



# LEXUS

Marc Kemps-Snijders  
Peter Wittenburg  
Max-Planck-Institute for Psycholinguistics



# Goal LEXUS

- driven by LMF AND **the actual needs of our users**
- all about 100 users at MPI and related projects
  - all with different lexical structures but re-using concepts
  - example: Shoebox categories
- unification is urgent since we work on
  - proper archival representation and
  - unified archival exploration



# Goal LEXUS

**Overview of ingested lexical entries**

Below the elements are listed which have been ingested into the system based on the dictionary file you have provided. By clicking on an element you will be directed to the detailed view of the entry.

Lexical Entry	
Lexeme: a	
Date (last edited): 24Mar/1997, sh v3.0.400 SHLEX SAMPLE	
Lexeme: a	
Date (last edited): 24Mar/1997	
Lexeme: a	
Date (last edited): 23Feb/1998	
Homonymy number: 1	
Lexeme: a	

Repräsentation:

Chinesisch | Pinyin | Deutsch | Gloss

我们先看看菜单。  
Wǒmen xiān kàn kàn càidān .

**LEXUS-Lexical Entry Viewer**

Lexical Object: Sense number

name	value
Encyclopedic info (E)	This is not linked to coconut groves but is used for mangrove, etc.
Class (C)	grove
Class (N)	dusun

TRUMAI		Language
Language		Trumai is an isolate language, that is, its genetic affiliations are unknown. Possibly the other languages that were related to Trumai already disappeared without any kind of documentation or record, which then would make Trumai the only existent representative of an entire linguistic family.
Geography		
People and Culture		
Project Description		
Team		
Data Examples		
Project website		
Links		

位	wèi	WUEI	
你	nǐ	nǐ	du
先	xiān	ciɛn	vorher, im Voraus; zuerst
先生	xiānsheng	ciɛnʂəŋ	Herr
刀叉	dāochā	tauʂhʰa	Messer und Gabel



# LEXUS functionality

- implementation of Lego bricks idea of LMF
- allows to create components recursively
  - of course the core model and extensions may be modelled
- allows to add data categories to components
- allows to grab data categories from DCR's
  - ISO DCR of course
  - Shoebox MDF DCR
  - urgently need MPI ISO compliant DCR
- multimedia support
- schema validation
- import/export(Shoebox, LMF XML)
- local and webbased operation
- customizable layout
- search
- eager to get comments and extend on our current ideas



# Future enhancements

- Definition and use of value ranges
- merging of lexical versions (collaboration on lexica)
- MPI and personal DCR's (ISO compliant)
- Import from XML sources

- paradigms
- scripts
- conditions
- operational triggers

waiting on specifications (LMF ?)

- interaction with ELAN annotation tool  
EAF based on Abstract Corpus Model  
which is an extension of AG





# Future enhancements

- Definition and use of value ranges
- merging of lexical versions (collaboration on lexica)
- MPI and personal DCR's (ISO compliant)
- Import from XML sources

- paradigms
- scripts
- conditions
- operational triggers
- interaction with EAF based on ... which is an extension

The screenshot shows the Elan software interface. At the top, there is a menu bar (File, Edit, Search, View, Play, Options, Help, Commands) and a toolbar with various playback and search icons. Below the menu is a video window showing two people in a room. To the right of the video is a table of annotations:

Nr	Annotation	Begin time	End time	Duration
3	that's the oranje single	00:00:13.310	00:00:14.540	00:00:01.230
4	then you follow the sign kleeft	00:00:15.330	00:00:17.450	00:00:02.120
5	you come down	00:00:17.570	00:00:18.750	00:00:01.180
6	you know eh after this trajanus plein	00:00:18.750	00:00:21.860	00:00:03.110

Below the table is a 'Text Viewer' window showing the transcript of the audio. The text is: 'so you go out of the Institute to the Saint Anna Straat. and then you go the other, Saint Anna Straat to this to the center of the town, to this big rotunde. and you follow then the sign Kleeft that's the oranje single then you follow the sign kleeft you come down you know eh after this trajanus plein you come down to the rhine eh valley yeah that's'.

At the bottom of the interface is a detailed linguistic analysis of the audio. It includes a waveform, a spectrogram, and a series of text layers for different linguistic levels:

- K-Spch**: there is another pl
- W-Spch**: then you follow the sign kleeft | you come down | you know eh after this trajanus plein | you come down to | rhine eh v
- W-Words**: then you follow the sign kleeft | you come down | you know eh after this Trajan Plein | you co do to the | Rhine
- W-POS**: adv pro v art n n | pro v adv pro v post prep dem n n | pr v ad pr art | n
- W-IPA**: ðen ju fɔfə ðə sa ŋ kleeft | ju kɔm daŋ | ju n ə aftə ðɔs trajanus pleŋ | ju kɔm daŋ tɔ | ra ŋ ə v



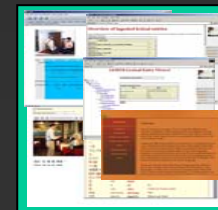
# Future enhancements

- Definition and use of value ranges
- merging of lexical versions (collaboration on lexica)
- MPI and personal DCR's (ISO compliant)
- Import from XML sources

- paradigms
- scripts
- conditions
- operational triggers

waiting on specifications (LMF ?)

- interaction with ELAN annotation tool  
EAF based on Abstract Corpus Model  
which is an extension of AG
- interaction within web-based exploration  
framework





# LMF Core model

LEXUS - Microsoft Internet Explorer

File Edit View Favorites Tools Help

## LEXUS-Lexicon Scheme Viewer

Lexical Extension:

LEXUS

- LMF core
  - LexicalEntry
    - lexicalentry
      - Form
      - sense

Model name: to be filled out  
Cardinality: 0..n  
Description: to be filled out  
Admin info: to be filled out  
Content features:

name	description	cardinality
------	-------------	-------------

Save

DCR  
[Lexica schemas](#)  
[Create schema](#)  
[Ingest schema](#)  
[View schemes](#)  
[Lexica contents](#)  
[search](#)

```
graph TD; lexicon --- global_info[global info]; lexicon --- lexical_entry[lexical entry]; lexical_entry --- form; lexical_entry --- sense;
```

Done Local intranet



# LMF model (shoebox import)

LEXUS - Microsoft Internet Explorer

File Edit View Favorites Tools Help

## LEXUS-Lexicon Scheme Viewer

LEXUS

- SHLEX SAMPLE
  - LexicalEntry
    - Lexeme
      - Citation form
      - Status
      - Lexeme
      - Homonym number
      - Date (last edited)
    - Subentry
      - Borrowed word (loan)
      - Morphology
      - Subentry
      - Phonetic form
      - Main entry cross-ref.
    - Part of speech
      - Non-animate dual
      - First plural
      - Second dual
      - First plural exclusive
      - Part of speech (n)
      - Non-animate singular
      - First singular
      - Reduplication form(s)
      - First plural inclusive
      - Part of speech
      - Third singular
      - Third plural
      - Second plural
      - First dual
      - Singular form
      - Third dual

You may move the elements around the form to adjust their position.

**Save**

**Lexeme**

**Description**

The Record marker for each record in a lexical entry. It contains the lexeme or headword (which is commonly monomorphemic). Since such a lexeme form is often not accessible for vernacular speakers if printed, use the %c field to provide a more readable form for vernacular speakers.

Notes (grammar)

This field is for any grammar notes that are pertinent to the lexeme that you wish to keep separate from general notes.

Date (last edited)

**Description**

A Shoebox bookkeeping field to help keep track of the last time an entry was edited. One per record (usually the last field) is adequate. Usually inserted automatically by Shoebox. The %stamp field is set up under the shoebox menu option: Database-Properties-Options tab.

Gloss (v)

This field is available for developing a monolingual dictionary. While somewhat near-synonymous with the morphological gloss (E) field, the gloss (v) field is given to the particular paradigm form in English.

Paradigm form gloss (E)

Used for glossing the vernacular paradigm form in English.

Variant comment (r)

Where regional language comments can be given for the variant form.

Paradigm

Used for specifying the morphological structure for complex lexemes.

Only/restrictions (r)

For denoting in the regional language any semantic and/or grammatical restrictions pertinent to the lexeme.

[DCR](#)

[Lexica schemas](#)

[Create schema](#)

[Ingest schema](#)

[View schemas](#)

[Lexica contents](#)

[search](#)



# Example entry containing multimedia

Lexicon Tool - Microsoft Internet Explorer

## LEXUS-Lexical Entry Viewer

[View schema](#) [View data sheet](#) [View layout](#)


**LEXUS**


- test
  - LexicalEntry
    - Lexeme
      - user defined cf: Peter
      - First dual: my value
      - abbreviation: my value 2
      - example file: <<file>>

**user defined cf**  
**Description**  
to be filled out  
Peter

**abbreviation**  
**Description**  
A designation formed by omitting words or letters from a longer form and designating the same concept.  
my value 2

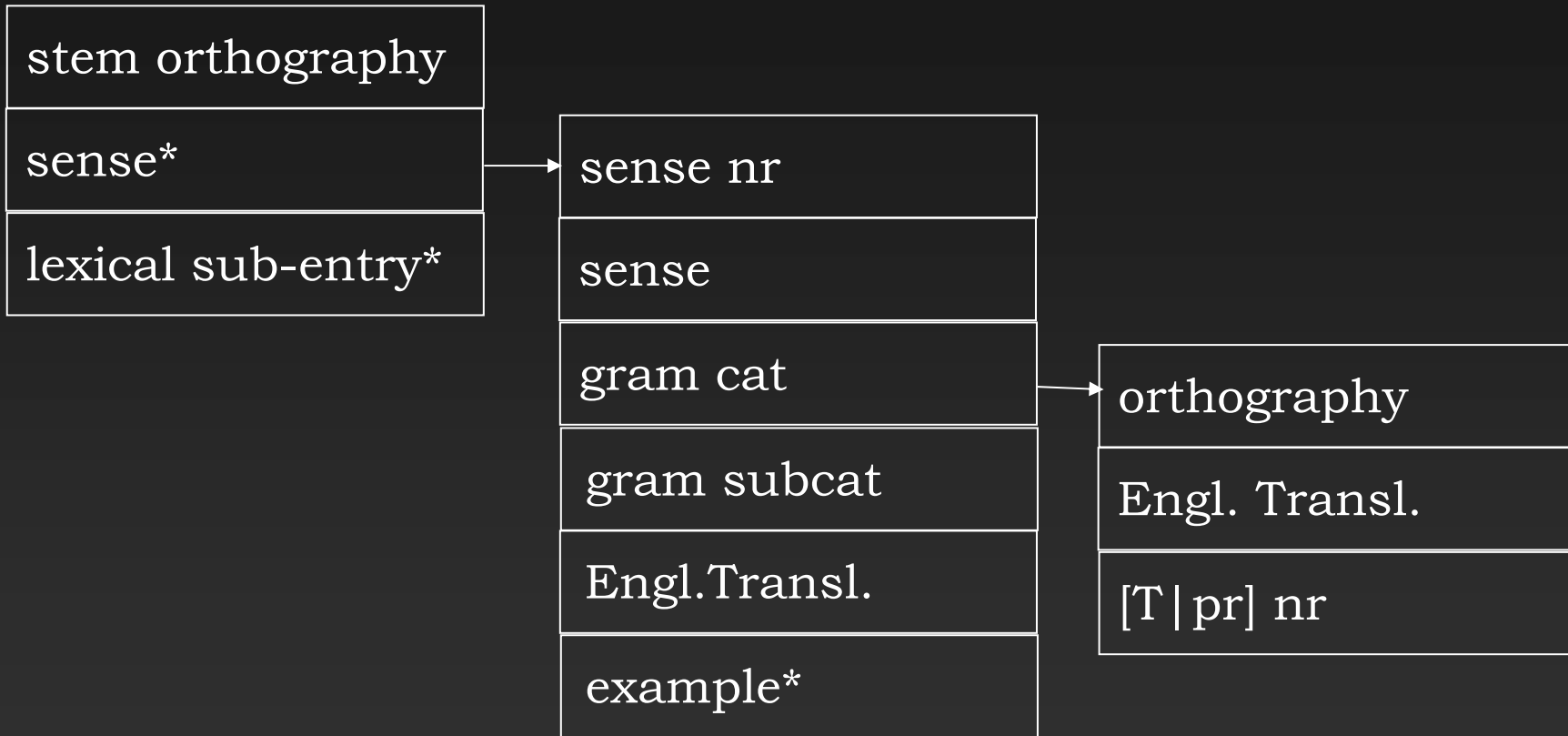
**First dual**  
**Description**  
Used to give the vernacular for this particular paradigm form. (It is better to use the new \pdl field set for this.)  
my value

**example video**  
**Description**  
This feature describes an example video  


**example file**  
**Description**  
This feature describes an example file  




# Example concrete lexicon



part of TEOP lexicon used in DOBES project

\* indicates 1:n relations