïka mokio *médicament (bébé, fille)* medicine (baby girl) galerie de photo
- haïka pūhenua *médicament (après accouchement)* medicine (after birth) galerie de photo
- haïka putuhulu toiki *médicament contre bobos* medicine (spots) galerie de photo
- haïka tekéo ika *médicament (empoisonnement)* medicine (fish poisoning) galerie de photo
- haïka tōto me 'ama *médicament (menstruation)* medicine (menstruation) galerie de photo


'apau kokomi

01



'apau kokomi

02



Page 1



# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

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

The screenshot displays the LEXUS interface with two overlapping windows. The background window shows a list of lexical entries for 'Poisson-pêche - Ta- ava'a'. The foreground window shows a detailed view of the entry 'fafa'ua', which includes a definition, usage notes, and a photograph of a manta ray. A third window on the right shows a list of entries with audio links.

**Audio**  
**Video**  
**Image**  
**Link to archived files**



# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

- Creation of lexica from scratch
- import lexica from other formats (Toolbox, XML)
- User definable views of word list and lexical entries
- Linking multimedia (fragments) to lexical entries
- Defining custom sort orders

Start letter	Characters
a	aA[-a][-A][^a][^A][\(\d\)a][^\(d\)a][^\(rt\)a][^\(
b	bB[-b][-B][^b][^B]
d	dD[-d][-D][^d][^D]
i	iI[-i][-I][^i][^I][^\(g\)i][^\(rt\)i][^\(w\)i]iïfîiï
j	jJ[-j][-J][^j][^J]
k	kK[-k][-K][^k][^K]
l	lL[-l][-L][^l][^L]
ld	[ld][LD][-ld][-Ld][^ld][^Ld]
m	mM[-m][-M][^m][^M]
n	nN[-n][-N][^n][^N]
ng	[ng][Ng][-ng][-Ng][^ng][^Ng]
r	rR[-r][-R][^r][^R]
rr	[rr][Rr][-Rr][-rr][^Rr][^rr]
rt	[rt][Rt][-rt][-Rt][^Rt][^rt]
u	uU[-u][-U][^u][^U][^\(w\)u]ùûáúáúáú
v	vV[-v][-V][^v][^V]
w	wW[-w][-W][^w][^W]
y	yY[-y][-Y][^y][^Y]ÿŸ



# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

- Creation of lexica from scratch
- import lexica from other formats (Toolbox, XML)
- User definable views of word list and lexical entries
- Linking multimedia (fragments) to lexical entries
- Defining custom sort orders
- use of non-ASCII characters / Unicode compatible





# LEXUS functionalities



MAX-PLANCK-GESELLSCHAFT

- Creation of lexica from scratch
- import lexica from other formats (Toolbox, XML)
- User definable views of word list and lexical entries
- Linking multimedia (fragments) to lexical entries
- Defining custom sort orders
- use of non-ASCII characters / Unicode compatible
- Creating complex filters and queries



- **Toolbox**
  - interaction between lexicon and annotation
  - Doesn't strongly enforce a structure
  - file-based
    - not dependent on an Internet connection
    - can modify lexicon with other tools
  - not as good with multimedia integration



- **Fieldworks Language Explorer (FLEX)**
  - successor of Toolbox
  - close integration of annotation and lexicon
  - can handle multimedia better than Toolbox, but still not great
  - no longer file-based



- **Lexique Pro**
  - Can be used to create nice-looking versions of Toolbox lexica
  - Is not usually used as a lexicon editor itself
  - Is meant to be used in combination with Toolbox or FLEx



# LEXUS: Strengths and Weaknesses



MAX-PLANCK-GESELLSCHAFT

- + Lexica can be shared with others (read or write)
- + Flash-based: cross-platform
- + Easy embedding of local or archived multimedia
- + Can create different representations of the same source (e.g. PDF export)
- + Open Source (GPL)



# LEXUS: Strengths and Weaknesses



MAX-PLANCK-GESellschaft

- GUI is a bit cumbersome (esp. view creation)
- Only usable online
- No interaction with annotation tool
- Flash-based: a dying platform (see iOS)



## **What's planned for the immediate future?**

- Offline Version
- Closer integration with ELAN
- Integrate LEXUS into TLA Authentication Infrastructure
- Add preview functionality to the Views Editor
- Improve Archivability of LEXUS lexica



# LEXUS in Action



MAX-PLANCK-GESELLSCHAFT





# Questions?



MAX-PLANCK-GESellschaft



<http://tla.mpi.nl/tools/tla-tools/lexus/>

Alexander.Koenig@mpi.nl