

> Digitale

Ressourcen in den GSHS ... am Beispiel LEXUS

GSHS LIBRARY CONFERENCE

Florenz 10.11.2006 <



Marc Kemps-Snijders, Jaqueliijn Ringersma, Peter Wittenburg

MPI for Psycholinguistics, Netherlands <

Dr. Susanne Alt

Centre National des Ressources Textuelles et Linguistiques - CNRS, France <



Lexus

> ein « flexibles, internet-basiertes Werkzeug für elektronische Wörterbücher »

Lexical Resource Search View extra willem wever

(WORKSPACE) LEXUS-Lexicon Scheme Viewer

cururui 2. *see cururui (wuu yiye), Date 21/Jul/2005*

Lexeme: tpii kããdi, *Gloss: humid weather, Date 01/Sep/2004*

Lexeme: tpii tpii kããdi kããdi, *Gloss: off and on rain and sunshine, Date 01/Sep/2004*

Lexeme: tpii yãã, *Gloss: fish type, Date 27/Jul/2005*

Lexeme: tpiicha, *Gloss: 1. self-control 2. upbrining, Date 23/Aug/2004*

Lexeme: Tpiinkënë, *Gloss: man's name, Date 21/Dec/2000*

Lexeme: tpiipi, *Gloss: steersman / captain*

Lexeme: tpiitpii wee, *Gloss: flying ants, Date 01/Sep/2004*

Lexeme: tpiitaa, *Gloss: Pig Island, near snake passage, Date 11/Aug/1999*

Lexeme: tpiikimbiye, *Gloss: 1. break in pieces 2. disperse (crowd), Date 01/Aug/2002*

Lexeme: tpiila, *Gloss: steered, Date 16/Aug/2002*

Lexeme: tpile, *Gloss: thing*

Lexeme: tpile qhi, *Gloss: valuable (chomo, nkaa or wuluwulu), Date 17/Aug/2004*

Lexeme: tpile nënë, *Gloss: sing sing, Date 19/Aug/2002*

Lexeme: tpile pë, *Gloss: snake, Date 10/Feb/2006*

Lexeme: tpile tii pee, *Gloss: flat-type bagi, Date 19/Aug/1999*

Lexeme: tpile tpoo, *Gloss: animals, Date 10/Feb/2006*

Lexeme: tpile wee, *Gloss: 1. 'sing-sing' - one style of traditional song 2. insect, beetle 3. perform_singsing, Date 10/Feb/2006*

Lexeme: tpile pë, *Gloss: snake (lit. thing_long), Date 17/Apr/1996*

Sample lexicon for Rossel Demo 10/Feb/2006

Lexeme: tpile wee

Gloss: 1. 'sing-sing' - one style of traditional song, 2. insect, beetle, 3. perform_singsing

Example: 1. 'there's no singsing', 2. tpile we daa tóó/kwo

Free Translation:

Note: 1. No clear relationship to beetles! This song style contrasts with yaa, ntamë and other styles both musically and lyrically. Tpile wee are only sung by men and boys, in performances that last all night. Performers wear special grass skirts and carry special baskets with bushfowl eggs and nuts to be given to the audience. Each village or region has its own tpile wee repertoire - these are songs made up by known personages, but elaborated with topical verses on each performance. They are named works, like operas, usually with the name of a natural species. Tpile wee often accompany pig feasts, and sometimes new house feasts., 2. includes cockroach, stag beetles, wasps, ants, mosquitoes, white ants (and anything with __ wee) but excludes flies, bees, case-insect

1. 2.



Elektronische Wörterbücher

- > kosten- und zeitaufwendige Ressourcen
 - unabhkömmlich in der Linguistik, Psycholinguistik, Sprachverarbeitung, Ethnologie...
 - breiter nicht-wissenschaftlicher Interessentenkreis
- > unterschiedliche Beschreibungsebenen
 - Morphologie, Syntax, Semantik, Terminologie, Belege, Notizen, ...
 - einsprachig oder mehrsprachig, verschiedene Kodierungssysteme
 - verschiedene Formate (XML, SBX, RDB, Word...)
- > komplexer Workflow
 - Belegsammlung, Textkorpora, Feldarbeit : multimediales Material
 - Strukturierung eines Wörterbuches : Rubriken, Wertelisten, Verweise etc.
 - Redaktion : oft über verteilte Arbeitsplätze, mehrere Personen
 - Veröffentlichung : elektronisch, Internet, Printversion
 - Aktualisierung, Vervollständigung, Korrektur : Kohärenz bewahren, Versionen verwalten
- > Praxis : ständige ad-hoc Neuentwicklungen von Tools



Zwei Beispiele...

> MPI Nijmegen

- lange Erfahrung in der Archivierung von Sprachdaten für linguistische Forschung
 - Textkorpora, Ton- und Videoaufnahmen, Lexika (Celex) etc.
 - Werkzeuge zur Lexikon-Erstellung : « Shoebox »
- neue Herausforderungen im Dobes-Projekt (> 20 Sprachen)

> CNRTL Nancy

- lange lexikographische Erfahrung (Trésor de la Langue Française)
 - Literaturkorpora, manuelle Belegverwaltung, Digitalisierung, NLP-Lexika
 - Werkzeuge zur Lexikon-Erstellung : DTDs, XML-Editoren, Skripte
- neue Herausforderungen im Lexikon-Vernetzungs-Projekt

> ähnliche Problematik

- Bedarf an vergleichbaren Strukturen und Formaten (Suche, Fusion, Publikation)
- Bedarf an wiederverwendbaren Werkzeugen zur Konzipierung, Redaktion und Datenpflege



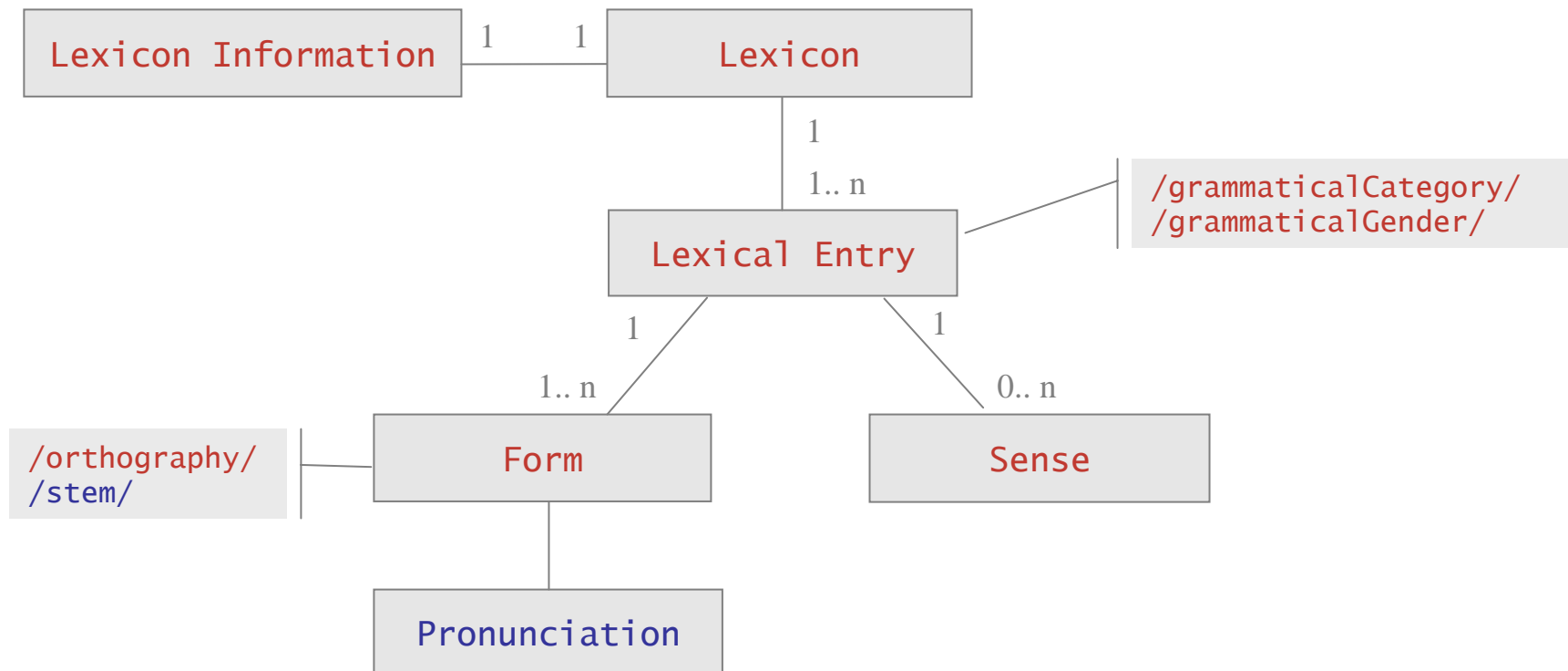
Rahmenbedingungen

- > ISO TC 37/SC 4
- > Arbeitsgruppe zu standardisierten Lexikonstrukturen
 - digitale Wörterbücher und maschinenlesbare Lexika (Sprachverarbeitung)
 - einsprachig oder mehrsprachig
- > grundlegende Idee
 - gemeinsame Informationsstrukturen identifizieren
 - als ausbaubare « Komponenten » standardisieren
 - umfassende linguistische Dokumentierung und Definition der Komponenten
- > bis 2004/2005
 - Draft mit Vorschlägen (Lexical Markup Framework)
 - weder Anwendung auf grosse Lexika noch Basis für Lexikon-Tool
- > Lexus



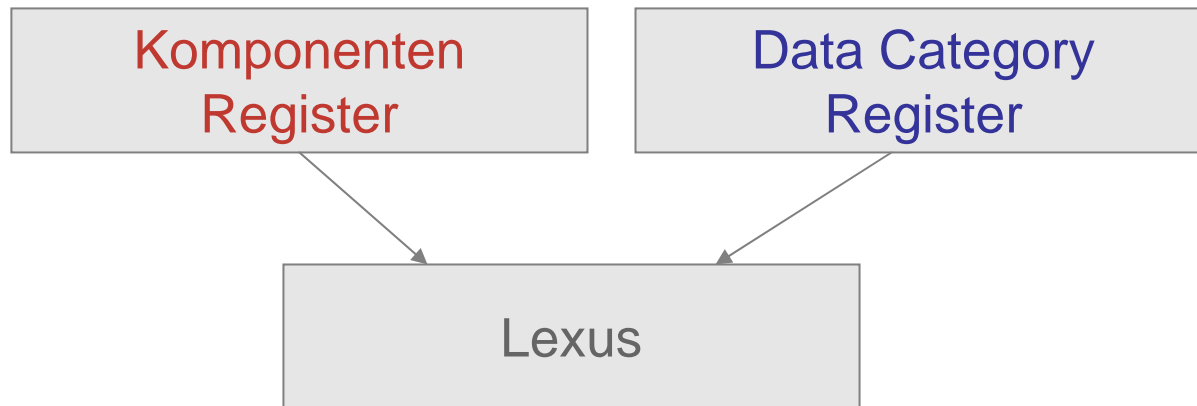
Lexical Markup Framework

> Komponenten, Attribute, benutzerdefinierte Extensionen





Lexus : Schnittstellen und Funktionen



- > Entwurf von ISO-konformen Wörterbuch-Strukturen
- > Redaktion
- > Verlinkung mit Ton- und Videodaten
- > HTML-Formatierung
- > Import, Konvertierung, Fusion, Export



Lexus : Benutzeroberfläche

The screenshot displays the Lexus Lexicon Scheme Viewer web application. The interface is divided into several sections:

- Top Navigation Bar:** Contains menu items: Workspace, Lexical Resource, Edit, Search, View, DCR, and the user name 'willem wever'.
- Left Sidebar:** A tree view under the heading 'LEXUS' with sub-items: LexicalDatabase, GlobalInfo, LexicalEntry, Form, Sense, and New Data Category.
- Main Content Area (Left):**
 - A dropdown menu is open, showing options: Copy, Paste, Delete, New Data Category, and New Component.
 - A 'New Data Category' dialog box is visible.
 - A secondary navigation bar includes: Workspace, Lexical Resource, Search.
 - Text: 'Please select the type of lexicon you would like to import' with radio buttons for 'Shoebox' (selected), 'Clan', and 'Wichita XML'.
 - Text: 'Please select whether you want to add a new lexicon or import data into existing lexicon..' with radio buttons for 'Create new lexicon' (selected) and 'Import data into existing lexicon..'
 - Section: 'Shoebox lexicon schema a'.
 - Text: 'This page allows you to import a shoebox lexicon. A new'.
 - Form fields for uploading files:
 - 'Please upload the Shoebox .typ structural file here.'
 - 'Please upload the Shoebox lexicon file here.'
 - 'Import' button.
- Main Content Area (Right):**
 - Text: 'Please enter your username and password.'
 - Form fields for 'Username:' (containing 'willem') and 'Password:'.
 - 'Submit' button.
 - Text: 'register as guest..' with a link.
 - Large text: 'LEXUS lexical resource tool'.



Lexus : auf einen Blick

> Werkzeug für Wörterbücher

- digitale Wörterbücher : Dobes-Projekt, Psycholing. Lexika, frz. WB aus Nancy, ...
- maschinenlesbare Lexika : Morphalou, Celex, ...
- klassischer Workflow, multimediale Vernetzung, Import/Fusion/Export

> Flexibel

- Aufruf von standardisierten Komponenten und Attributen
- benutzerabhängige Erweiterung der Lexikon-Strukturen
- benutzerdefiniertes Layout

> Internetbasiert

- über Standardbrowser abrufbar (Feldarbeit), Offline-Version
- Password und persönliches Arbeitsumfeld (Workspace)
- Verlinkung mit online-Dokumenten



Lexus... und dann ?

- > ausbaubar für spezifische Lexika
 - etymologische Wörterbücher
 - mehrsprachige Wörterbücher
 - Frequenzlisten, Wortlisten...
- > ausbaubar für onomasiologische Ressourcen
 - Terminologien, Fachwortschätze
 - Ontologien
 - Thesauri, Indexierungsvokabular, ...
- > ausgebaut wird nicht das Werkzeug selbst, sondern
 - die Komponenten- und Attributeregister
 - Extension für etymologische Informationen
 - Relationen zwischen (Teil-)einträgen für mehrsprachige WB
 - Terminologien : TMF (ISO-Standard derselben Familie)
 - Dokumentation wird automatisch aus dem Registerformat generiert



Lexus... und dann ?

> potentiell und unkompliziert erweiterbar auf

- digitale Wörterbücher : Dobes-Projekt, Psycholing. Lexika, frz. WB aus Nancy, ...
- maschinenlesbare Lexika : Morphalou, Celex, ...
- klassischer Workflow, multimediale Vernetzung, Import/Fusion/Export

> Flexibel

- Aufruf von standardisierten Komponenten und Attributen
- benutzerabhängige Erweiterung der Lexikon-Strukturen
- benutzerdefiniertes Layout

> Internetbasiert

- über Standardbrowser abrufbar (Feldarbeit), Offline-Version
- Password und persönliches Arbeitsumfeld (Workspace)
- Verlinkung mit online-Dokumenten